Influencer Detection meets Network AutoRegression – Influential Regions in the Bitcoin Blockchain

Simon Trimborn *[1], Hanqiu Peng[2], and Ying Chen[2,3]

[1]Amsterdam School of Economics, University of Amsterdam
[2]Department of Mathematics, National University of Singapore
[3]Risk Management Institute, National University of Singapore

December 8, 2022

Abstract

Known as an active global virtual money network, Bitcoin blockchain with millions of accounts has played an ever-growing important role in fund transition, digital payment and hedging. We propose a method to Detect Influencers in Network AutoRegressive models (DINAR) via sparse-group regularization to detect regions influencing others cross-border. For a granular analysis we analyze if the transaction record size plays a role for the dynamics of the cross-border transactions in the network. With two-layer sparsity, DINAR enables discovering 1) the active regions with influential impact on the global digital money network and 2) if changes in the transaction record size impact the dynamic evolution of Bitcoin transactions. We illustrate the finite sample performance of DINAR along with intensive simulation studies and investigate its asymptotic properties. In the real data analysis on Bitcoin blockchain from Feb 2012 to December 2021, we found that in the earlier years (2012-2016) network effects came surprisingly from Africa and South America. In 2017 Asia and Europe dominate whereas from 2018 effects majorly originate from North America. The effects are robust in regard to different groupings, evaluation periods and choice of regularization parameters.

Keywords: Bitcoin Blockchain, Network Dynamics, Two-Layer sparsity

JEL classification: C55, C58, C60, G17

*Corresponding author, phone: +31 643 611 771, E-Mail: trimborn.econometrics@gmail.com

1 Introduction

Powered by the blockchain technology, Bitcoin (BTC) brought an innovative financial asset class into the market. Just as banking customers in the United States send USD denominated transactions to settle their financial obligations, BTC blockchain users send BTC denominated transactions to each other, but as a borderless decentralized digital currency. BTC has grown into an active global virtual money network with millions of accounts. The number of BTC transactions has increased incredibly too. According to blockchain.com, there have been more than 730 million transactions completed by 1 May 2022. The average daily number of transactions was 91 in 2009, further rose to 69,084 in 2014 and peaked with more than 300,000 average daily transactions in 2019 and 2020, an annual increase of 300% from 2009 onwards.

Despite its impressive growth, public's attention is on the potential massive risks of BTC suffered from e.g. sudden price drops and liquidation risk. Several studies have reported high volatility and tail risk in BTC (Elendner et al., 2017; Feng et al., 2018), frequent jumps (Scaillet et al., 2018), informed trading (Feng et al., 2017), bubbles, sudden drops in market value (Hafner, 2018) and other common market risk factors (Liu et al., 2019). For example, the price dropped by 50% and 63% over two and four weeks in December 2013 and January 2018 respectively. An even worse case would happen if all users stopped interacting on the BTC blockchain, which would cause a suffer of losing all invested capitals to every user. This insolvency has already happened to 1705 cryptocurrencies (CCs), according to deadcoins.com. Although BTC seems safer as a much more liquid CC, risk is indeed higher than conventional financial instruments. Motivated by the huge market risks yet simultaneously desirable disruptive functions, there has been a demand on BTC exchanges as well as the construction of tunnels between BTC and traditional financial markets. This further triggered public concern on the impact of BTC on systemic risk. Unfortunately, one knows little about the BTC users' behaviour. While investors' behaviour can be studied via Limit Order Books of exchanges, the evolution of BTC price that is determined by the users' behaviour in the BTC blockchain is rarely studied. The anonymity of the BTC blockchain while providing high-end privacy protection masks the purposes of BTC transactions and often also the frequency of transactions. The dynamic evolution of the virtual money flows recorded in the BTC blockchain can provide a number of insightful implications about the users' behaviours. As a payment network, transactions in blockchain need to be carried out first to enable new transactions by other users. If the locations can be extracted from users' IP address, a combination of the blockchain and geographical information creates a valuable chance to study the influence of certain regions and certain types of users on the growth of the fast growing network.

For the borderless BTC, this question naturally has to be addressed at the global level. Europe and North America have for many years been considered as leaders in the financial markets. However the recent frantic enthusiasm for crypto mining in certain areas, in particular China, Japan and Korea as reported in the media, calls

into question the regional composition in the BTC blockchain network.

The first question of our study is:
Q1: Which regions influence other regions transaction behaviour on the BTC blockchain network?
We are interested in studying the regional effect in BTC blockchain given identifying leading regions provides insights for e.g. regulators who are interested in systemic risks. For exchanges where BTC are traded, Makarov and Schoar (2019) found cross-country arbitrage opportunities which widen up in times of strong BTC price increases. Sabah (2020) studied the relation between the geographical distribution and acceptance of BTC and the relation to the price performance. For the BTC blockchain, where the virtual currency is sent from user to user directly, Ron and Shamir (2013) analysed the transaction behaviours of the accounts from the emergence of Bitcoin until 13 May 2012. Foley et al. (2019) analysed the BTC blockchain for evidence of illegal activity and found 46% of BTC transactions are associated to respective cases. While studies like Lischke and Fabian (2016) have looked into the geographical distribution and the official vendor BTC transactions, little is known about the effect of regions on each other.

To investigate the regional interactions of the entire global network, we group the BTC transactions by continent. Inside each of these groups, the data are further split into 10 groups according to the transaction record size, resulting in all in 60 groups. These serve as a proxy for the wealth of a BTC investor, which is used to classify the users. The heuristic behind this is that only Bitcoiners with large holdings are able to execute large transactions, while small Bitcoiners contribute to small transactions. As surveyed by Trimborn and Yu (2022), heuristics are commonly applied to construct blockchain networks which can be analysed with analytic methods. This analysis is targeted at investigating if e.g. small/large transactions are linked to small/large transactions. Detecting the dynamic interactions between the regional and transaction record size groups can help to answer the second question:
Q2: Do transactions have an influence on transactions of similar size in other regions or do they impact transactions of different size?

Our analysis of the dynamic network activity finds the presence of serial cross-correlation. The existence of serial correlations motivates the adoption of Vector AutoRegressive (VAR) models for analysing the transactions of the BTC network. Already since Ord (1975), VAR has been used to investigate spatial interactions in networks. Pesaran et al. (2004) investigates the exposure of economies to each other, Chudik and Pesaran (2011) study Infinite-dimensional VARs under the assumption that each node is related to a small number of neighbouring nodes and a large number of non-neighbouring ones. Creal et al. (2013) propose Generalized AutoRegressive models and study the relation between exchange rates and credit risk ratings. Zhu et al. (2017) develop the Network vector AutoRegressive (NAR) model, where the connectivity of the network is represented by an adjacency matrix that is a given or pre-determined binary matrix, see also Zhou et al. (2017). Both papers assume that the dynamic network connectivity is controlled by one network parameter, which, in

combination with the given adjacency matrix, circumvents the dimensionality problem with large-scale networks. Though simple, modelling with one single network parameter and, more importantly, a known adjacency matrix, is a strong and unrealistic constraint for studying the BTC blockchain. While the geographical origin of a transaction can be identified, the geographical destination of a transaction is unknown, requiring an estimation of the adjacency matrix of the network.[1] This motivates using a flexible VAR model with unknown adjacency matrix, which encounters the overfitting problem for high dimensional networks. The estimation is often inefficient or even infeasible, unless one imposes some lower-dimensional structural assumptions, e.g. sparsity in the parameter space, see Basu and Michailidis (2015). A flexible modeling also aides to tackle the next research question of this study:

Q3: How is the network structure of the BTC blockchain related to the market state and developments in the cryptocurrency ecosystem?

Regularization approaches were originally designed for the univariate case in regressions, but have recently been brought to a vector time series context including the high-dimensional VAR models. In an investigation of large Vector AutoRegressive models with exogenous variables (VARX), Nicholson et al. (2017) propose five kinds of penalties. Song and Bickel (2011) assume a sparse structure for the lags and apply group sparsity to the columns of the parameter matrix. These studies build on the l_1/l_2-norm penalties proposed by Hoerl and Kennard (1988), Tibshirani (1996) and Zou and Hastie (2005), also known as ridge regression, the lasso, and naïve elastic net. Yuan and Lin (2006) develop the group sparsity method for regression models. The spline-lasso (Guo et al., 2016) allows for smoothly changing coefficients, which is motivated by the fused lasso (Tibshirani et al., 2005), encouraging locally constant coefficients within groups. Adopting both the l_2-norm and the l_1-norm in the regression context, Simon et al. (2013) develop an algorithm to search for the solution with group lasso penalization while allowing for individual penalizations inside of the groups.

We propose a method to Detect Influencers in Network AutoRegressive (DINAR) models via sparse-group regularization to study the dynamics in the BTC blockchain. The entries of the adjacency matrix are considered unknown and not necessarily binary, introducing a flexibility of the existence and level of connectivity in the network. By doing this, we essentially assume that only a few nodes are active. Moreover, diverse magnitudes of the parameters within the groups, with some being zero, implies the existence of individual sparsity. The sparsity assumption is necessary due to the huge dimensionality in combination with the limited data availability. The DINAR model adopts two kinds of sparsity. Group sparsity is applied to the columns (nodes) to identify the influential groups in certain continents, referred to as active nodes.

[1]The BTC blockchain uses relaying nodes to distribute the transactions to each participant. The IP address of the relaying node can be observed and provides an approximation of the origin of a transaction. Since the ownership of the funds is recorded in a public database distributed to each user and not in the accounts only, the destination of the transaction is not observable. For more details on the procedure used to observe this information, refer to Section 2.

Individual sparsity is imposed on the individual parameters in an active node, indicating that the active node does not have an effect on every other group. The proposed DINAR estimator with this two-layer sparsity enables discovering 1) the active regions with influential impact on the global digital money network and 2) if changes in the transaction record size impact the dynamic evolution of Bitcoin transactions. The DINAR estimator is designed for identifying influential regions and their groups within the region. For the optimization of DINAR, we develop an algorithm for the two-layer sparsity for high-dimensional networks. We investigate the performance of DINAR in a large simulation study on various numbers of observations and different strength of persistence which relates to different magnitude of parameters. We compare DINAR against LASSO, SCAD and BGR (Tibshirani, 1996; Fan and Li, 2001; Bańbura et al., 2010). The results of the simulation study show that DINAR is overall more accurate in identifying the influential groups, in particular for high dimensions. We also apply the 3 competing models to the real data investigation on the BTC blockchain data. The results show that neither LASSO, SCAD nor BGR are capable of identifying the influential groups. Indeed they either estimate only the autoregressive parameters or provide a messy adjacency parameter matrix estimate with much more parameters than DINAR whereas the AIC and BIC is usually better for DINAR.

This research is related to previous studies, yet there are several differences. DINAR regularizes the VAR specification for the individual parameter and the columns of the parameter matrix. The latter are all parameters associated with a single node and represent its effect on the system. The econometric models of Zhu et al. (2017), Davis et al. (2016), Ahelegbey et al. (2016), Bianchi et al. (2019), Billio et al. (2019) only focus on the individual parameters which limits the analysis to node-to-node insights. DINAR additionally allows for node-to-system analysis, hence it provides the influential nodes in the network. Further DINAR is based on regularization techniques. The work on Bayesian VARs by e.g. Bańbura et al. (2010), Ahelegbey et al. (2016), Bianchi et al. (2019), Billio et al. (2019) imposes different kinds of priors such as normal inverted Wishard, a combination of graph theory with the Minnesota and normal Wishard prior, graph theory and hyper-inverse Wishard and a nonparametric LASSO prior. DINAR provides the adjacency matrix estimator with sparse-group sparsity. Song and Bickel (2011) use a lasso type sparsity as a pre-selector for unregularized time series modelling. In the DINAR framework, we derive the least square estimator under two penalties and introduce a two-step algorithm into the DINAR model to obtain the estimator numerically. Simon et al. (2013) derive a related algorithm for univariate regression models, whereas we develop the algorithm for high-dimensional VAR type models. The interconnections strongly challenge the algorithm compared to the univariate case, making it computationally intensive. Lastly, the DINAR model contributes to the literature on econometric modelling for BTC networks. Makarov and Schoar (2021) investigate the geographic distribution of BTC miners and how the concentration of mining capacity varies with BTC price. In this study we analyse the time dependent impact of regional transaction activity on the same continent and other regions on the BTC blockchain. We investigate which regions

influence the others, how this relates to the state of the market and which conclusions can be drawn from the changes in influential regions. For the study we consider the real data of the BTC transactions from 25 February 2012 to 31 December 2021. The results demonstrate the spatial connections and dynamic changes in the BTC blockchain. In particular, it is shown that:

- In the early years (2012 to 2016) influential network effects originated from Africa and South America. This is unexpected because neither hosts a major financial centre nor is strongly covered by the media on cryptocurrency.
- In 2017 influential effects started to arise from Asia and Europe
- Even though a vast number of transactions originate from North America, before 2018 the continent did not show influential network effects though was influential in 2018, 2019 and 2020. At the same time financial institutions, frequently based in North America, started offering products comprising BTC which might explain the stronger role of this continent and its impact on the transaction behaviour on the BTC blockchain.
- Europe and North America are not strongly influenced by detected influential regions, if at all. Instead they tend to influence each other or themselves.
- Taking into account that most Bitcoin mining farms are in Asia, it is surprising to some extent that Asia is not the sole driver but operates Bitcoin for Europe, North America, Africa, Oceania and South America, fostering the importance of these regions in the blockchain.

This paper is organized as follows. Section 2 describes the BTC transaction data. Section 3 presents the method to Detect Influencers in Network AutoRegressive (DINAR) models via sparse-group regularization. In Section 4 we illustrate its advantage over various other estimators in a large simulation study. Section 5 applies the DINAR model as well as 3 competing models to real Bitcoin transaction data and presents an interpretation and discussion. Section 6 presents some conclusions. The data and codes to carry out the numerical calculations are available on the corresponding author's GitHub account.

2 Data description

We consider the BTC blockchain from 25 February 2012 to 31 December 2021 (3597 days with 3489 observed days). The raw data are published on the blockchain at 10-minute frequency[2] with attributes of transaction record size (for the sake of brevity we will refer to it as 'transaction size'), account ID, accounts participating in the

[2]Note that on the Bitcoin blockchain, the transactions are not published at the moment they occur. The miners collect the records and publish them as a block at a 10-minute frequency.

transactions, the timestamp of the transaction, source Blockchain.info and Google Cloud. Blockchain.info provides additional information on the IP address from the relaying party of the origin of the transaction which is used to label the region. We acquired this information from 25 February 2012 until 17 July 2017, after which the IP address are no longer published by blockchain.info. To extend the data further, whenever an account, identified via the account ID, transacts again, we assign the known IP address to it. Users of Bitcoin are encouraged not to reuse accounts and to create a new one for each transaction. This practice challenges the mapping procedure, however still a reasonable number of transactions and transacted amount of the total can be identified, see Table 1. The percentages are derived by comparing the average transaction size of each of the 6 continents in the time period 2014 to 2017 against the identified transaction size of the respective continent in the outlined time periods. We chose this comparison because the overall transaction size on the BTC blockchain did not change much after 2017, see Figure 1. For a while the transaction size even shrank. Hence a comparison with the data from 2014 to 2017 is reasonable. We observe that for the remainder of 2017 the mapping rate is very high and decreases for various continents in 2018. In the real data analysis in section 5, we will exclude continents from the analysis for certain years when their mapping rate is too low. For the decision on which continent to exclude, we also take into account the transaction size, illustrated in Figure 2.

Figure 1: Daily log aggregated transaction size on the BTC blockchain from 2009 until the end of 2021.

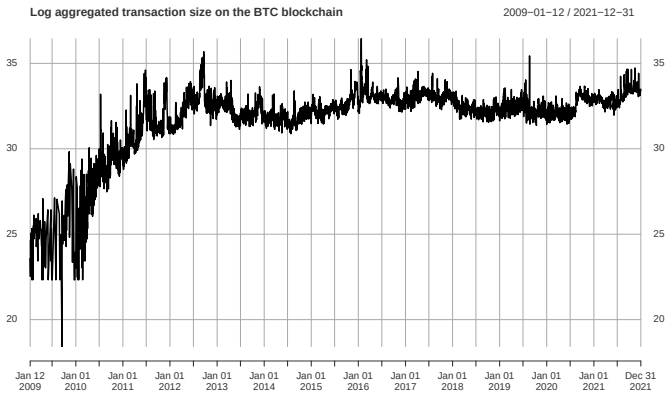


Table 1: Percentage of matched transaction volume over 3 time periods relative to the identified transaction volume over the time period 2014 - 2017.

	AF	AS	EU	NA	OC	SA
2017-07 - 2017-12	33.00	36.02	95.17	35.17	13.90	22.58
2018-01 - 2018-12	15.74	2.62	53.24	5.82	1.39	4.19
2019-01 - 2021-12	3.33	1.46	28.14	1.54	0.96	3.70

We group the data into 6 continents: Africa (AF), Asia (AS), Europe (EU), North America (NA), Oceania (OC) and South America (SA). The continent is identified

depending on the IP address compared with a dataset of IP address from MaxMind Inc. We follow Reid and Harrigan (2013) in tracking the approximate location of the origin of the transaction [3]. The node that informs first about a transaction is close to the location where the transaction takes place, thus one can approximately identify the location where the transaction originates. This approach only works as long as the running node does not use an anonymizing technology. Lischke and Fabian (2016) document that about 1.6% of all BTC nodes use anonymizing technology such as Tor networks.

Each continental group is further categorized according to the transaction size. Due to anonymity, characterizing BTC users is not easy. We thus group the users according to the size of the daily transactions associated with the accounts. The heuristic behind this is that only Bitcoiners with large holdings are able to execute large transactions, while small Bitcoiners contribute to small transactions. As surveyed by Trimborn and Yu (2022), heuristics are commonly applied to construct blockchain networks which can be analysed with analytic methods. Inside each continental grouping, the data are separated into 10 size groups, depending on the deciles of the sizes of the transactions. The first group, indicated by a 1 placed after the abbreviation for the continent, has the smallest transactions, corresponding to the 0%–10% percentile, while the tenth group, with the largest transactions, is indicated by a 10, and corresponds to the 91%–100% percentile. Later, for robustness analysis, we also consider a 3 group per continent setting, where users are clustered into three size groups corresponding to 0%–30%, 31%–70%, and 71%–100% percentiles for small, medium, and big investors, respectively.

The identification of the originating continent and building of the groups we conducted based on the raw data with 10-min frequency. Except for Europe and North America, there are 1% and 25% zeros, meaning no transactions. A lack of liquidity can be challenging for the model estimation. We overcome the liquidity problem by aggregating the raw data to a daily frequency.

For the further analysis, we consider the log transactions. To avoid $-\infty$ in the data for cases without any transactions in a continental grouping within a day, we add 1 Satoshi [4] to each transaction. Given the large numbers under consideration, the bias effect of the correction is negligible.

Figure 2 displays the evolution of the daily log aggregated transaction sizes over all groups in each continent. The left plot shows the data over the time period February 2012 to July 2017, the right one from July 2017 until December 2021. Since the data from July 2017 are based on the matching procedure, we display them in a separate plot. We first analyse and comment on the data until July 2017 since the data from then on are based on the matching procedure which induces a potential

[3]The location of the relaying node gets observed, which is geographically close to the origin of the transaction. Since the information is saved in the blockchain and each user has a copy of it, no information on receiving node gets recorded. Consequently the final destination is not traceable.

[4]The BTC transactions are reported in Satoshi values, the smallest fraction of a BTC, where 1 BTC $= 100,000,000$ Satoshi.

bias. Europe and North America on average have the largest transactions and the dynamic pattern is quite steady. Asia and Oceania contain a few days (8 and 19) without transactions, even after aggregating to daily frequency. They are also more volatile than Europe and North America. Africa and South America are the most volatile and have a relatively larger number of days, 364 and 241, respectively, without transactions. The interpretation of Figure 2 is further supported by the descriptive statistics, presented in Table 2. Inferring from the mean and standard deviation, Europe and North America, Asia and Oceania, Africa and South America indeed show a related behaviour. Also the minimum values indicate the existence of zero transactions, a lack of liquidity, in some areas, supporting the previous analysis. For the matched data from July 2017, one can observe that all time series are decreasing which is attributed to a lower matching rate the further one goes into the future. We can also observe that the matched transactions in South America and Oceania quickly go towards zero. Hence users in these continents observe the suggested rule by Satoshi Nakamoto, not to reuse accounts, better than users in other continents. In North America, Europe and Asia we are still able to match reasonable high amounts of transactions to IP addresses. The aggregated transaction value is also shrinking though for Europe and North America it remains on a steady level from 2019 onwards whereas for Asia it continues to decrease.

Figure 2: Time series of daily log aggregated transactions. The time period is 25 February 2012 until 17 July 2017 (left plot) and 18 July 2017 until 31 December 2021 (right plot) in the 6 continents Africa, Asia, Europe, North America, Oceania, South America.

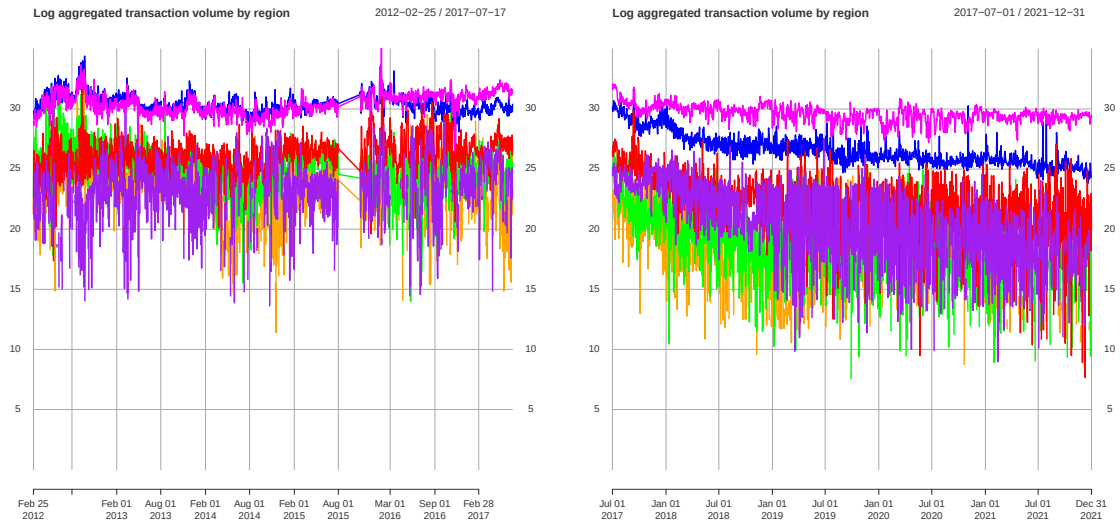


Table 3 provides the average daily transaction volumes in USD in each decile of each continent. To avoid potential bias, the table is derived from the data from February 2012 to July 2017. For the conversion to USD we consider the daily closing price of BTC reported on YahooFinance. One sees that the daily transaction volumes can be very low, especially in Africa and South America, see the lowest decile. At

the same time, the transactions in the top decile of Europe have a mean transaction volume of over 144 million USD per day. In Africa and South America for the same decile, it still ranges to over 200,000 and 425,000 USD on daily average. Apparently there are quite high transaction volumes, especially when considering that BTC is still an emerging asset.

For deeper insights into the features of the data of the groups in each continent, the empirical distribution of the log of the sizes of the transactions is displayed as a densityplot in Figure 3, again for the data from February 2012 to July 2017. For each continent, the left plot corresponds to the first group, namely group 1 with the smallest transactions, and the right one to group 10 with the largest transactions, leading to an increasing pattern within each continent. The narrow box width of Europe and North America suggests a smooth evolution of the transaction sizes with few spikes. There are hardly any occurrences of zero transactions, indicating a healthy liquidity in these regions. This indicates a more mature market in Europe and North America, hence a clearer structure within an estimated model is to be expected. Asia and Oceania are relatively more dispersely distributed. The daily transaction sizes are more volatile, inferred from the size of the center box and the length of the whiskers. South America becomes again extreme in the sense of showing longer whiskers, translating to a larger variation of the sizes of the transactions within each group. Even in group 10 with the highest transaction sizes, there are days without any transactions. Africa follows a very different pattern from the other continents. The respective densityplots indicate high volatilities with frequent drops to zero transaction volume. The divergences between the groups eventually suggests, for the modelling, an adjacency matrix with a flexible choice of parameters.

Table 2: Descriptive statistics of the log accumulated transactions of the 6 regions Africa (AF), Asia (AS), Europe (EU), North America (NA), Oceania (OC), South America (SA).

	AF	AS	EU	NA	OC	SA
mean	18.65	26.01	30.43	30.41	24.76	20.48
sd	9.47	2.19	0.95	0.85	3.08	8.15
skewness	-1.34	-7.17	-1.03	-0.74	-5.34	-1.96
kurtosis	3.06	85.75	11.73	15.79	43.06	5.24
min	0.00	0.00	22.04	21.78	0.00	0.00
max	28.91	31.65	35.38	34.40	31.21	31.85

There remains the question if there is any dynamic dependence between the BTC transactions. As reflected by the lag 1 cross-correlations, Figure 4, shows that there is an autocorrelation effect for the groups of each continent. The effect appears strongest in Europe and North America. Oceania and Asia, on the other hand, have weaker serial dependences, as can be seen from the more shallow red. Frequently it appears that the effect in the highest group 10, is stronger than in their corresponding lowest group 1. The remaining continents, Africa and South America, share similar serial dependence between the groups, which is also stronger than the one in

Figure 3: Daily log transactions of the 10 groups displayed as density plots, with the left density plot representing group 1 and the right one group 10 of the respective continent. The time period is 25 February 2012 until 17 July 2017 in the 6 continents Africa, Asia, Europe, North America, Oceania, South America. The first 8 density plots for Africa range to 0 due to the little number of transactions in this continent in several days.

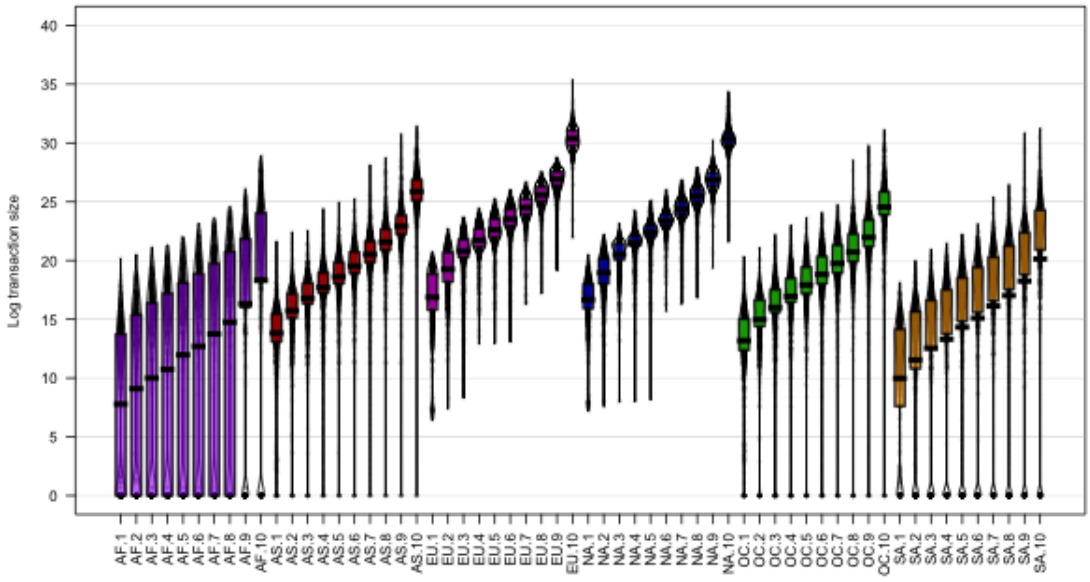
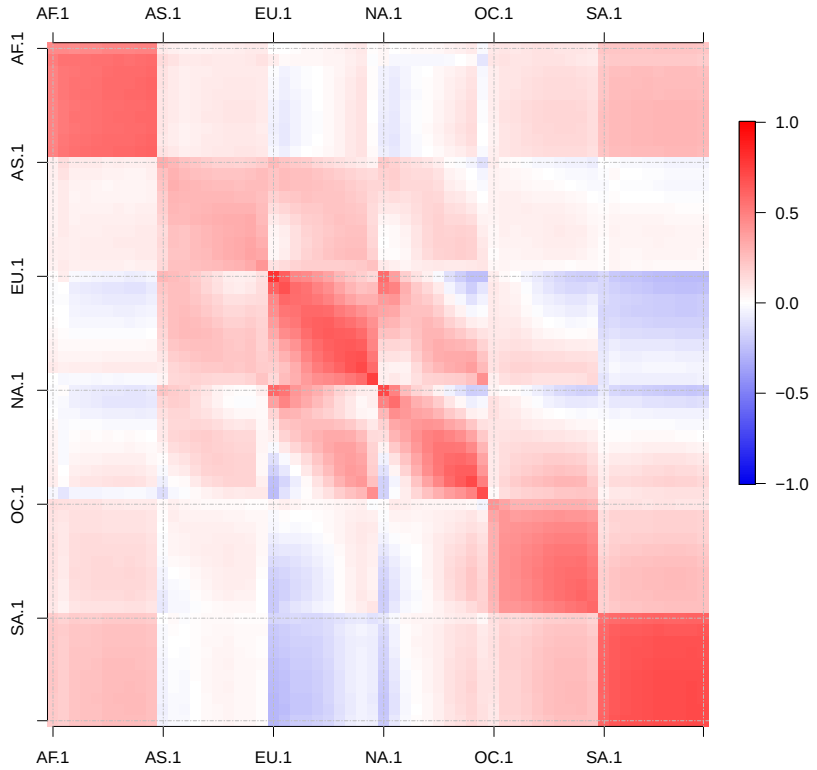


Figure 4: Lag 1 cross-correlations between the size of the transactions – ordered in 10 groups – in the 6 regions: Africa (AF), Asia (AS), Europe (EU), North America (NA), Oceania (OC), and South America (SA). Each block on the diagonal represents the lag 1 dependence within a continent, while the off-diagonal blocks represent the inter-continental effects.



Asia and Oceania. Moreover, there are network effects within the BTC blockchain, again reflected by the lag 1 cross-correlations between the groups and the regions, see Figure 4. The diagonal block of the heat map shows the lead–lag dependence among the groups within the same continent, while the off-diagonal shows the intra-continental cross-dependence. Europe and North America exhibit a stronger cross dependence, both inter-continent and intra-continent, in terms of their influence on the others (lead) and being affected by the others (lag). The network effect is much less between the other continents. South America and Africa exhibit a connection within themselves and simultaneously sparse cross-dependence with the other continents. This indicates that these continents are self-dependent. We observe frequently a stronger intra-continental relationship between the groups than between groups of different continents. This suggests that users may send and therefore trigger more often transactions within their own geographical location. This makes sense from the standpoint that payment systems are more frequently used for services in the geographical vicinities of the end-users.

The magnitude of the cross dependence differs, suggesting flexibility in the dy-

Table 3: Mean daily transaction value (in USD) in the deciles of the transactions of the 6 regions Africa (AF), Asia (AS), Europe (EU), North America (NA), Oceania (OC), South America (SA).

	AF	AS	EU	NA	OC	SA
.1	7.18	52.52	920.34	333.46	11.71	4.06
.2	34.53	205.43	5931.73	2470.26	56.53	19.53
.3	89.94	448.50	18743.82	7341.04	147.04	57.45
.4	197.07	1039.13	39903.12	15438.44	342.60	131.56
.5	431.86	2074.87	83607.85	34735.14	771.13	302.91
.6	947.88	4386.15	177291.31	80343.77	1683.52	713.64
.7	2179.41	17316.84	390226.22	193029.66	3853.24	1970.66
.8	5834.54	66690.19	1011731.37	531628.26	10173.54	4955.48
.9	20647.78	235919.06	3677787.40	2040912.93	40651.43	17731.31
.10	425972.68	2789979.66	144060061.43	71684792.44	573119.69	211864.35

Table 4: AIC, HQ and BIC evaluation criteria for a VAR model selection procedure derived on the data from 2012 to 2021.

	Selected Lag	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5	Lag 6	Lag 7
AIC	2.00	-118.91	-119.18	-119.00	-118.59	-118.22	-117.84	-117.39
HQ	1.00	-116.60	-114.60	-112.14	-109.46	-106.82	-104.17	-101.45
BIC	1.00	-112.44	-106.34	-99.80	-93.03	-86.29	-79.55	-72.74

namic modeling parameters. In the heat map, there are a number of zeros, displayed as blank fields, and values close to zero, which implies sparsity in the dynamic structure.

It remains to investigate the best possible lag structure for the network. We fit VAR models to the data to identify the best lag structure to use during our modeling procedure. The results are presented in Table 4. BIC and HQ criterion favour a model of order 1 whereas AIC suggest a model of order 2. Since our interest is rather in identifying the true model, we place higher relevance upon the BIC. Further it can be observed that the AIC is only slightly better for lag 2 than for lag 1, hence the improvement is marginal. In the Appendix we also provide the lag selection of AIC, BIC and HQ for a VAR modeling procedure on a yearly basis, see Table 19. Again BIC and HQ suggest the lag order 1 and AIC suggests often a model with further lags. Since we place stronger emphasis upon the identification of the true model, we follow the suggestion of the BIC and HQ. Consequently we will work with a lag structure of order 1 in this study.

3 Detecting Influencers in Network AutoRegression

We propose a method to Detect Influencers in Network AutoRegressive (DINAR) models via sparse-group regularization to describe the dynamic dependence in a network with an unknown and sparse adjacency matrix. The adjacency matrix reflects both the connectivity with non-zero values and their strengths, shown as the magnitudes among the nodes. The serial dependence on its own lagged value is not regularized. To detect the dynamic dependence, a two-layer sparsity is imposed on both group and individual effects. We develop a regularized least squares estimator and a gradient descent algorithm for modelling the high dimensional network.

3.1 Specification

Let N denote the size of the network and $Y_{i,t}$ denote the transaction size of Node i, $1 \leq i \leq N$ at time t, $1 \leq t \leq T$, where T is the length of the time period. The DINAR model is defined as follows

$$Y_{i,t} = \beta_0 + \beta_1 Y_{i,(t-1)} + \sum_{j=1}^{N} a_{ij} Y_{j,(t-1)} + \varepsilon_{i,t} \tag{1}$$

where the parameter β_1 controls the autoregressive dependence. The adjacency matrix $A = (a_{ij})_{1 \leq i,\ j \leq N}$ represents the connectivity. The elements of A reflect both the connectivity between Node i and the lagged value of Node j, if nonzero, but also the strength of the dynamic influence of Node j's lag on Node i. The adjacency matrix is assumed to be sparse, with few non-zero entries, highlighting active groups and nodes. If $a_{ij} \neq 0$, Node j is active and has influence on Node i. For $a_{ij} = 0$, Node j has no influence on Node i. If $a_{ij} = 0$ for all i, then Node j is inactive. It is unknown which elements are zeros and which are not. Since the autoregressive dependence is parametrized by β_1, the diagonal elements of A are forced to be zeros (i.e. $a_{ii} = 0$, $1 \leq i \leq N$). In addition, $\varepsilon_{i,t}$ is white noise s.t. $\mathrm{E}(\varepsilon_{i,t}) = 0$, $\mathrm{E}(\varepsilon_{i,s}\varepsilon_{i,\tau}) = 0$, $\mathrm{Var}(\varepsilon_{i,t}) = \sigma_i^2$, $1 \leq i \leq N$ and $1 \leq t,\ s,\ \tau \leq T$.

Define $\mathbf{Y}_t = (Y_{1t}, \cdots, Y_{Nt})^\top \in \mathbb{R}^N$, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \cdots, \varepsilon_{Nt})^\top$ and write I_N for the N dimensional identity matrix. The DINAR model (1) can be represented in compact matrix form:

$$\mathbf{Y}_t = \mathbf{1}\beta_0 + (I_N\beta_1 + A)\mathbf{Y}_{t-1} + \boldsymbol{\varepsilon}_t. \tag{2}$$

Our interest is to detect 1) the active groups and 2) the active elements within the active groups, namely to estimate the adjacency matrix A under sparsity. The large size of the network challenges the estimation of the $N \times N$ adjacency matrix, due to the limited data availability with T as the number of observations.

3.2 Regularization and Inference

Under the two-layer sparsity assumption, also referred to as sparse-group, the estimation is achieved by carrying out a nonlinear regularized optimization:

$$\min_{\theta} \frac{1}{2N} \sum_{t=2}^{T} ||\mathbf{Y}_t - \mathbf{1}\beta_0 - (I_N\beta_1 + A)\mathbf{Y}_{t-1}||_F^2 \tag{3}$$
$$+ \sum_{i=1}^{N} (1-\alpha)\lambda ||A_{\cdot i}||_F + \sum_{i=1}^{N}\sum_{j\neq i}^{N} \alpha\lambda |a_{ij}|$$

where $\theta = (\beta_0, \beta_1, A)^\top$. Note that the estimator (3) is a multivariate version of the univariate estimator introduced in Simon et al. (2013). Group sparsity is applied to the columns of the adjacency matrix. The matrix A is partitioned to $A_{\cdot i}$ with all the elements being 0 except for the ith column, i.e. $A_{\cdot i} = \{A | a_{kj} = 0 \, \forall \, k \wedge (j \neq i)\}$, and $A_{\cdot -i} = \{A | a_{kj} = 0 \, \forall \, k \wedge (j = i)\}$ with the ith column being 0. Individual sparsity is further applied only to the nonzero columns, namely, the active groups. If a group is inactive, the entire corresponding columns of the adjacency matrix will be shrunk to zero. Here α and λ are the tuning parameters and $||A||_F = \sqrt{\sum_{i,j} a_{ij}^2}$ refers to the Frobenius norm. The term $(1-\alpha)\lambda$ controls the group sparsity and $\alpha\lambda$ the individual sparsity. The group-regularization term applies the Frobenius norm because for a vector it is equal to the 2-norm which is non-differentiable at 0. Hence it allows for an estimation of some groups as 0. Similarly the parameters within groups which were not estimated as 0 by the first regularization are estimated with L^1 norm to allow for parameter selection.

We study the small sample behaviour of the estimator in section 4. We also study the asymptotic properties of the estimator under fixed dimensionality N. For the theorems and proofs, please refer to the Appendix, Section I.4 and I.5.

3.3 Implementation

To solve the optimization problem, we develop a gradient descent algorithm and iteratively apply it to each column of A. It is an active-set algorithm, which optimizes the parameters within a group while holding the other parameters fixed. After the algorithm iterated over all groups, it starts again at the first group and reoptimizes the parameters. This procedure is repeated until the parameters no longer change but by a diminutive value. For details on the algorithm, please refer to the Appendix, Section I.

The estimation relies on the choice of α and λ as defined in the regularized optimization (3). The mixing parameter α controls if relatively stronger regularization will be applied to the groups or individual parameters. It can be chosen data-driven though we define it as $\alpha = 1/N$ which ensures that relatively the same regularization

is applied to each group as to any single parameter. If α would be chosen as larger than $1/N$, a relatively stronger regularization would be applied to the individual parameters than to the groups and vice versa for α lower than $1/N$.

The choice of λ on the other hand is data-driven. We design a sequence of λ's starting from the one which sets all parameters in the model equal to 0. Then, a halving sequence is created until the value 0 is reached. A popular approach to find the best λ out of the created sequence is cross-validation, though it ignores the serial dependence in time series, see Nicholson et al. (2017). Hence, we use a forward-looking criterion by selecting λ such that the out-of-sample AICs and BICs are minimized. This approach was also used in Bańbura et al. (2010), Song and Bickel (2011) and Nicholson et al. (2017).

4 Simulation

4.1 Simulation settings

In this section, we investigate the finite sample performance of the proposed DINAR model. In particular, we evaluate parameter estimation accuracy, transaction prediction errors, and sparsity detection of the adjacency matrices along with various simulated networks. We consider five scenarios to demonstrate how the finite sample performance would be influenced by the network size N, the dynamic evolution period T and a growth in the number of groups. We design a small dynamic network with $N = 10$ groups, a medium one with $N = 20$, and a large one with $N = 60$. Each type of networks further involve three different length of time periods, namely short with $T = 100$ transactions, medium with $T = 500$, and long with $T = 1000$. The parameter vector β_0 will not be considered here because in the real data application we demean the data before the estimation procedure. This makes the inclusion of β_0 in the real data application not necessary and to make the settings comparable, we omit it during the data generation process and estimation in the simulation study. Each simulated data is constructed as

$$\mathbf{Y}_t = \{\beta_1 I_N + A\}\mathbf{Y}_{t-1} + \boldsymbol{\epsilon}_t \tag{4}$$

where $\boldsymbol{\epsilon}_t \sim N(0, I_N)$. Attributes $(Y_{1t}, \cdots, Y_{Nt})$, $1 < t \leq T$ are generated with serial dependence β_1 on its own past values and cross dependence A on others' past values.

We consider adjacency matrices of 3 different strength of persistence. We control the persistence by checking the maximal eigenvalue of the companion matrix since the roots of the polynomial of a process correspond to the eigenvalues of the companion matrix. The maximal eigenvalue of the companion matrix of the respective adjacency matrices are 0.75 (low persistence), 0.84 (medium persistence) and 0.94 (strong persistence). A value of 1 would indicate that the process does not 'forget' any past information and a value larger 1 would result in an explosive process. The

matrices are constructed as follows:

- Low persistence: The autoregressive parameter β_1 is 0.5 and 4 of the groups have parameters different from 0. These group have alternating parameters of magnitude 0.15.
- Medium persistence: Similar to the previous setting, the autoregressive parameter β_1 is 0.5 and 4 of the groups have parameters different from 0. These group have alternating parameters of magnitude 0.20.
- Strong persistence: Similar to the previous setting, the autoregressive parameter β_1 is 0.5 and 4 of the groups have parameters different from 0. These group have alternating parameters of magnitude 0.25.

For the case an increasing number of groups with increasing dimensionality, we construct the matrices as follows, which ensures again low, medium and strong persistence. Note that we lower the magnitude of the serial dependence parameter which allows us to introduce additional groups with high parameter magnitude whereas the persistence remains controlled.

- Low persistence: For $N = 20$ the autoregressive parameter β_1 is 0.35 and 6 of the groups have parameters different from 0. These group have alternating parameters of magnitude 0.15. For $N = 60$ the autoregressive parameter β_1 is 0.2 and 10 of the groups have parameters different from 0. These group repeat the parameter vector $0.15, 0, 0$, hence one more 0 is introduced compared to the $N = 20$ setting.
- Medium persistence: For $N = 20$ the autoregressive parameter β_1 is 0.35 and 6 of the groups have parameters different from 0. These group have alternating parameters of magnitude 0.2. For $N = 60$ the autoregressive parameter β_1 is 0.15 and 10 of the groups have parameters different from 0. These group repeat the parameter vector $0.2, 0, 0$, hence one more 0 is introduced compared to the $N = 20$ setting.
- Strong persistence: For $N = 20$ the autoregressive parameter β_1 is 0.3 and 6 of the groups have parameters different from 0. These group have alternating parameters of magnitude 0.25. For $N = 60$ the autoregressive parameter β_1 is 0.05 and 10 of the groups have parameters different from 0. These group repeat the parameter vector $0.25, 0, 0$, hence one more 0 is introduced compared to the $N = 20$ setting.

We compare the DINAR model against 3 alternative methods, namely LASSO (Tibshirani, 1996), SCAD (Fan and Li, 2001) and BGR (Bańbura et al., 2010). The former two methods are designed for parameter selection whereas they do not have a layer for network identification as DINAR does. The latter is has a bayesian

VAR specification which shrinks parameters towards zero. It serves as a comparison between non-sparse and sparse methods. Similarly to DINAR, the λ sequence for LASSO and SCAD is chosen so that the largest λ sets all parameters in the models equal to 0 and from then on a halving sequence is created until the λ approaches 0. The tapering-off parameter in SCAD is set to 3.7, as recommended by Fan and Li (2001). The models are estimated with a coordinate-wise descent algorithm. The BGR model is estimated with the BigVAR package, Nicholson et al. (2022).

Each of the experiments is repeated $V = 100$ times. In each replication $1 \leq v \leq V$, we obtain the penalized least square estimates $(\hat{\beta}_1^{(v)}, \hat{A}^{(v)})$ under both group and individual sparsity. The following measurements are adopted to evaluate the performance in terms of pattern identification and accuracy:

- To evaluate the pattern identification, we computed the False Negative (FN) and False Positive (FP) rates on the estimated sets. FN refers to active set's being falsely identified as null, namely, under-detection or overly sparse. FP refers to the set's being wrongly identified as active, namely, overdetection or overly dense. It is natural that the lower these two measures are, the better the performance. Given the two-layer sparsity, there are then 4 metrics: FN.g and FP.g for groups, and FN.e and FP.e for individual elements. In the case of perfect detection, namely, all 4 metrics are zero, we conclude that the true pattern was identified.

- Accuracy is measured using the Root Mean Squared Error (RMSE). There are two different metrics. 'RMSE para' refers to the estimation accuracy, computed based on the difference between the true and estimated parameters. 'RMSE out' refers to the forecast accuracy, which is calculated based on the residuals between the observed values of the time series and the predicted values based on the model. It is an out-of-sample measure based on the testing dataset. In both of the accuracy metrics, a low value indicates good accuracy.

The Figures 5, 6, 7 and 8 summarize the performance of DINAR and competing estimators along with the 6 measurements in the 54 experiments, separated according to model specification. The Figures 5 and 6 show the setting where the number of groups remain steady with increasing dimensionality, the Figures 7 and 8 the ones for increasing number of groups with increasing dimensionality. We evaluate the simulations with AIC and BIC, Figures 5, 7 show the results for BIC and Figures 6, 8 for AIC. The roseplots show the performance of the simulation study for the 3 dimensions, $\{10, 20, 60\}$, for the number of observations, $\{100, 500, 1000\}$, and the 3 strength of persistence, low, medium (med) and strong (str). Each roseplot is separated into 3 sections, one for each covered dimensionality. Within each section, 3 subsections are assigned for the number of observations: $T = \{100, 500, 1000\}$. For each of these subsections, 3 further subsections are provided for the strength of persistence. Within these subsections, 4 columns of rectangles are provided, named A, B, C, D. The naming convention refers to DINAR (A), SCAD (B), LASSO (C) and BGR (D).

The entire circle has 6 tracks, each of which represents another evaluation criterion. The most outer track is referred to as 1, and the most inner track is referred to as 6. We leave a gap for improved representation between track 5 and 6. The 4 most outer tracks are for the False Negative and False Positive criteria: FN.g (1), FP.g (2), FN.e (3), FP.e (4). The FN and FP rates vary between 0 and 1, whereas the color palette goes from white (0) to red (1). No False Negatives and no False Positives are the best possible outcomes; hence, the more white or shallow red the rectangles are, the better. Track 5 reports the 'RMSE para' with a color palette from white (0) to blue (maximum value of 'RMSE para'). Again, the lower the value, the better; hence, white or shallow blue rectangles are preferable. The 'RMSE out' is reported via track 6. The color palette goes from white (0) to green (maximum value of 'RMSE out'). Since these evaluation criteria reflect the error terms, the smaller the values are, the better. Thus, white and shallow green is preferable.

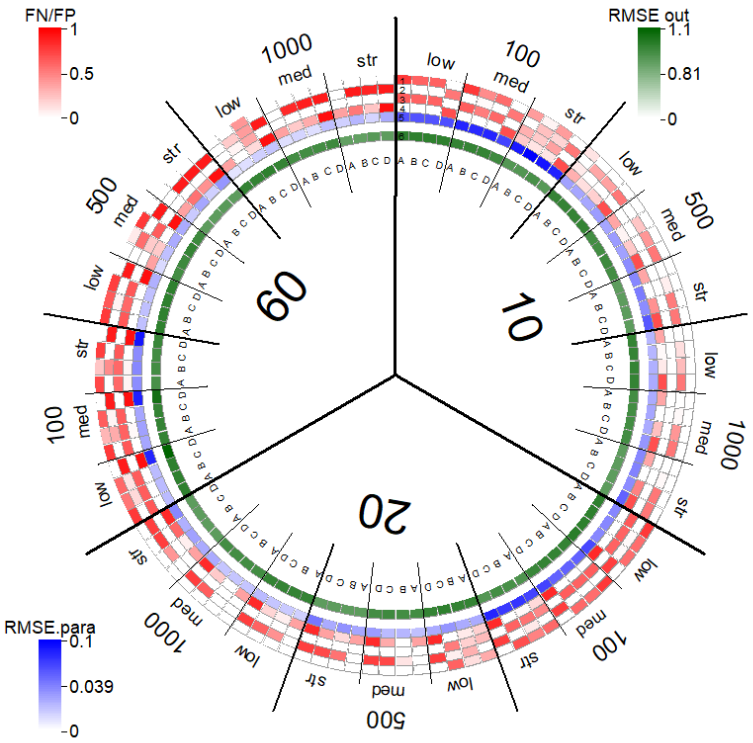
4.2 Simulation evaluation

We select the models in the simulation study on out-of-sample data of same length than during the estimation with AIC and BIC. The Figures 5, 6, 7 and 8 summarize the results and show that for 100 observations and low persistence with 500 observations, the evaluation with AIC gives better results than with BIC. From 500 observations onwards with medium and strong persistence, AIC and BIC commonly perform similar. Though BIC is performing slightly better in terms of FN.g, FP.g, FN.e and FP.e for 1000 observations.

As we observe from the Roseplot 5, 6 and the Tables 9, 10, 11, 14, 15, 16 (in Appendix), DINAR has a good ability to recover the underlying structure. For all 3 number of time series, $N = \{10, 20, 60\}$, and $T = 100$, we observe that none of the 4 models do particularly well. All models favour too sparse parameter matrices, except for BGR which overparametrizes the model strongly. For lower dimensionality the models, DINAR, SCAD, LASSO and BGR have about the same 'RMSE out'. With increasing dimensionality, $N = \{20, 60$, DINAR performs increasingly better. Notably the result with AIC outperforms the competing models stronger than the one evaluated with BIC. So for $T = 100$ we can conclude that all models do equally poor in identifying the true model but DINAR does better in terms of prediction accuracy when the dimensionality increases. For higher number of observations, $T = \{500, 1000\}$, we observe that DINAR has a FP.e comparable to LASSO or just better than it for most scenarios and dimensions. However SCAD is frequently outperforming DINAR in terms of FP.e for $N = 10$, but for higher dimensions, $N = 20$ and $N = 60$ the difference becomes smaller though still in favor of SCAD. However DINAR is performing tremendously better when it comes to the identification of the correct groups. Its FP.g and FN.g are usually around zero or zero exactly for $T = \{500, 1000\}$. But for the $N = 60$, which we consider in the application of this study, DINAR is performing remarkably better than the competing models. This indicates that DINAR is better suited for model identification in higher dimen-

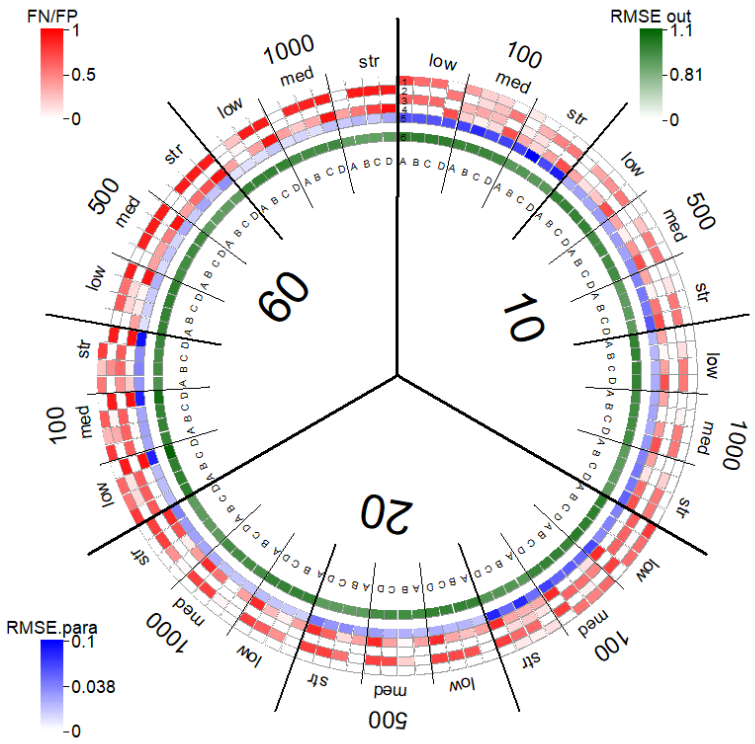
sions than the other models. As for FN.g and FP.g, the measurements to evaluate the identification of the influential groups, DINAR is outperforming all 3 competing models. Since at times DINAR has a worse FP.e than SCAD, this indicates that DINAR identifies the influential groups but does not necessarily estimate parameters which have zero magnitude as being 0. Indeed we observed that various parameters are estimated as extremely close to zero but not zero exactly. This is likely due to DINAR attempting to model the underlying variations as accurately as possible in which it is successful as can be seen from the 'RMSE out' being similar to the other models. However these models consider non-influential groups as being influential. In particular for the highest dimension, $N = 60$, the competing models are far from the true underlying structure. On the contrary the estimation with DINAR results in a model which is much closer to the true underlying model though it tends to estimate parameters within influential groups as non-zero which are zero in the true model. Though the identification of the influential groups is the key target of this study and DINAR does remarkably better in this regards than the competing models. These observations are valid for evaluation with AIC and BIC. In line with the goals of this study, DINAR is performing better in this simulation.

Figure 5: Roseplot for the first set of simulations evaluated with BIC



So far we kept the number of influential groups fixed at 4 for each simulation. In Figure 7, 8 and Tables 12, 13, 17, 18, we show the results for an increasing number of influential groups with increasing dimensionality. The settings of the simulation remain as before, just the parameter magnitudes change to ensure the persistence

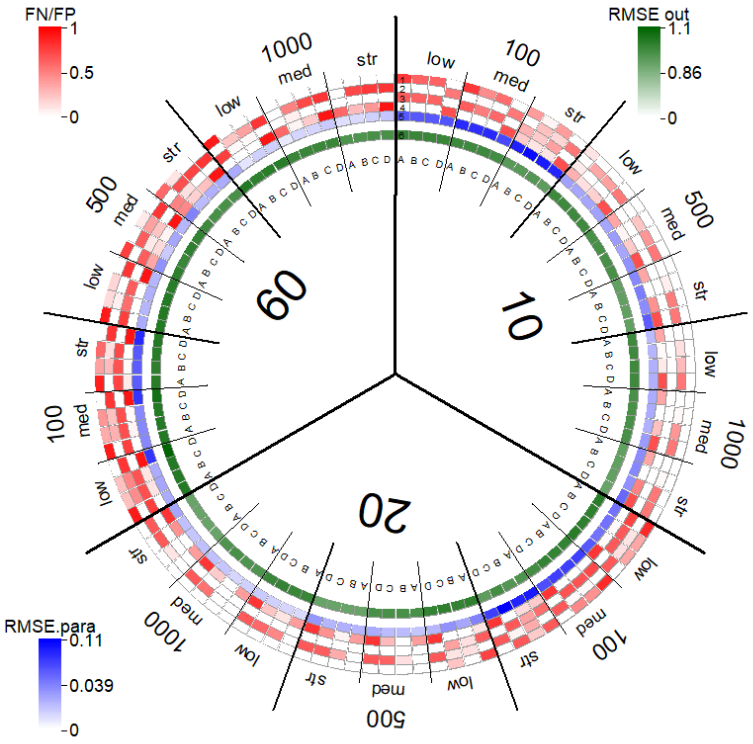
Figure 6: Roseplot for the first set of simulations evaluated with AIC



remains on a similar level as before. The results indicate that again DINAR is doing much better in regards of identifying the influential groups. Similar to before it is outperformed at times by SCAD. When the dimensionality grows to $N = 60$ and for evalution with AIC, then DINAR is doing much better than any of the competing models for $T = \{500, 1000\}$. Given the high dimensionality, it is within expectation that all models perform equally poor when estimating the model with 100 observations. The BIC has a stronger penalty term than the AIC and due to this situation and the larger number of parameters different from 0 than before, frequently the BIC criterion chooses a model for DINAR which performs poorly. However, as illustrated with the AIC criterion, DINAR is able to identify a better model than competing methods. Hence the conclusions from the previous case remain valid.

In terms of accuracy, we observe that the 'RMSE para' are usually better for DINAR. They are consistently lower than for the three alternative methods. This result is in line with the fact that DINAR has better FN and FP values than the alternative methods. Even for the cases when LASSO has a lower FP.e, the 'RMSE para' is similar for LASSO and DINAR which indicates that the magnitude of the wrongly selected parameters for DINAR must be very close to 0 which only marginally harms the parameter identification performance of DINAR. The 'RMSE out' are comparable for all methods which suggests that LASSO, SCAD and BGR over-parameterize the adjacency matrix since they do not provide better prediction accuracy out-of-sample. Thus DINAR recovers the underlying structure better and gives overall a similar

Figure 7: Roseplot for the simulations with increasing number of groups when dimensionality increases, evaluated with BIC



out-of-sample performance.

The key target of this study is the identification of influential groups in high dimensions. As the simulation study showed, DINAR is outperforming all other methods in this regards. Hence, overall the simulation results suggest that DINAR outperforms the alternative methods in terms of identifying influential groups in high dimensions.

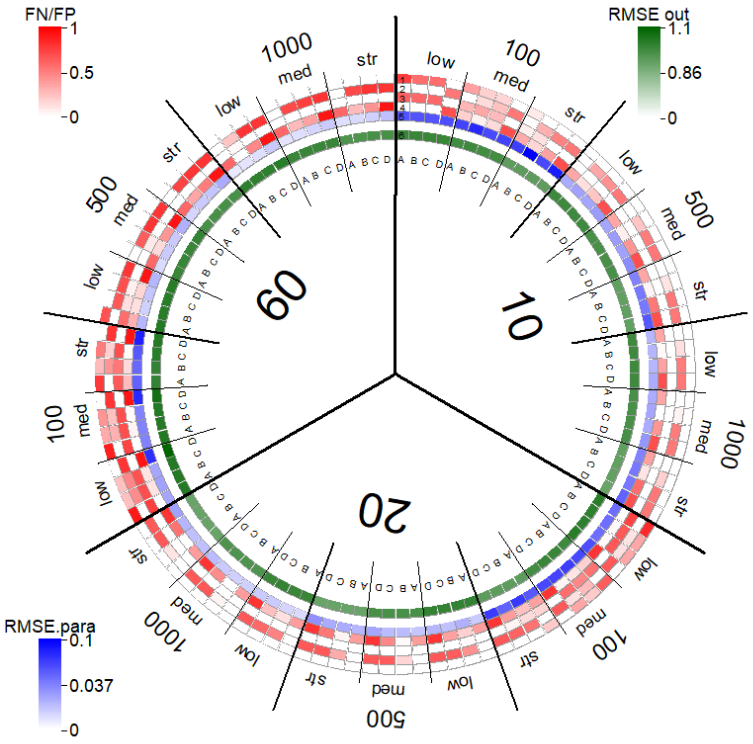
5 Real Data: Bitcoin Transaction Analysis

In this section, we analyse the BTC blockchain and implement the DINAR model to detect the regional and size effects in the global virtual currency transactions in the BTC blockchain network.

5.1 Estimation procedure

We use the BTC transaction data described in Section 2 from February 2012 to December 2021. To provide a better interpretation, the data is demeaned and scaled

Figure 8: Roseplot for the simulations with increasing number of groups when dimensionality increases, evaluated with AIC



with GARCH volatility. To ensure that the dynamics of each volatility process are represented, a GARCH(7,7) model is fitted. The reason for the scaling is to make the parameters comparable and to ensure that the LASSO type model does not favour time series with a larger volatility due to varying scale of the data. Hence the data are scaled to 0 mean and volatility 1. As such the magnitudes of the parameters become comparable. The intercept β_0 is not required in the estimation since the data are demeaned. We are modelling the transactions on a daily basis as follows:

$$Y_{i,t} = \beta_1 Y_{i,(t-1)} + \sum_{j=1}^{N} a_{ij} Y_{j,(t-1)} + \varepsilon_{i,t} \tag{5}$$

where the parameters are defined as in (1). We focus on the estimation of the unknown adjacency matrix $A = \{a_{ij}\}$ for network connectivity and the parameter β_1 of serial dependence. To understand the time related dependence in the network, we split the activity into years and perform the estimation independently for each year. In total, there are 10 samples; each contains the daily transactions of the 60 groups within the particular year.

The estimation relies on the choice of α and λ as defined in the regularized optimization (3). The mixing parameter α is set to be $1/N$, where N is the number

of groups: in this case $N = 60$. The choice of λ on the other hand is data-driven. Although cross-validation is a standard technique, it ignores the serial dependence in time series, see Nicholson et al. (2017). Hence, we use a forward-looking criterion by selecting λ such that the out-of-sample AICs and BICs are minimized on the next year's data, a year-to-year approach. This approach was also used in Bańbura et al. (2010), Song and Bickel (2011) and Nicholson et al. (2017). As an example, for the period of 2015, the DINAR estimation is conducted on the sample period from 1 January 2015 to 31 December 2015. The hyperparameter λ is selected such that the forecasts for the next 90, 180, 270, 365 days from 1 January 2016, computed with the adjacency matrix and β_1 estimated in 2015, has the minimal out-of-sample AIC and BIC among all the alternatives. We chose 90, 180, 270 and 365 days to provide a short and long term analysis which also serves as robustness analysis for the results. Since DINAR is a method dependent on the evaluation period, a consistent choice over all periods is crucial for a meaningful comparison. To select λ, we carried out this year-to-year estimation exercise for each period from 2012 to 2020 until we reached the end of the sample, i.e. 2021. By this procedure the selected model will be evaluated for its forecasting accuracy. For evaluation of the network effects over time, this is a viable choice. A different evaluation scheme might result in a different model choice. Accounting for that we also provide and analyse the networks of different λ-values from the model search. Further Table 5 shows the selected number of parameters and groups of the optimal model chosen by AIC and BIC accounting for the different length of evaluation periods. Due to its stronger penalty for complex models, BIC frequently favours a sparser model. However, most of the time the metrics agree on a similar model choice which is in particular the case when the evaluation period is shorter. The longer the evaluation period more often AIC favours a model with more parameters and consequently more active groups. Since the information criteria agree most of the time on the same model and we are rather interested in a consistent model choice, the main evaluation metric in the application will be the BIC. Additionally the BIC displays higher stability in terms of the optimal model choice regardless for shorter or longer evaluation periods.

5.2 Evaluation

We evaluate the estimation performance using metrics including the out-of-sample AIC and BIC for the next 90, 180, 270, 365 observations of the coming year. These metrics strike a balance between model fit and the complexity of the model, hence are well suited to evaluate sparsity approaches.

We illustrate the serial cross-dependence with chord diagrams, see Figures 9 and 11. They demonstrate the dynamic connectivity in the global BTC blockchain network. A chord diagram displays the direction and magnitude of the influence of each node by showing the magnitude by means of the circle and the destination of the signal by the chord. The wider the space on the circle, the larger the magnitude and hence the higher the dynamic impact in the network. A chord diagram does

not differentiate between positive and negative influences. The sum of the absolute values of the parameters (magnitude) is displayed on the circle. Moreover, the colour of the chord corresponds to the colour of the continent to which the effect is directed. For an example, consider Figure 9e, where EU.1 is outstanding with a magnitude of 12.5 and about one-third of the magnitude directly influencing the other European groups. The remaining magnitude mostly reflects an influence on North American, South American, Oceanian and Asian groups.

5.3 Results and interpretation: 10-groups

DINAR is applied to the grouped transactions (10 groups per continent) and we tackle the problem of answering if user groups defined by transactions are related in a time dependent manner between years. In the year-to-year analysis, zero entries in the adjacency matrix indicate that the past transactions of the corresponding group have no influence on the future transactions of another group. If there are only zero entries in one column, this indicates the lack of network connectivity of the particular regional size group with all the other groups. On the other hand, a group with a non-zero entry in the adjacency matrix is considered as an active group as it is able to influence the dynamic evolution of the virtual money flows.

We focus on the years where the adjacency matrices are not zero, in other words, where network effects appeared. Figure 9 illustrates the active network connectivity based on the 10 groups per continent over the whole sample period evaluated upon 180 days out of sample. For an enlarged version of each plot, see Figures 14, 15, 16, 17 in the Appendix. Comparing with Table 5, we see that the number of groups and parameters chosen for 180 by BIC are relatively stable over the other evaluation periods as well. It shows there are network effects for all years but 2013. 2012 was the year when BTC received increasing attention. Its price doubled before it skyrocketed in 2013 with its price reaching over 1000 USD for the first time in November 2013. For this first year of the analysis only the smallest group in Africa had networks effects and these effects were quite similar in magnitude[5] for any other group as can be seen from the magnitude of the parameters being more or less equal (width of the chords), see Figure 9a. In 2012 BTC was not that popular yet and it is interesting to observe that more of the network connections are with South America, Asia and Oceania, rather than with Europe or North America. Checking the robustness of the result, we compare with the modeling choice of LASSO and SCAD, see Tables 7 and 8. LASSO and SCAD both choose a model only containing the autoregressive effects, as can be seen from the number of parameters matching the dimensionality and the number of groups being 0. Even though the model is sparser, BIC and AIC are both worse for LASSO and SCAD compared to the more complex solution of DINAR, which indicates that for DINAR the goodness-of-fit improved strongly compared to its complexity. With the price spiking in 2013 no network effects were detected in the BTC blockchain,

[5]Recall that the data are standardized, hence the magnitudes of the parameters are comparable in their values.

though again in 2014 effects arose. Africa had effects on itself as well as Asia and South America, whereas Europe had network effects mostly on itself, Oceania and North America. The effects on themselves are expected because as was displayed in Figure 4, there are strong intra-continent correlations, in particular in Africa and South America. Interestingly the detected network effects in 2014 can be split between areas. Most notably the continents which are heavyweights in the international financial system, North America and Europe, do not receive strong network effects from Africa and vice versa Europe is connected to North America and Oceania, whereas the latter includes Australia and New Zealand. Figure 2 showed that akin to the traditional financial system, the plethora of transactions in the BTC blockchain originate from North America and Europe. In combination with the observed network effects from 2014, this leads to the conclusion that even early in its existence, the user behaviour on the BTC blockchain hinted towards similar structures as they exist in the traditional financial system. However it is remarkable that network effects originated from Africa, which plays a less significant role in the global financial markets. In 2015 only network effects originating from South America were detected which impacted mostly itself, Oceania, Asia and Africa, though hardly Europe and North America. It appears that as before the continents which have less transactions originating from them, compare Figure 2, show network connections with each other. However Europe and North America, only impact themselves via their respective autoregressive behaviours.

In 2016 the situation changed strongly. Akin to 2012, the price doubled on the exchanges before it skyrocketed at the end of 2017. The group representing the smallest transactions originating from Europe, is now the only group showing network effects. Notably most of the effects go to other groups in Europe, followed by South America, Oceania and Asia. North America and Africa hardly receive any network effects and remain mostly represented by their own autoregressive effects. Interestingly North American transactions appear to be not impacted by other groups and also not to impact others. This remains surprising since the vast majority of BTC transactions originate from North America. It appears that the transaction network in North America is mostly disentangled from users in other geographies. The results for 2017 include North America, Europe and Asia. Other geographies had to be excluded from the analysis since identifiable transactions dropped too frequently to 0, see Figure 2. This year showed a surprisingly different structure than before. Asia has network effects with itself, North America and Europe, whereas North America is mostly impacting Europe and Asia, itself only marginally. 2017 was also the year of the price spike of BTC, when it reached 20000 USD per Bitcoin for the first time during a highly volatile market environment. It is well documented, see e.g. Griffin and Shams (2020), that this time period was subject to market manipulation. The exchange Bitfinex, which played a central role in the price manipulation, is located in Asia, which might be one contributing reason to this phenomenon. However, the actions of a few actors on the blockchain and markets are unlikely to result in such strong network connectivity, hence it is likely that a structural change took place in 2017.

From 2018 onwards too few transactions originating from Asia could be identified

which made it impossible to continue including the continent into the analysis. In 2018 and 2019, North America showed network effects originating from its largest transactions group impacting itself and Europe. It is worth pointing out that North America does play now a network role because previously its transactions had hardly impact on other continents and groups. Since North America plays a central role in the international financial system and in the past few years more and more financial institutions started offering product containing BTC, this network effects could reflect the increasing activity of the financial institutions in this market. The same observation and interpretation can be made for 2020, whereas now also Europe shows networks effects which mostly relate back to other groups within Europe.

We have often observed strong network effects from a continent impacting groups of its own geography. This observation is expectable since we observed in Figure 4 strong intra-continent correlations. This structure could be due to the rational that users of a payment system usually spend their funds in locations or with companies / shops which are geographically closer to themselves. It is more common for consumers to facilitate local transactions than international ones.

We observed particularly in the earlier days that network effects coming from South America and Africa and later on effects from Europe and North America dominated. The groups and continents involved in the more recent years seem surprising to some extent, because media reports often focus on the roles of CCs in Asia rather than in Europe or North America, especially in terms of mining. But comparing the time series plots, Figure 3, it is obvious the volumes of transactions in Europe and North America are higher than in other regions. Also these are the regions where financial institutions offer products involving BTC. This gives a good rationale for the effects coming from these two regions, even from smaller groups like EU.1. Further support for this finding comes from the surprising number of null values in Africa, South America and Oceania, which (for South America and Oceania) occur more often in 2016 and 2017. Explanations for these values may be that users from these regions switched to other CCs, since in this period a bunch of altcoins (CCs other than BTC) became important. Secondly, the number of transactions in the BTC blockchain increased strongly (150%) in this time, see Figure 10, and simultaneously the maximum block size of 1 Megabyte was reached. Since each block of transactions has a limit on the possible number of included transactions, it is likely that certain users from Europe, from where network effects originated in 2016, see Figure 9e, completely dominated the transaction chain in this period. Via the willingness to pay transaction fees, the miners' decisions about privileging one transaction over another by including it into the next block can be influenced. The miners have an incentive to include small transactions which pay high transaction fees, since by this action they can maximize their personal profit. Hence the respective transactions would be prioritized, which leads to the conclusion that high value transactions originated from the continents detected in the analysis. This provides good evidence for the economic reasons described for our finding. The limit on the possible number of transactions included in each block led the developers to introduce a BTC without this restriction, called Bitcoin Cash (BCH) on 1 August 2017. This event was a fork of the BTC

source code in which the code was amended so that it would fit the features wished for. As a result, BTC and BCH exist as individual CCs.

Finally, it can be inferred that we observe strong changes in the network over the years. In the earlier years the network effects more often originated from Africa, South America and Europe, whereas later on, when financial products including BTC emerged, North America dominated the network effects.

Table 5: Number of parameters and identified groups when using AIC and BIC for out-of-sample performance whereas a group is counted as active if 25% of the parameters are different from 0. The results are displayed for the models found in the respective years for DINAR. The first 4 columns show the results evaluated on 90 days out-of-sample, the second 4 columns the results evaluated on 180 days out-of-sample, third 4 are for 270 days and the last 4 are for an entire year of data.

	90 days				180 days				270 days				one year			
	para		group		para		group		para		group		para		group	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
2012	119	60	1	0	119	119	1	1	119	119	1	1	119	119	1	1
2013	60	60	0	0	237	60	3	0	237	60	3	0	237	237	3	3
2014	178	60	2	0	178	178	2	2	178	178	2	2	178	178	2	2
2015	119	60	1	0	119	119	1	1	119	119	1	1	119	119	1	1
2016	119	60	1	0	119	119	1	1	119	119	1	1	119	119	1	1
2017	88	59	2	1	88	88	2	2	88	88	2	2	88	88	2	2
2018	39	39	1	1	39	39	1	1	39	39	1	1	39	39	1	1
2019	39	39	1	1	39	39	1	1	286	39	14	1	286	39	14	1
2020	58	58	2	2	58	58	2	2	286	58	14	2	286	58	14	2

For the analysis of the out-of-sample fit, we compare the fit of DINAR against alternative models. This comparison is performed together with the robustness check, reported in section 5.6. Considering the regularization parameters, we compare their values in each year for each evaluation period in Table 6. The high λ penalties in the early years indicate that a lot of noise is present in the data and the identified network effects had to be strong to surpass the regularization applied. This gives further evidence for the identified groups and parameters representing the underlying network. We can observe that the regularization parameters shrink strongly over time which indicates that network effects were spurious in the year-to-year analysis and they represent a larger share of the volatility in the data.

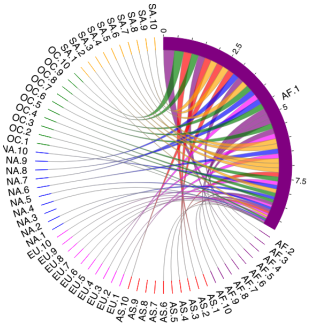
5.4 Robustness check with different λ values

The Figures 12 and 13 provide deeper insights into the network within the years. We display the resulting adjacency matrices for 3 different levels of λ in the respective years. We show the results for λ_{14}, λ_{15} and λ_{16} because for up to λ_{13} frequently no network effects are visible and from λ_{17} the model structure often becomes unclear again. Additionally during the model selection exercise the best model was found to be in this range.

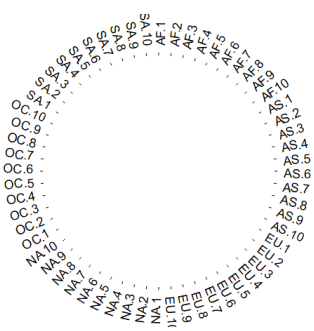
We observe that the groups which are identified as being the originators for net-

Figure 9: Adjacency matrices and serial dependence parameter in analysis with 10 groups and evaluation period length of 180 days.

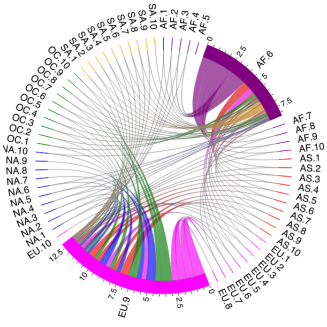
(a) Adjacency matrices for the 10 groups in 2012



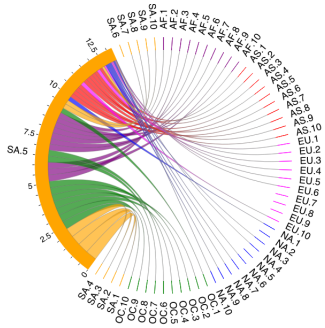
(b) Adjacency matrices for the 10 groups in 2013



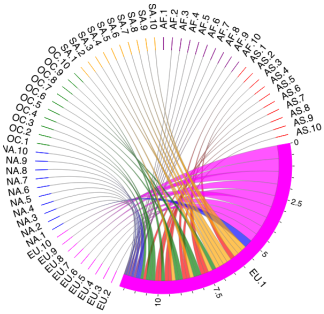
(c) Adjacency matrices for the 10 groups in 2014



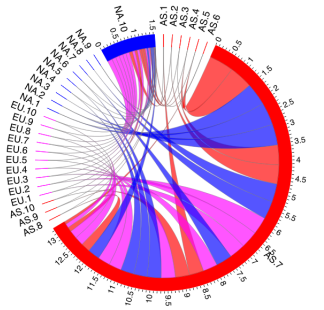
(d) Adjacency matrices for the 10 groups in 2015



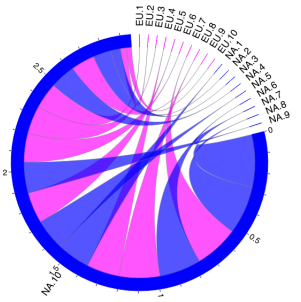
(e) Adjacency matrices for the 10 groups in 2016



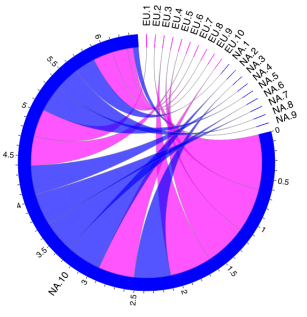
(f) Adjacency matrices for the 10 groups in 2017



(g) Adjacency matrices for the 10 groups in 2018



(h) Adjacency matrices for the 10 groups in 2019



(i) Adjacency matrices for the 10 groups in 2020

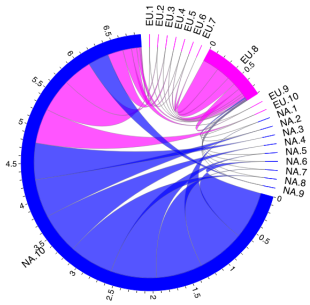
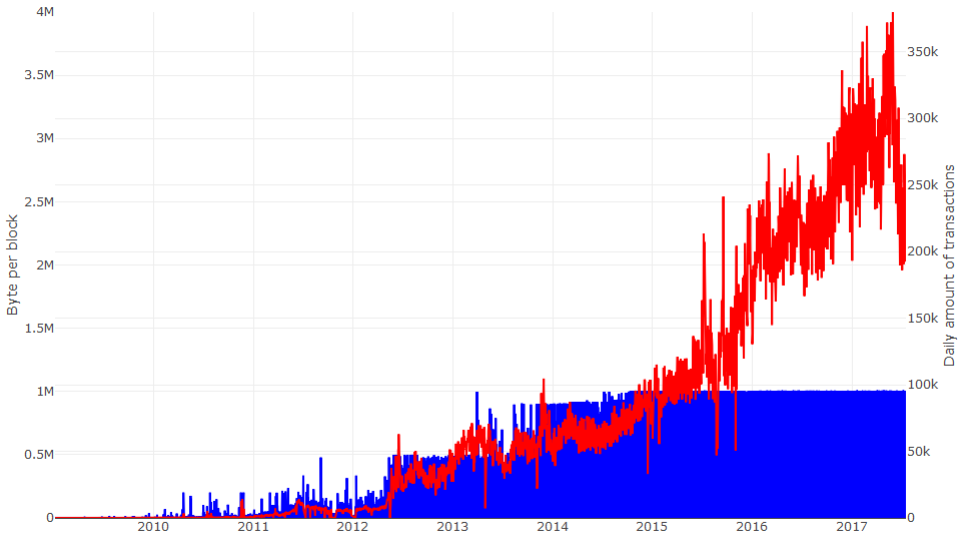


Figure 10: Daily Bitcoin Transactions (line) and the Block Size of Bitcoin (filled area) in the time period 4 January 2009 to 17 July 2017.



work effects remain strong and important even when the λ value decreases and more groups become active in the network. This gives further evidence that the identified groups are important for the network. We also observe that in the earlier years, larger λ values already result in the identification of network effects. However from 2017 onwards, larger λ do not uncover network effects. Taking into account the optimal λ for each year, it is interesting that the next smaller λ often displays network effects coming from North America which are not part of the best model. This is remarkable since North America has various of the largest financial centres. Even for the next λ after the optimal one, the network effects coming from North America are much smaller in magnitude than for other groups and continents. This gives further evidence for the observation that North America does not play an important role in the BTC blockchain. This, however, changes from 2018 onwards, which is the time period during which more and more financial institutions started offering financial products including BTC or being circled around BTC.

It is also an interesting observation that Africa and South America play important roles in the BTC blockchain network, considering both continents have the lowest transaction volume and are not financial centres. Asia on the other hand was expected to show effects because it has an important role in the BTC community since many miners and CC companies are located in Asia. Though only in 2017 network effects are coming from Asia. Also for the λ right after the optimal one, Asia usually does not play an important role if any. Usually the magnitude of the parameters coming from Asian groups is small and in 2014 and 2015, the network with the λ right after the optimal one, does not even include any group from Asia. Hence the network effect of this continent on the BTC blockchain is even less than the one from North America, even though it is rational to expect the opposite.

Table 6: Regularization parameters for individual parameters and groups of parameters when using BIC for the selection of the optimal model based upon out-of-sample performance. The first 2 columns show the regulatization parameters for evaluation on 90 days out-of-sample, the second 2 columns for evaluation on 180 days out-of-sample, third 2 are for 270 days and the last 4 are for an entire year of data.

	90 days		180 days		270 days		one year	
	$\alpha * \lambda$	$(1-\alpha) * \lambda$	$\alpha * \lambda$	$(1-\alpha) * \lambda$	$\alpha * \lambda$	$(1-\alpha) * \lambda$	$\alpha * \lambda$	$(1-\alpha) * \lambda$
2012	0.0369	2.1771	0.0185	1.0886	0.0185	1.0886	0.0185	1.0886
2013	0.0208	1.2263	0.0208	1.2263	0.0208	1.2263	0.0104	0.6131
2014	0.0238	1.4056	0.0119	0.7028	0.0119	0.7028	0.0119	0.7028
2015	0.0359	2.1199	0.0180	1.0599	0.0180	1.0599	0.0180	1.0599
2016	0.0249	1.4683	0.0124	0.7341	0.0124	0.7341	0.0124	0.7341
2017	0.0082	0.2369	0.0041	0.1184	0.0041	0.1184	0.0041	0.1184
2018	0.0132	0.2503	0.0132	0.2503	0.0132	0.2503	0.0132	0.2503
2019	0.0079	0.1501	0.0079	0.1501	0.0079	0.1501	0.0079	0.1501
2020	0.0080	0.1515	0.0080	0.1515	0.0080	0.1515	0.0080	0.1515

5.5 Robustness check with alternative grouping: 3-groups

In order to understand whether the regional and size effects are robust to the grouping, we carried out robustness checks with alternative three groups. For the three groups, users in each continent are further split into small, medium, and large groups, according to the sizes of their transactions. The other model settings remain the same as previous. The resulting adjacency matrices for the three groups per continent are illustrated in Figure 11, which again were obtained by minimizing the out-of-sample BIC on 180 days forecast.

Due to the change in the grouping, naturally the network can change though persistence in the results is a strong sign for the successful identification of underlying structures. As Figure 11 shows, for 2012 again just one group has network effects on all the other continents, similarly to the 10-grouping setting. However this setting attributes the networks effects to South America instead of Africa. Though both settings agree that there are no network effects in 2013. For 2014 the result is robust. Similarly to the 10-group setting, the network effects originate from Europe and Africa, with the identified groups comprising the previously identified ones in the 10-group setting. Hence the result is robust to the grouping.

For 2015 and 2016 the 3-group results in no network effects, whereas both times the 10-group setting uncovered effects from South America and Europe. From 2017 the data matched with the IP address from the previous 5 years are used. Recall that for the 10-group setting we had to exclude some continents because too few transactions could be identified originating from these locations. Due to the less granular setting, we are now able to still include Oceania and Africa, which we had to exclude earlier. Note that similarly to the 10-group setting, Asia shows strong network effects in this period though also Oceania with even stronger effects gets uncover. Oceania could not be identified in the 10-group setting because we had to exclude it due to too few

transactions in various groups. In the 10-group setting we also observed that network effects originating from North America were always present from 2018 onwards. We detect the same result for the 3-group setting, hence we conclude the result is robust in regards to the grouping.

We found most of the result to be robust towards the granularity of the grouping, with the exception of 2012, 2015 and 2016, where the results differed by a different group being selected and, for the latter two years, no network effects present. However, for the other 6 years the results from the 10-group setting were detected as well, hence we conclude that the results are robust in regards to the grouping.

5.6 Robustness check against alternative methods

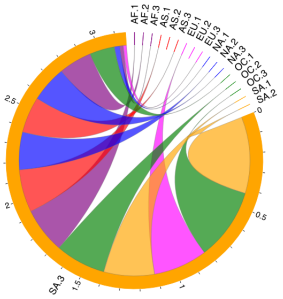
Akin to the simulation study, we compare the results of DINAR against 3 alternative methods. We compare against LASSO, SCAD and BGR whereas we evaluate the estimation performance by AIC and BIC on out-of-sample data of different period length for the next year's observations. Further we compare how many parameters were selected and how many groups were identified as active. We consider a group as active if 25% of the parameters are different from 0. The results are reported in Tables 7 and 8.

We observe that DINAR has a better AIC and BIC than the optimal LASSO solution in each year. Notably LASSO often chooses a model which has only autoregressive parameters, which can be inferred from the observation that the number of parameters equals the dimensionality of the data and the number of identified groups is 0. Only for 2019 and 2020, LASSO chooses models with a number of active parameters, but even then DINAR has a better AIC and BIC, hence provides the better solution despite being sparser. For SCAD the same interpretation holds but for the last few years. For 2018, 2019 and 2020, SCAD chooses a more complex model than DINAR with more active groups but the AIC and BIC outperform. This indicates that the parameter estimation from SCAD must have been more accurate than for DINAR which would result in a better goodness-of-fit. This interpretation is supported by the fact that DINAR relies on a LASSO-type penalization, and we observed that LASSO showed a worse AIC and BIC. By its nature BGR as a non-sparse model estimates all parameters different from 0. The solution of any of the three sparse models outperforms BGR.

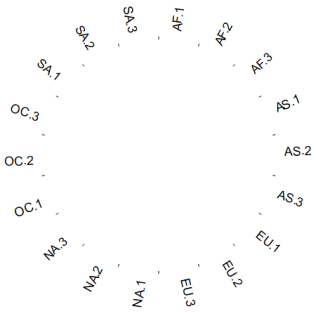
Comparing against the results for the 3-grouping, see Tables 20 and 21, we observe similar results as before. Again DINAR outperforms LASSO but gets outperformed by SCAD twice, namely in 2016 and 2018, whereas for the latter only marginally. Again we attribute this result to the fact that DINAR relies on a LASSO-type penalization, whereas SCAD has a different kind of regularization function, which can result in differences in the parameter estimates and hence may result in a better goodness-of-fit. The solution of any of the three sparse models outperforms BGR.

Figure 11: Adjacency matrices and serial dependence parameter in analysis with 3 groups and evaluation period length of 180 days.

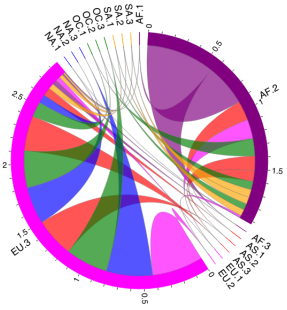
(a) Adjacency matrices for the 3 groups in 2012



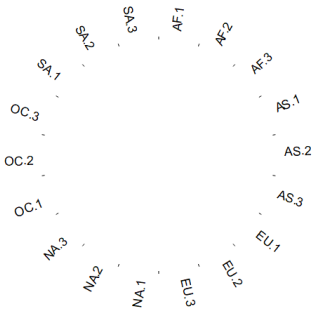
(b) Adjacency matrices for the 3 groups in 2013



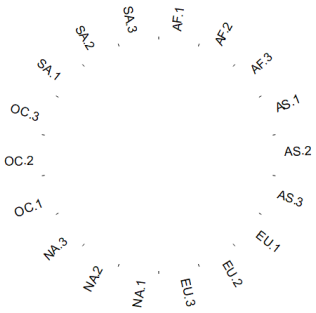
(c) Adjacency matrices for the 3 groups in 2014



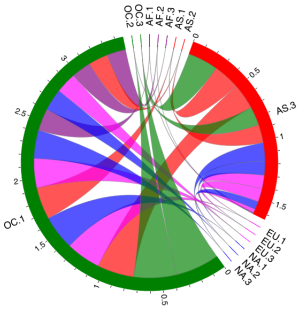
(d) Adjacency matrices for the 3 groups in 2015



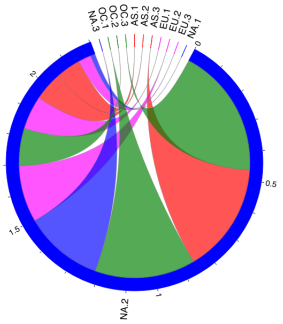
(e) Adjacency matrices for the 3 groups in 2016



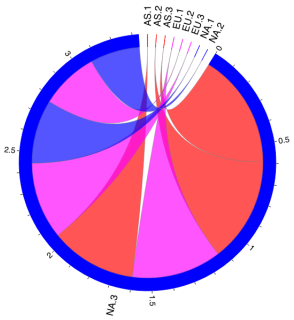
(f) Adjacency matrices for the 3 groups in 2017



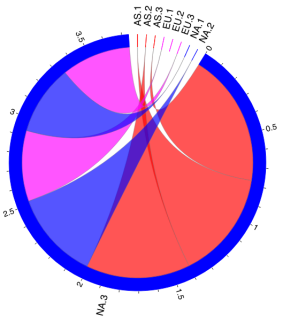
(g) Adjacency matrices for the 3 groups in 2018



(h) Adjacency matrices for the 3 groups in 2019



(i) Adjacency matrices for the 3 groups in 2020



Overall DINAR outperforms the competing models in terms of AIC and BIC over different length of evaluation periods, namely 90, 180, 270 and 365 days. Due to the almost always better results from DINAR, we conclude that the solution outperforms the competing models.

6 Conclusion

Cryptocurrencies have become interesting asset classes. BTC, being the elephant in the room, is traded all over the globe and virtually uncorrelated to any other asset class, which in principle is good for purposes of diversification. Besides the trading data on the exchanges, the blockchain displays a second layer of transactions, which are the actual shifts of funds directly between users without a middleman. The anonymity of the blockchain challenges the analysis, even though understanding the state of the network is important to understand a cryptocurrency. For the analysis of the blockchain, its huge dimensionality is challenging. We have proposed a method to Detect Influencers in Network AutoRegressive (DINAR) models via sparse-group regularization to analyse the time dependent network relations between the users of the BTC blockchain. We investigated the finite sample performance of DINAR in a large simulation study against 3 alternative methods which illustrates the outperformance of DINAR in terms of uncovering the underlying structure, in particular to identify the important network nodes. In the real data analysis, we discovered DINAR has a better modeling fit by AIC and BIC compared to the 3 alternative methods. The better fit is accompanied by DINAR uncovering the a underlying network structure, which allows us to infer about the impact of transactions between continents. With the developed algorithm to derive the adjacency matrix of DINAR, we find spatial connections in the BTC blockchain. We found in the year-to-year analysis that in the early years (2012 to 2016), network effects originated from continents such as Africa and South America, which is unexpected considering neither hosts a major financial centre nor is strongly covered by the media on cryptocurrency. Though the results are robust against different groupings, evaluation time periods and choice of regularization parameters. Additionally the DINAR model outperforms competing models during these years. From 2017 onwards network effects originate from Asia, Europe and for 2018, 2019 and 2020 majorly from North America which did not show any network effects in the years before. At the same time financial institutions, frequently based in North America, started offering products comprising BTC which might explain the stronger role of this continent and its impact on the transaction behaviour on the BTC blockchain. Taking into account that most Bitcoin mining farms are in Asia, it is surprising to some extent that Asia is not the sole driver but operates Bitcoin for Europe, North America, Africa, Oceania and South America, fostering the importance of these regions in the blockchain.

7 Acknowledgements

We would like to thank the editor, associate editor and three anonymous referees for their valuable comments on this article.

Table 7: BIC and AIC for out-of-sample performance and the number of parameters and identified groups whereas a group is counted as active if 25% of the parameters are different from 0. The results are displayed for the models found in the respective years for DINAR, LASSO, SCAD and BGR. The first 4 columns show the results evaluated on 90 days out-of-sample, the second 4 columns the results evaluated on 180 days out-of-sample.

		90 days				180 days			
		BIC	AIC	para	groups	BIC	AIC	para	groups
DINAR	2012	-12156	-12306	60	0	-20503	-20883	119	1
	2013	-16321	-16471	60	0	-26209	-26401	60	0
	2014	-16253	-16403	60	0	-30018	-30586	178	2
	2015	-17328	-17478	60	0	-31381	-31761	119	1
	2016	-22014	-22164	60	0	-35730	-36109	119	1
	2017	-5790	-5937	59	1	-10073	-10354	88	2
	2018	-4766	-4863	39	1	-8882	-9006	39	1
	2019	-5455	-5552	39	1	-9758	-9882	39	1
	2020	-5562	-5707	58	2	-10530	-10715	58	2
LASSO	2012	-12152	-12302	60	0	-20461	-20653	60	0
	2013	-16321	-16471	60	0	-26208	-26399	60	0
	2014	-16255	-16405	60	0	-29949	-30141	60	0
	2015	-17317	-17467	60	0	-31224	-31416	60	0
	2016	-22018	-22168	60	0	-35577	-35768	60	0
	2017	-5621	-5696	30	0	-9481	-9577	30	0
	2018	-4720	-4770	20	0	-8709	-8773	20	0
	2019	-5214	-5264	20	0	-9159	-9763	189	13
	2020	-5212	-5467	102	7	-10094	-10615	163	13
SCAD	2012	-12152	-12302	60	0	-20461	-20653	60	0
	2013	-16320	-16470	60	0	-26253	-26448	61	0
	2014	-16254	-16404	60	0	-29946	-30138	60	0
	2015	-17317	-17467	60	0	-31224	-31416	60	0
	2016	-22016	-22166	60	0	-35573	-35765	60	0
	2017	-5621	-5696	30	0	-9481	-9577	30	0
	2018	-4814	-4894	32	0	-9023	-9125	32	0
	2019	-5350	-5560	84	4	-9805	-10073	84	4
	2020	-5723	-5903	72	4	-10826	-11056	72	4
BGR	2012	3328	-5672	3600	60	-3157	-14652	3600	60
	2013	-1036	-10035	3600	60	-8947	-20442	3600	60
	2014	-1046	-10045	3600	60	-13312	-24807	3600	60
	2015	-2009	-11008	3600	60	-14150	-25645	3600	60
	2016	-6443	-15443	3600	60	-17964	-29459	3600	60
	2017	-2161	-4410	900	30	-6056	-8930	900	30
	2018	-3288	-4288	400	20	-7348	-8625	400	20
	2019	-3826	-4826	400	20	-7899	-9177	400	20
	2020	-4201	-5200	400	20	-9022	-10299	400	20

Table 8: BIC and AIC for out-of-sample performance and the number of parameters and identified groups whereas a group is counted as active if 25% of the parameters are different from 0. The results are displayed for the models found in the respective years for DINAR, LASSO, SCAD and BGR. The first 4 columns show the results evaluated on 270 days out-of-sample, the second 4 columns the results evaluated on the entire following year out-of-sample.

		270 days				one year			
		BIC	AIC	para	groups	BIC	AIC	para	groups
DINAR	2012	-31078	-31506	119	1	-42610	-43073	119	1
	2013	-36942	-37158	60	0	-46934	-47854	237	3
	2014	-38792	-39433	178	2	-39412	-40055	178	2
	2015	-44011	-44439	119	1	-56897	-57358	119	1
	2016	-65052	-65480	119	1	-72990	-73454	119	1
	2017	-14376	-14693	88	2	-17449	-17791	88	2
	2018	-13707	-13848	39	1	-18802	-18954	39	1
	2019	-13788	-13929	39	1	-18227	-18379	39	1
	2020	-14726	-14935	58	2	-20099	-20326	58	2
LASSO	2012	-30835	-31051	60	0	-42175	-42409	60	0
	2013	-36940	-37156	60	0	-46832	-47065	60	0
	2014	-38395	-38610	60	0	-39004	-39220	60	0
	2015	-43571	-43787	60	0	-56229	-56462	60	0
	2016	-64960	-65176	60	0	-72872	-73105	60	0
	2017	-13250	-13358	30	0	-16122	-16239	30	0
	2018	-13411	-13483	20	0	-18375	-18453	20	0
	2019	-13239	-13919	189	13	-17790	-18722	239	17
	2020	-14329	-14916	163	13	-19750	-20584	214	17
SCAD	2012	-30835	-31051	60	0	-42175	-42409	60	0
	2013	-37046	-37266	61	0	-46977	-47214	61	0
	2014	-38393	-38609	60	0	-39002	-39218	60	0
	2015	-43571	-43787	60	0	-56229	-56462	60	0
	2016	-64959	-65175	60	0	-72871	-73105	60	0
	2017	-13286	-13945	183	7	-16122	-16239	30	0
	2018	-13919	-14034	32	0	-19080	-19204	32	0
	2019	-13995	-14297	84	4	-18750	-19077	84	4
	2020	-15267	-15588	89	6	-20921	-21268	89	6
BGR	2012	-12694	-25649	3600	60	-23617	-37647	3600	60
	2013	-18913	-31868	3600	60	-28511	-42471	3600	60
	2014	-20992	-33947	3600	60	-21568	-34575	3600	60
	2015	-25676	-38630	3600	60	-38100	-52039	3600	60
	2016	-46577	-59531	3600	60	-53655	-67685	3600	60
	2017	-10166	-13404	900	30	-13070	-16578	900	30
	2018	-12277	-13716	400	20	-17502	-19061	400	20
	2019	-11950	-13389	400	20	-16460	-18020	400	20
	2020	-13149	-14588	400	20	-18555	-20114	400	20

References

Ahelegbey, D. F., M. Billio, and R. Casarin (2016). "Bayesian Graphical Models for Structural Vector Autoregressive Processes". *Journal of Applied Econometrics* 31.2, pp. 357–386.

Bańbura, M., D. Giannone, and L. Reichlin (2010). "Large Bayesian vector auto regressions". *Journal of Applied Econometrics* 25.1, pp. 71–92.

Basu, S. and G. Michailidis (2015). "Regularized estimation in sparse high-dimensional time series models". *The Annals of Statistics* 43.4, pp. 1535–1567.

Bianchi, D., M. Billio, R. Casarin, and M. Guidolin (2019). "Modeling systemic risk with Markov Switching Graphical SUR models". *Journal of Econometrics* 210.1, pp. 58–74.

Billio, M., R. Casarin, and L. Rossini (2019). "Bayesian nonparametric sparse VAR models". *Journal of Econometrics* 212.1, pp. 97–115.

Chudik, A. and M. H. Pesaran (2011). "Infinite-dimensional VARs and factor models". *Journal of Econometrics* 163.1, pp. 4–22.

Creal, D., S. J. Koopman, and A. Lucas (2013). "Generalized Autoregressive Score Models with Applications". *Journal of Applied Econometrics* 28.5, pp. 777–795.

Davis, R. A., P. Zang, and T. Zheng (2016). "Sparse Vector Autoregressive Modeling". *Journal of Computational and Graphical Statistics* 25.4, pp. 1077–1096.

Elendner, H., S. Trimborn, B. Ong, and T. M. Lee (2017). "The Cross-Section of Crypto-Currencies as Financial Assets: Investing in Crypto-currencies beyond Bitcoin". *Handbook of Blockchain, Digital Finance and Inclusion: Cryptocurrency, FinTech, InsurTech, and Regulation*. Ed. by D. Lee Kuo Chuen and R. Deng. 1st ed. Vol. 1. Elsevier, pp. 145–173.

Fan, J. and R. Li (2001). "Variable selection via nonconcave penalized likelihood and its oracle properties". *Journal of the American statistical Association* 96.456, pp. 1348–1360.

Feng, W., Y. Wang, and Z. Zhang (2017). "Informed trading in the Bitcoin market". *Finance Research Letters*.

Feng, W., Y. Wang, and Z. Zhang (2018). "Can cryptocurrencies be a safe haven: a tail risk perspective analysis". *Applied Economics* 50.44, pp. 4745–4762.

Foley, S., J. R. Karlsen, and T. J. Putniņš (2019). "Sex, Drugs, and Bitcoin: How Much Illegal Activity Is Financed through Cryptocurrencies?" *The Review of Financial Studies* 32.5, pp. 1798–1853.

Friedman, J., T. Hastie, and R. Tibshirani (2010). "A note on the group lasso and a sparse group lasso".

Griffin, J. M. and A. Shams (2020). "Is Bitcoin really untethered?" *The Journal of Finance* 75.4, pp. 1913–1964.

Guo, J., J. Hu, B.-Y. Jing, and Z. Zhang (2016). "Spline-Lasso in High-Dimensional Linear Regression". *Journal of the American Statistical Association* 111.513, pp. 288–297.

Hafner, C. (2018). "Testing for Bubbles in Cryptocurrencies with Time-Varying Volatility". *Journal of Financial Econometrics*.

Hoerl, A. and R. Kennard (1988). "Ridge regression". *in Encyclopedia of Statistical Sciences* 8, pp. 129–136.

Lischke, M. and B. Fabian (2016). "Analyzing the Bitcoin Network: The First Four Years". *Future Internet* 8.1.

Liu, Y., A. Tsyvinski, and X. Wu (2019). "Common Risk Factors in Cryptocurrency". 25882.

Makarov, I. and A. Schoar (2019). "Trading and arbitrage in cryptocurrency markets". *Journal of Financial Economics*.

Makarov, I. and A. Schoar (2021). "Blockchain analysis of the bitcoin market".

Nicholson, W., D. Matteson, and J. Bien (2017). "VARX-L: Structured Regularization for Large Vector Autoregressions with Exogenous Variables". *International Journal of Forecasting* 33.3, pp. 627–651.

Nicholson, W., D. Matteson, J. Bien, and I. Wilms (2022). "BigVAR: Dimension Reduction Methods for Multivariate Time Series".

Ord, K. (1975). "Estimation Methods for Models of Spatial Interaction". *Journal of the American Statistical Association* 70.349, pp. 120–126.

Pesaran, M. H., T. Schuermann, and S. M. Weiner (2004). "Modeling Regional Interdependencies Using a Global Error-Correcting Macroeconometric Model". *Journal of Business & Economic Statistics* 22.2, pp. 129–162.

Reid, F. and M. Harrigan (2013). "An Analysis of Anonymity in the Bitcoin System". *Security and Privacy in Social Networks*. Ed. by Y. Altshuler, Y. Elovici, A. B. Cremers, N. Aharony, and A. Pentland. Springer New York, pp. 197–223.

Ron, D. and A. Shamir (2013). "Quantitative Analysis of the Full Bitcoin Transaction Graph". *Financial Cryptography and Data Security*. Ed. by A.-R. Sadeghi. Lecture Notes in Computer Science 7859. Springer Berlin Heidelberg, pp. 6–24.

Sabah, N. (2020). "Cryptocurrency accepting venues, investor attention, and volatility". *Finance Research Letters* 36, p. 101339.

Scaillet, O., A. Treccani, and C. Trevisan (2018). "High-frequency jump analysis of the bitcoin market". *Journal of Financial Econometrics*.

Simon, N., J. Friedman, T. Hastie, and R. Tibshirani (2013). "A Sparse-Group Lasso". *Journal of Computational and Graphical Statistics* 22.2, pp. 231–245.

Song, S. and P. J. Bickel (2011). "Large vector auto regressions". *arXiv preprint arXiv:1106.3915*.

Tibshirani, R. (1996). "Regression Shrinkage and Selection via the Lasso". *Journal of the Royal Statistical Society. Series B (Methodological)* 58.1, pp. 267–288.

Tibshirani, R., M. Saunders, S. Rosset, J. Zhu, and K. Knight (2005). "Sparsity and smoothness via the fused lasso". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.1, pp. 91–108.

Trimborn, S. and L. Yu (2022). "Blockchain meets network analytics: a tale of heuristics, location and fraud detection".

Wang, H., G. Li, and C.-L. Tsai (2007). "Regression Coefficient and Autoregressive Order Shrinkage and Selection via the Lasso". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 69.1, pp. 63–78.

Wu, T. T. and K. Lange (2008). "Coordinate descent algorithms for lasso penalized regression". *The Annals of Applied Statistics* 2.1, pp. 224–244.

Yuan, M. and Y. Lin (2006). "Model selection and estimation in regression with grouped variables". *Journal of the Royal Statistical Society, Series B* 68, pp. 49–67.

Zhou, J., Y. Tu, Y. Chen, and H. Wang (2017). "Estimating Spatial Autocorrelation With Sampled Network Data". *Journal of Business & Economic Statistics* 35.1, pp. 130–138.

Zhu, X., R. Pan, G. Li, Y. Liu, and H. Wang (2017). "Network vector autoregression". *The Annals of Statistics* 45.3, pp. 1096–1123.

Zou, H. and T. Hastie (2005). "Regularization and variable selection via the elastic net". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.2, pp. 301–320.

I Appendix

I.1 Gradient Descent

In every iteration, the parameters of a particular group are optimized, while the remaining parameters are fixed.

Specifically, let $A_{\cdot i}$ be the ith column/group to be optimized in an iteration step. The remaining parameters in $A_{\cdot -i}$ are held fixed when optimizing the ith column. We construct the partial residuals of $\mathbf{Y}_t$, which contain the dependence unexplained by the already optimized parameters:

$$\begin{aligned} r_{t,-\beta_0} &= \mathbf{Y}_t - (I_N\beta_1 + A)\mathbf{Y}_{t-1}, \\ r_{t,-\beta_1} &= \mathbf{Y}_t - \mathbf{1}\beta_0 - A\mathbf{Y}_{t-1}, \\ r_{t,-A_{\cdot i}} &= \mathbf{Y}_t - \mathbf{1}\beta_0 - (I_N\beta_1 + A_{\cdot -i})\mathbf{Y}_{t-1}, \end{aligned}$$

The following are the loss functions:

$$\begin{aligned} L(r_{-\beta_0};\beta_0) &= \frac{1}{2N}\sum_{t=2}^{T}||r_{t,-\beta_0} - \mathbf{1}\beta_0||_F^2. \\ L(r_{-\beta_1};\beta_1) &= \frac{1}{2N}\sum_{t=2}^{T}||r_{t,-\beta_1} - I_N\beta_1\mathbf{Y}_{t-1}||_F^2, \\ L(r_{-A_{\cdot i}};A_{\cdot i}) &= \frac{1}{2N}\sum_{t=2}^{T}||r_{t,-A_{\cdot i}} - A_{\cdot i}\mathbf{Y}_{t-1}||_F^2. \end{aligned}$$

To simplify the notation, let $\theta_1 = \beta_0$, $\theta_2 = \beta_1$ and $\theta_k = A_{\cdot i}$, $k = 3, \cdots, N+2$. We rewrite the optimization in this particular iterative step as

$$\hat{\theta}_k = \underset{\theta_k}{\operatorname{argmin}}\, L(r_{-\theta_k};\theta_k) + (1-\alpha)\lambda||\theta_k||_F + \sum_{i=1}^{N}\alpha\lambda|\theta_{k,i}| \tag{6}$$

where for $k = \{1, 2\}$ the penalty term is set to $\lambda = 0$, namely no sparsity penalization is applied for β_0 and β_1.

There is no closed form solution for the non-convex optimization problem in (3). We introduce a two-step gradient descent algorithm to numerically estimate β_0, β_1 and A in the DINAR framework. We derive the updating function for each iteration step l. Using a Taylor expansion, we formulate an upper bound for $L(r_{-\theta_k^{(l)}};\theta_k^{(l)})$ depending on the $\theta_k^{(l-1)}$ that has been optimized in the previous iteration step $l-1$.

The minimization problem can be equivalently solved by minimizing

$$M(\theta_k^{(l)}) = L(r_{-\theta_k^{(l-1)}}; \theta_k^{(l-1)}) + (\theta_k^{(l)} - \theta_k^{(l-1)})^\top \nabla L(r_{-\theta_k^{(l-1)}}; \theta_k^{(l-1)}) + \frac{1}{2\xi}||\theta_k^{(l)} - \theta_k^{(l-1)}||_F^2 \tag{7}$$

$$+ (1-\alpha)\lambda||\theta_k^{(l)}||_F + \sum_{i=1}^{N} \alpha\lambda|\theta_{k,i}^{(l)}|,$$

where ξ is small enough so that the quadratic term dominates the Hessian of the loss function. This approach is also known as the majorize-minimization approach, Wu and Lange (2008). Note in the case of $k = \{1, 2\}$, λ is set to 0.

The first term of Equation (7) does not depend on $\theta_k^{(l)}$, thus it can be further simplified to

$$M(\theta_k^{(l)}) \propto \frac{1}{2\xi}||\theta_k^{(l)} - \{\theta_k^{(l-1)} - \xi\nabla L(r_{-\theta_k^{(l-1)}}; \theta_k^{(l-1)})\}||_F^2 \tag{8}$$

$$+ (1-\alpha)\lambda||\theta_k^{(l)}||_F + \sum_{i=1}^{N} \alpha\lambda|\theta_{k,i}^{(l)}|.$$

The loss function is embedded into the thresholding function of the Lasso as follows:

$$S(z, \alpha\lambda) = \text{sign}(z) \circ (|z| - \alpha\lambda)_+,$$

where $\circ$ denotes the Hadamard product. This leads to $\hat{\theta}_k = 0$ if

$$||S\left\{\theta_k^{(l-1)} - \xi\nabla L(r_{-\theta_k^{(l-1)}}; \theta_k^{(l-1)}), \xi\alpha\lambda\right\}||_F \leq \xi(1-\alpha)\lambda$$

and otherwise

$$\left\{1 + \xi(1-\alpha)\lambda/||\theta_k^{(l)}||_F\right\}\theta_k^{(l)} = S\left\{\theta_k^{(l-1)} - \xi\nabla L(r_{-\theta_k^{(l-1)}}; \theta_k^{(l-1)}), \xi\alpha\lambda\right\}$$

The solution to (8) satisfies

$$\theta_k^{(l)} = \left(1 - \frac{\xi(1-\alpha)\lambda}{||S(\theta_k^{(l-1)} - \xi\nabla L(r_{-\theta_k^{(l-1)}}; \theta_k^{(l-1)}), \xi\alpha\lambda)||_F}\right)_+ S(\theta_k^{(l-1)} - \xi\nabla L(r_{-\theta_k^{(l-1)}}; \theta_k^{(l-1)}), \xi\alpha\lambda). \tag{9}$$

I.2 Algorithm

The two-layer sparsity has both group and individual terms which are inseparably connected. Friedman et al. (2010) outline an idea for an algorithm that would be

applicable in such situations. Yet the idea was designed for the univariate case and the groups are defined on the rows. In our multivariate case, we define the group on the columns, as we are looking for leading groups which influence the future values of other groups. This makes the groups dependent on each other.

Simon et al. (2013) formulated the algorithm for univariate regression models, which translates to a regression on a vector of length $(T-1)$. Sparsity of the rows instead of the columns would result in an optimization problem which requires less computation time to find a solution since each group can be optimized independently from the others. However this does not allow for a network interpretation. Because of the interdependency of the groups, the optimization problem cannot be written in a vectorized form. The complexity of the model challenges the algorithm, resulting in a longer runtime. We propose a new algorithm customized for the multivariate case with groups defined on the columns. The algorithm initializes with all parameters set to be 0. It iterates through each group of parameters by starting with the parameters β_0, β_1 to control for the effects of the intercept and autoregressive dependence before optimizing on the groups in the adjacency matrix A. The algorithm optimizes at the update step width ξ, before the current group θ_k gets updated. The update of θ_k is performed until an a priori chosen vectorized threshold value ϵ_2 is reached. When θ_k has been updated, the next group gets optimized until a full walk through all the groups of parameters has been performed. This procedure repeats until it converges. In detail, the algorithm works as described in Algorithm 1.

The parameter ϵ_1 can be set to any value in $(0, 1)$. Its value controls the density of the grid in which the search for the updating value of parameters takes place. The smaller it is, the faster the algorithm, so one can use it to speed up the computationally intensive method. The entire algorithm works under a chosen mixing parameter α and a penalty parameter λ. The algorithm converges when a vectorized threshold parameter ϵ_3 is satisfied.

The algorithm depends on the hyperparameter λ. It controls the level of penalization, which balances the sparseness of the model against the fit. We derive first which level of λ sets all groups to 0 by following the approach of Simon et al. (2013). The path is started with λ_{max} and from there on a halving sequence is created. In the spirit of Simon et al. (2013), the mixing parameter α is set to be $\alpha = 1/N$, which gives equal importance to group and individual sparsity.

The settings for running Algorithm 1 are as described in Algorithm 2.

Algorithm 1 DINAR optimization algorithm

Input: Data $\mathbf{Y}_t$ for all $t = 1, \ldots, N$
Output: Adjacency matrix A

Initialization $\beta_0 = 0$, $\beta_1 = 0$, $A = 0$, $m = 1$
Set $\theta_1 = \beta_0, \theta_2 = \beta_1, \theta_k = A_{.i}, i = 1, \ldots, N, k = i + 2$
while $\text{vec}\{A^{(m)} - A^{(m-1)}\} < \epsilon_3$, $\beta_0^{(m)} - \beta_0^{(m-1)} < \epsilon_3$ **or** $\beta_1^{(m)} - \beta_1^{(m-1)} < \epsilon_3$ **do**
  for $k = 1, \ldots, N + 2$ **do**
    $l = 2$
    while $\theta_k^{(l)} - \theta_k^{(l-1)} < \epsilon_2$ **do**
      $\xi = 1$
      while ξ small enough such that it holds $L(r_{-U}; U) \leq L(r_{-\theta_k^{(l-1)}}; \theta_k^{(l-1)}) + (U - \theta_k^{(l-1)})^\top \nabla L(r_{-\theta_k^{(l-1)}}; \theta_k^{(l-1)}) + \frac{1}{2\xi}||U - \theta_k^{(l-1)}||^2$ **do**
        $z = \theta_k^{(l-1)} - \xi * \nabla L(r_{-\theta_k^{(l-1)}}; \theta_k^{(l-1)})$
        $S = \text{sign}(z) \circ (|z| - \alpha\lambda)_+$
        $U = \{1 - \xi(1 - \alpha)\lambda / ||S||\}_+ S$
        $\xi = \epsilon_1 * \xi$
      end while
      $\theta_k^{(l)} = U_{l-1} + \frac{(l-1)}{(l+2)}(U_l - U_{l-1})$
      $l = l + 1$
    end while
  end for
  $m = m + 1$
end while

Algorithm 2 DINAR algorithmic procedure

1. Run **Algorithm 1** for each λ with J iterations
2. Fix identified groups from step 1.
3. To obtain warm starting values, run **Algorithm 1** without penalization for identified groups
4. Utilize results from 3. as starting values
5. Run **Algorithm 1** with λ sequence

I.3 Simulation results & additional Figures

Table 9: Simulation $N = 10$ with $T = 100, 500, 1000$ for low, medium, strong persistence evaluated with BIC.

	Persistence	Models	FN.g	FP.g	FN.e	FP.e	RMSE para	RMSE out
100	low	DINAR	0.8	0	0.66	0	0.07	0.89
		SCAD	0.62	0.01	0.62	0	0.07	0.89
		LASSO	0.62	0.01	0.62	0	0.07	0.89
		BGR	0	0.55	0	0.71	0.07	0.88
	medium	DINAR	0.79	0	0.64	0.01	0.09	0.87
		SCAD	0.47	0.05	0.54	0.02	0.09	0.86
		LASSO	0.58	0.01	0.61	0	0.09	0.87
		BGR	0	0.55	0	0.71	0.08	0.83
	strong	DINAR	0.42	0	0.34	0.24	0.09	0.75
		SCAD	0.08	0.26	0.23	0.13	0.1	0.74
		LASSO	0.4	0.2	0.38	0.13	0.1	0.77
		BGR	0	0.55	0	0.71	0.09	0.71
500	low	DINAR	0.13	0	0.1	0.34	0.04	0.86
		SCAD	0	0.37	0.01	0.15	0.04	0.86
		LASSO	0.02	0.39	0.02	0.2	0.04	0.87
		BGR	0	0.55	0	0.71	0.04	0.86
	medium	DINAR	0	0	0	0.37	0.04	0.81
		SCAD	0	0.24	0	0.08	0.04	0.81
		LASSO	0	0.42	0	0.22	0.04	0.81
		BGR	0	0.55	0	0.71	0.05	0.81
	strong	DINAR	0	0	0	0.37	0.05	0.69
		SCAD	0	0.08	0.01	0.03	0.05	0.69
		LASSO	0	0.53	0	0.45	0.05	0.7
		BGR	0	0.55	0	0.71	0.07	0.69
1000	low	DINAR	0	0	0	0.37	0.03	0.86
		SCAD	0	0.15	0	0.05	0.03	0.86
		LASSO	0	0.28	0	0.12	0.03	0.86
		BGR	0	0.55	0	0.71	0.04	0.86
	medium	DINAR	0	0	0	0.37	0.04	0.81
		SCAD	0	0.05	0	0.02	0.04	0.8
		LASSO	0	0.43	0	0.27	0.04	0.81
		BGR	0	0.55	0	0.71	0.04	0.8
	strong	DINAR	0	0	0	0.37	0.05	0.69
		SCAD	0	0.05	0	0.02	0.05	0.69
		LASSO	0	0.54	0	0.52	0.05	0.69
		BGR	0	0.55	0	0.71	0.06	0.69

Table 10: Simulation $N = 20$ with $T = 100, 500, 1000$ for low, medium, strong persistence evaluated with BIC.

	Persistence	Models	FN.g	FP.g	FN.e	FP.e	RMSE para	RMSE out
100	low	DINAR	0.8	0	0.66	0	0.05	0.89
		SCAD	0.59	0.05	0.64	0.01	0.05	0.89
		LASSO	0.59	0.05	0.64	0.01	0.05	0.89
		BGR	0	0.76	0	0.85	0.07	0.91
	medium	DINAR	0.79	0	0.66	0.01	0.06	0.87
		SCAD	0.58	0.01	0.64	0	0.07	0.87
		LASSO	0.61	0.01	0.64	0	0.06	0.87
		BGR	0	0.76	0	0.85	0.08	0.85
	strong	DINAR	0.6	0	0.49	0.18	0.08	0.76
		SCAD	0.51	0.12	0.57	0.06	0.08	0.79
		LASSO	0.71	0.03	0.64	0.01	0.08	0.8
		BGR	0	0.76	0	0.85	0.09	0.73
500	low	DINAR	0.36	0	0.29	0.22	0.04	0.87
		SCAD	0.16	0.32	0.21	0.11	0.04	0.87
		LASSO	0.62	0.09	0.54	0.04	0.04	0.88
		BGR	0	0.76	0	0.85	0.04	0.87
	medium	DINAR	0	0	0	0.39	0.03	0.81
		SCAD	0	0.13	0.04	0.02	0.03	0.81
		LASSO	0	0.72	0	0.35	0.03	0.81
		BGR	0	0.76	0	0.85	0.04	0.81
	strong	DINAR	0	0	0	0.39	0.04	0.69
		SCAD	0	0.58	0	0.17	0.05	0.69
		LASSO	0	0.72	0	0.4	0.04	0.7
		BGR	0	0.76	0	0.85	0.06	0.69
1000	low	DINAR	0	0	0	0.39	0.02	0.86
		SCAD	0	0.45	0	0.09	0.02	0.86
		LASSO	0	0.62	0	0.2	0.02	0.86
		BGR	0	0.76	0	0.85	0.03	0.86
	medium	DINAR	0	0	0	0.39	0.03	0.8
		SCAD	0	0.03	0.01	0.01	0.03	0.8
		LASSO	0	0.64	0	0.22	0.03	0.81
		BGR	0	0.76	0	0.85	0.03	0.8
	strong	DINAR	0	0	0	0.39	0.04	0.69
		SCAD	0	0.44	0	0.1	0.04	0.69
		LASSO	0	0.76	0	0.6	0.03	0.69
		BGR	0	0.76	0	0.85	0.05	0.68

Table 11: Simulation $N = 60$ with $T = 100, 500, 1000$ for low, medium, strong persistence evaluated with BIC.

	Persistence	Models	FN.g	FP.g	FN.e	FP.e	RMSE para	RMSE out
100	low	DINAR	0.8	0	0.66	0	0.03	0.89
		SCAD	0.55	0.18	0.66	0.01	0.03	0.89
		LASSO	0.6	0.13	0.66	0.01	0.03	0.89
		BGR	0	0.92	0	0.95	0.09	1.07
	medium	DINAR	0.8	0	0.66	0	0.04	0.87
		SCAD	0.32	0.35	0.64	0.04	0.04	0.87
		LASSO	0.66	0.07	0.66	0	0.04	0.87
		BGR	0	0.92	0	0.95	0.09	0.99
	strong	DINAR	0.74	0	0.61	0.07	0.05	0.79
		SCAD	0.25	0.48	0.62	0.08	0.05	0.8
		LASSO	0.77	0.01	0.66	0	0.05	0.8
		BGR	0	0.92	0	0.95	0.1	0.84
500	low	DINAR	0.8	0	0.66	0	0.03	0.89
		SCAD	0.67	0.06	0.61	0.01	0.03	0.89
		LASSO	0.76	0	0.66	0	0.03	0.89
		BGR	0	0.92	0	0.95	0.03	0.89
	medium	DINAR	0.04	0	0.03	0.38	0.02	0.81
		SCAD	0.11	0.59	0.23	0.24	0.03	0.83
		LASSO	0.78	0	0.66	0	0.04	0.86
		BGR	0	0.92	0	0.95	0.04	0.83
	strong	DINAR	0	0	0	0.4	0.02	0.69
		SCAD	0	0.9	0.01	0.58	0.04	0.69
		LASSO	0	0.91	0	0.44	0.02	0.71
		BGR	0	0.92	0	0.95	0.04	0.7
1000	low	DINAR	0	0	0	0.4	0.01	0.86
		SCAD	0	0.22	0.08	0.01	0.02	0.86
		LASSO	0.44	0.41	0.37	0.17	0.02	0.88
		BGR	0	0.92	0	0.95	0.02	0.87
	medium	DINAR	0	0	0	0.4	0.02	0.8
		SCAD	0	0.85	0	0.24	0.02	0.8
		LASSO	0	0.9	0	0.36	0.02	0.81
		BGR	0	0.92	0	0.95	0.03	0.81
	strong	DINAR	0	0	0	0.4	0.02	0.68
		SCAD	0	0.91	0	0.54	0.03	0.68
		LASSO	0	0.87	0	0.25	0.02	0.71
		BGR	0	0.92	0	0.95	0.04	0.69

Table 12: Simulation $N = 20$ with $T = 100, 500, 1000$ for low, medium, strong persistence for an increasing number of groups with increasing dimensionality and evaluated with BIC.

	Persistence	Models	FN.g	FP.g	FN.e	FP.e	RMSE para	RMSE out
100	low	DINAR	0.86	0	0.74	0	0.06	0.94
		SCAD	0.37	0.14	0.67	0.04	0.06	0.93
		LASSO	0.37	0.14	0.67	0.04	0.06	0.93
		BGR	0	0.67	0	0.8	0.07	0.94
	medium	DINAR	0.84	0	0.73	0.03	0.08	0.9
		SCAD	0.46	0.05	0.69	0.01	0.08	0.9
		LASSO	0.49	0.04	0.69	0.01	0.08	0.9
		BGR	0	0.67	0	0.8	0.07	0.84
	strong	DINAR	0.67	0.01	0.58	0.23	0.1	0.78
		SCAD	0.23	0.35	0.38	0.17	0.09	0.76
		LASSO	0.67	0.06	0.67	0.03	0.11	0.83
		BGR	0	0.67	0	0.8	0.08	0.71
500	low	DINAR	0.76	0	0.66	0.13	0.05	0.92
		SCAD	0.02	0.15	0.16	0.04	0.04	0.9
		LASSO	0.25	0.14	0.39	0.03	0.05	0.92
		BGR	0	0.67	0	0.8	0.03	0.89
	medium	DINAR	0	0	0	0.42	0.03	0.8
		SCAD	0	0.15	0.04	0.02	0.03	0.8
		LASSO	0	0.63	0	0.3	0.03	0.81
		BGR	0	0.67	0	0.8	0.04	0.8
	strong	DINAR	0	0	0	0.42	0.03	0.67
		SCAD	0	0.35	0.01	0.08	0.04	0.67
		LASSO	0	0.66	0	0.44	0.03	0.68
		BGR	0	0.67	0	0.8	0.04	0.67
1000	low	DINAR	0	0	0	0.42	0.02	0.89
		SCAD	0	0.49	0	0.14	0.02	0.89
		LASSO	0	0.61	0	0.25	0.02	0.89
		BGR	0	0.67	0	0.8	0.02	0.89
	medium	DINAR	0	0	0	0.42	0.02	0.8
		SCAD	0	0.02	0	0	0.02	0.8
		LASSO	0	0.57	0	0.23	0.03	0.81
		BGR	0	0.67	0	0.8	0.03	0.8
	strong	DINAR	0	0	0	0.42	0.03	0.67
		SCAD	0	0.13	0	0.02	0.03	0.67
		LASSO	0	0.66	0	0.45	0.03	0.67
		BGR	0	0.67	0	0.8	0.04	0.67

Table 13: Simulation $N = 60$ with $T = 100, 500, 1000$ for low, medium, strong persistence for an increasing number of groups with increasing dimensionality and evaluated with BIC.

	Persistence	Models	FN.g	FP.g	FN.e	FP.e	RMSE para	RMSE out
100	low	DINAR	0.91	0	0.77	0	0.04	0.98
		SCAD	0.27	0.47	0.7	0.12	0.04	0.98
		LASSO	0.26	0.48	0.7	0.13	0.04	0.97
		BGR	0	0.82	0	0.93	0.08	1.11
	medium	DINAR	0.91	0	0.76	0	0.05	0.96
		SCAD	0.36	0.33	0.7	0.06	0.05	0.96
		LASSO	0.42	0.27	0.71	0.05	0.05	0.96
		BGR	0	0.82	0	0.93	0.09	1.04
	strong	DINAR	0.88	0.01	0.74	0.1	0.07	0.94
		SCAD	0.39	0.28	0.69	0.05	0.07	0.92
		LASSO	0.6	0.16	0.73	0.02	0.07	0.95
		BGR	0	0.82	0	0.93	0.09	0.96
500	low	DINAR	0.91	0	0.77	0	0.04	0.97
		SCAD	0.17	0.09	0.59	0.02	0.03	0.96
		LASSO	0.19	0.01	0.65	0	0.04	0.97
		BGR	0	0.82	0	0.93	0.03	0.95
	medium	DINAR	0.78	0	0.65	0.36	0.05	0.93
		SCAD	0	0.65	0.11	0.16	0.02	0.88
		LASSO	0.11	0.39	0.26	0.07	0.04	0.92
		BGR	0	0.82	0	0.93	0.03	0.89
	strong	DINAR	0.61	0	0.51	0.49	0.05	0.85
		SCAD	0.12	0.71	0.15	0.21	0.03	0.81
		LASSO	0	0.74	0.01	0.26	0.03	0.82
		BGR	0	0.82	0	0.93	0.04	0.82
1000	low	DINAR	0.91	0	0.76	0	0.04	0.97
		SCAD	0	0.26	0.05	0.02	0.01	0.93
		LASSO	0.01	0.38	0.04	0.04	0.02	0.94
		BGR	0	0.82	0	0.93	0.02	0.94
	medium	DINAR	0	0	0	0.61	0.01	0.87
		SCAD	0	0.51	0.09	0.08	0.02	0.87
		LASSO	0	0.74	0	0.24	0.02	0.88
		BGR	0	0.82	0	0.93	0.02	0.88
	strong	DINAR	0	0	0	0.61	0.02	0.8
		SCAD	0	0.73	0.01	0.29	0.02	0.8
		LASSO	0	0.8	0	0.43	0.02	0.8
		BGR	0	0.82	0	0.93	0.03	0.81

Table 14: Simulation $N = 10$ with $T = 100, 500, 1000$ for low, medium, strong persistence evaluated with AIC.

	Persistence	Models	FN.g	FP.g	FN.e	FP.e	RMSE para	RMSE out
100	low	DINAR	0.77	0	0.63	0.02	0.07	0.89
		SCAD	0.58	0.03	0.6	0.02	0.07	0.89
		LASSO	0.59	0.03	0.6	0.01	0.07	0.89
		BGR	0	0.55	0	0.71	0.07	0.88
	medium	DINAR	0.39	0	0.32	0.22	0.08	0.84
		SCAD	0.21	0.18	0.37	0.08	0.09	0.85
		LASSO	0.19	0.27	0.3	0.17	0.07	0.85
		BGR	0	0.55	0	0.71	0.08	0.83
	strong	DINAR	0.12	0.01	0.09	0.36	0.08	0.72
		SCAD	0.01	0.33	0.16	0.18	0.1	0.73
		LASSO	0.01	0.51	0.04	0.39	0.07	0.72
		BGR	0	0.55	0	0.71	0.09	0.71
500	low	DINAR	0	0.01	0	0.38	0.04	0.86
		SCAD	0	0.37	0.01	0.16	0.04	0.86
		LASSO	0	0.44	0	0.26	0.04	0.86
		BGR	0	0.55	0	0.71	0.04	0.86
	medium	DINAR	0	0	0	0.37	0.04	0.81
		SCAD	0	0.25	0	0.08	0.04	0.81
		LASSO	0	0.53	0	0.44	0.04	0.81
		BGR	0	0.55	0	0.71	0.05	0.81
	strong	DINAR	0	0.01	0	0.38	0.05	0.69
		SCAD	0	0.14	0	0.06	0.06	0.69
		LASSO	0	0.54	0	0.54	0.06	0.69
		BGR	0	0.55	0	0.71	0.07	0.69
1000	low	DINAR	0	0	0	0.37	0.03	0.86
		SCAD	0	0.16	0	0.05	0.03	0.86
		LASSO	0	0.52	0	0.38	0.03	0.86
		BGR	0	0.55	0	0.71	0.04	0.86
	medium	DINAR	0	0	0	0.37	0.04	0.81
		SCAD	0	0.06	0	0.02	0.04	0.8
		LASSO	0	0.53	0	0.43	0.04	0.81
		BGR	0	0.55	0	0.71	0.04	0.8
	strong	DINAR	0	0.02	0	0.38	0.05	0.69
		SCAD	0	0.21	0	0.13	0.06	0.69
		LASSO	0	0.55	0	0.6	0.06	0.69
		BGR	0	0.55	0	0.71	0.06	0.69

Table 15: Simulation $N = 20$ with $T = 100, 500, 1000$ for low, medium, strong persistence evaluated with AIC.

	Persistence	Models	FN.g	FP.g	FN.e	FP.e	RMSE para	RMSE out
100	low	DINAR	0.79	0	0.65	0.01	0.05	0.89
		SCAD	0.59	0.05	0.64	0.01	0.05	0.89
		LASSO	0.59	0.05	0.64	0.01	0.05	0.89
		BGR	0	0.76	0	0.85	0.07	0.91
	medium	DINAR	0.62	0	0.51	0.16	0.06	0.85
		SCAD	0.51	0.08	0.6	0.03	0.07	0.86
		LASSO	0.58	0.03	0.62	0.01	0.06	0.87
		BGR	0	0.76	0	0.85	0.08	0.85
	strong	DINAR	0.16	0	0.13	0.35	0.06	0.72
		SCAD	0.08	0.61	0.23	0.32	0.09	0.74
		LASSO	0.1	0.61	0.17	0.37	0.06	0.74
		BGR	0	0.76	0	0.85	0.09	0.73
500	low	DINAR	0	0	0	0.39	0.03	0.86
		SCAD	0	0.68	0.02	0.28	0.03	0.86
		LASSO	0	0.73	0.01	0.37	0.03	0.86
		BGR	0	0.76	0	0.85	0.04	0.87
	medium	DINAR	0	0	0	0.39	0.03	0.81
		SCAD	0	0.2	0.03	0.05	0.03	0.81
		LASSO	0	0.73	0	0.39	0.03	0.81
		BGR	0	0.76	0	0.85	0.04	0.81
	strong	DINAR	0	0	0	0.39	0.04	0.69
		SCAD	0	0.58	0	0.18	0.05	0.69
		LASSO	0	0.76	0	0.64	0.04	0.69
		BGR	0	0.76	0	0.85	0.06	0.69
1000	low	DINAR	0	0	0	0.39	0.02	0.86
		SCAD	0	0.45	0	0.09	0.02	0.86
		LASSO	0	0.66	0	0.29	0.02	0.86
		BGR	0	0.76	0	0.85	0.03	0.86
	medium	DINAR	0	0	0	0.39	0.03	0.8
		SCAD	0	0.07	0	0.02	0.03	0.8
		LASSO	0	0.76	0	0.57	0.03	0.8
		BGR	0	0.76	0	0.85	0.03	0.8
	strong	DINAR	0	0.04	0	0.41	0.04	0.69
		SCAD	0	0.46	0	0.12	0.04	0.69
		LASSO	0	0.76	0	0.7	0.04	0.69
		BGR	0	0.76	0	0.85	0.05	0.68

Table 16: Simulation $N = 60$ with $T = 100, 500, 1000$ for low, medium, strong persistence evaluated with AIC.

	Persistence	Models	FN.g	FP.g	FN.e	FP.e	RMSE para	RMSE out
100	low	DINAR	0.8	0	0.66	0	0.03	0.89
		SCAD	0.55	0.18	0.66	0.01	0.03	0.89
		LASSO	0.6	0.13	0.66	0.01	0.03	0.89
		BGR	0	0.92	0	0.95	0.09	1.07
	medium	DINAR	0.79	0	0.65	0.01	0.04	0.87
		SCAD	0.32	0.35	0.64	0.04	0.04	0.87
		LASSO	0.66	0.07	0.66	0	0.04	0.87
		BGR	0	0.92	0	0.95	0.09	0.99
	strong	DINAR	0.54	0	0.45	0.23	0.04	0.75
		SCAD	0.25	0.48	0.62	0.08	0.05	0.8
		LASSO	0.77	0.01	0.66	0	0.05	0.8
		BGR	0	0.92	0	0.95	0.1	0.84
500	low	DINAR	0	0	0	0.4	0.02	0.86
		SCAD	0.01	0.68	0.18	0.11	0.02	0.87
		LASSO	0.03	0.57	0.17	0.08	0.02	0.88
		BGR	0	0.92	0	0.95	0.03	0.89
	medium	DINAR	0	0	0	0.4	0.02	0.81
		SCAD	0	0.89	0.02	0.38	0.03	0.81
		LASSO	0	0.91	0	0.54	0.02	0.81
		BGR	0	0.92	0	0.95	0.04	0.83
	strong	DINAR	0	0	0	0.4	0.02	0.69
		SCAD	0	0.91	0	0.59	0.04	0.69
		LASSO	0	0.92	0	0.71	0.03	0.69
		BGR	0	0.92	0	0.95	0.04	0.7
1000	low	DINAR	0	0	0	0.4	0.01	0.86
		SCAD	0	0.25	0.07	0.03	0.02	0.86
		LASSO	0	0.91	0	0.4	0.02	0.86
		BGR	0	0.92	0	0.95	0.02	0.87
	medium	DINAR	0	0	0	0.4	0.02	0.8
		SCAD	0	0.85	0	0.24	0.02	0.8
		LASSO	0	0.9	0	0.36	0.02	0.81
		BGR	0	0.92	0	0.95	0.03	0.81
	strong	DINAR	0	0	0	0.4	0.02	0.68
		SCAD	0	0.91	0	0.54	0.03	0.68
		LASSO	0	0.92	0	0.78	0.02	0.68
		BGR	0	0.92	0	0.95	0.04	0.69

Table 17: Simulation $N = 20$ with $T = 100, 500, 1000$ for low, medium, strong persistence for an increasing number of groups with increasing dimensionality and evaluated with AIC.

	Persistence	Models	FN.g	FP.g	FN.e	FP.e	RMSE para	RMSE out
100	low	DINAR	0.86	0	0.74	0	0.06	0.94
		SCAD	0.37	0.14	0.67	0.04	0.06	0.93
		LASSO	0.37	0.14	0.67	0.04	0.06	0.93
		BGR	0	0.67	0	0.8	0.07	0.94
	medium	DINAR	0.66	0	0.57	0.21	0.07	0.87
		SCAD	0.24	0.28	0.5	0.13	0.08	0.88
		LASSO	0.37	0.14	0.6	0.06	0.08	0.89
		BGR	0	0.67	0	0.8	0.07	0.84
	strong	DINAR	0.23	0	0.2	0.39	0.07	0.72
		SCAD	0.01	0.57	0.17	0.28	0.08	0.72
		LASSO	0.02	0.63	0.08	0.41	0.06	0.72
		BGR	0	0.67	0	0.8	0.08	0.71
500	low	DINAR	0	0	0	0.42	0.02	0.89
		SCAD	0	0.45	0.06	0.21	0.03	0.9
		LASSO	0	0.65	0.01	0.39	0.03	0.89
		BGR	0	0.67	0	0.8	0.03	0.89
	medium	DINAR	0	0	0	0.42	0.03	0.8
		SCAD	0	0.19	0.03	0.04	0.03	0.8
		LASSO	0	0.66	0	0.48	0.03	0.8
		BGR	0	0.67	0	0.8	0.04	0.8
	strong	DINAR	0	0	0	0.42	0.03	0.67
		SCAD	0	0.35	0.01	0.08	0.04	0.67
		LASSO	0	0.67	0	0.54	0.04	0.67
		BGR	0	0.67	0	0.8	0.04	0.67
1000	low	DINAR	0	0	0	0.42	0.02	0.89
		SCAD	0	0.51	0	0.14	0.02	0.89
		LASSO	0	0.61	0	0.27	0.02	0.89
		BGR	0	0.67	0	0.8	0.02	0.89
	medium	DINAR	0	0	0	0.42	0.02	0.8
		SCAD	0	0.06	0	0.01	0.02	0.8
		LASSO	0	0.66	0	0.47	0.02	0.8
		BGR	0	0.67	0	0.8	0.03	0.8
	strong	DINAR	0	0	0	0.42	0.03	0.67
		SCAD	0	0.13	0	0.02	0.03	0.67
		LASSO	0	0.67	0	0.57	0.03	0.67
		BGR	0	0.67	0	0.8	0.04	0.67

Table 18: Simulation $N = 60$ with $T = 100, 500, 1000$ for low, medium, strong persistence for an increasing number of groups with increasing dimensionality and evaluated with AIC.

	Persistence	Models	FN.g	FP.g	FN.e	FP.e	RMSE para	RMSE out
100	low	DINAR	0.91	0	0.77	0	0.04	0.98
		SCAD	0.27	0.47	0.7	0.12	0.04	0.98
		LASSO	0.26	0.48	0.7	0.13	0.04	0.97
		BGR	0	0.82	0	0.93	0.08	1.11
	medium	DINAR	0.9	0	0.75	0.03	0.05	0.96
		SCAD	0.36	0.33	0.7	0.06	0.05	0.96
		LASSO	0.42	0.27	0.71	0.05	0.05	0.96
		BGR	0	0.82	0	0.93	0.09	1.04
	strong	DINAR	0.8	0.01	0.67	0.3	0.06	0.91
		SCAD	0.29	0.44	0.56	0.17	0.06	0.9
		LASSO	0.54	0.21	0.7	0.05	0.07	0.94
		BGR	0	0.82	0	0.93	0.09	0.96
500	low	DINAR	0.7	0	0.59	0.34	0.03	0.96
		SCAD	0	0.66	0.14	0.15	0.02	0.94
		LASSO	0	0.71	0.1	0.19	0.02	0.94
		BGR	0	0.82	0	0.93	0.03	0.95
	medium	DINAR	0.01	0	0.01	0.61	0.02	0.88
		SCAD	0	0.65	0.11	0.16	0.02	0.88
		LASSO	0	0.81	0	0.43	0.02	0.88
		BGR	0	0.82	0	0.93	0.03	0.89
	strong	DINAR	0	0	0	0.61	0.02	0.81
		SCAD	0.01	0.76	0.02	0.37	0.02	0.81
		LASSO	0	0.81	0	0.53	0.03	0.81
		BGR	0	0.82	0	0.93	0.04	0.82
1000	low	DINAR	0	0	0	0.61	0.01	0.93
		SCAD	0	0.26	0.05	0.02	0.01	0.93
		LASSO	0	0.82	0	0.49	0.01	0.93
		BGR	0	0.82	0	0.93	0.02	0.94
	medium	DINAR	0	0	0	0.61	0.01	0.87
		SCAD	0	0.73	0	0.29	0.01	0.87
		LASSO	0	0.77	0	0.39	0.02	0.87
		BGR	0	0.82	0	0.93	0.02	0.88
	strong	DINAR	0	0	0	0.61	0.02	0.8
		SCAD	0	0.74	0	0.3	0.02	0.8
		LASSO	0	0.81	0	0.45	0.02	0.8
		BGR	0	0.82	0	0.93	0.03	0.81

I.4 Theoretical properties

We derive the asymptotic properties of the estimator (1). Recall that the matrix A is assumed to have a sparse 'network' structure. The residual term $\epsilon_t = (\epsilon_{1,t}, \cdots, \epsilon_{N,t})^\top$ is a vector that is assumed to be independently and identically distributed with $\epsilon_t \sim (0, \Sigma)$. For ease of notation, we denote $B = [\mathbf{1}\beta_0, (I_N\beta_1 + A)]$ as a $N \times (N+1)$ parameter matrix. To make the results wider applicable, we will derive the asymptotic properties by applying the regularization function upon the entire matrix B. In the model as it is applied in this study, λ is set to 0 for β_0 and β_1. Hence the derivations in this section cover the general regularization case and we denote $p_{\alpha\lambda,(1-\alpha)\lambda}(B) = \sum_{i=1}^{N+1}(1-\alpha)\lambda||B_{.i}||_F + \sum_{i=1}^{N}\sum_{j=1}^{N+1}\alpha\lambda|b_{ij}|$. In accordance with this study, we treat α as a fixed predefined value. We assume that the model is stationary and ergodic, with all roots of the polynomial $\boldsymbol{I}_N - BZ$ lying outside the unit ball. Note that N is fixed during the investigation of the asymptotic properties. We assume the following regularity conditions hold:

1. The observations Y_t for all t are i.i.d. with probability density $f(Y, B)$. It shall hold

$$\mathrm{E}\left[\frac{\partial log f(Y,B)}{\partial B_{ij}}\right] = 0 \text{ for all } i = 1, \ldots, N \text{ and } j = 1, \ldots, N+1$$

   and

$$I_{i_1j_1,i_2j_2} = \mathrm{E} - \frac{\partial^2 log f(Y,B)}{\partial B_{i_1j_1}\partial B_{i_2j_2}}$$

2. The Fisher Information matrix $I(B)$ is finite and positive definite at $B = \mathcal{B}$ with $\mathcal{B}$ the true parameter matrix.

3. There exists an open subset ω in the parameter space Ω of B that contains the true parameter matrix $\mathcal{B}$. For almost all Y_t the density $f(Y, B)$ admits all third derivatives $\frac{\partial^3 log f(Y,B)}{\partial B_{i_1j_1}\partial B_{i_2j_2}\partial B_{i_3j_3}}$ for all B in the open subset. There exist functions $M_{i_1j_1,i_2j_2,i_3j_3}$ such that

$$\left|\frac{\partial^3 log f(Y,B)}{\partial B_{i_1j_1}\partial B_{i_2j_2}\partial B_{i_3j_3}}\right| \leq M_{i_1j_1,i_2j_2,i_3j_3}(Y) \text{ for all } B \in \omega$$

   whereas $m_{i_1j_1,i_2j_2,i_3j_3} = \mathrm{E}\left[M_{i_1j_1,i_2j_2,i_3j_3}(Y)\right] < \infty$.

Note that T will go to infinity which impacts the values of $\alpha\lambda$ and $(1-\alpha)\lambda$, hence we denote $\alpha\lambda_T$ and $(1-\alpha)\lambda_T$ from here onwards. Denote $g_{max,T} = \max(\frac{\partial p_{\alpha\lambda_T,(1-\alpha)\lambda_T}}{\partial B_{ij}} : B_{ij} \neq 0)$, which is the maximal regularization applied to any B_{ij}. $g_{max,T}$ will only take on the value 0 if $\lambda_T \to 0$. Also in case of a dense system, $g_{max,T}$ would be 0 but this contradicts the assumption of this study of a sparse parameter matrix. We denote the true parameter matrix as $\mathcal{B}$. The proofs to the results are given in detail in Appendix I.5.

Theorem 1. *Assume that the assumptions for model (2) hold. If* $\max\{\frac{\partial^2 p_{\alpha\lambda_T,(1-\alpha)\lambda_T}}{\partial B_{ij}^2(B_{ij})} : B_{ij} \neq 0\} \to 0$*, then there exists a local maximizer* $\widehat{B}$ *for (3) such that* $||\widehat{B} - \mathcal{B}||_F = \mathcal{O}_p(T^{-1/2} + g_{max,T})$.

When the hypotheses of Theorem 1 are fulfilled, a proper choice of the regularization parameters $\alpha\lambda$ and $(1-\alpha)\lambda$ ensures the existence of a local maximizer of (3), which converges at speed $\sqrt{T}$. If $\alpha\lambda_T, (1-\alpha)\lambda_T \to 0$, then the estimator is root-T consistent.

Next, we show that the estimator possesses the sparsity property and hence is capable of selecting the model parameters in a sparse system. Denote by $CL(\cdot)$ the constrained likelihood. In what follows, assume without loss of generality that the true parameter matrix $\mathcal{B}$ contains a submatrix of dimension $N_1 \times N_1$ whose elements are different from 0 in the upper left corner. The remaining elements are equal to 0. Let $B_{N_1N_1}$ indicate the respective submatrix and $B_{-N_1-N_1}$ the remaining elements of the respective matrix.

Lemma 1. *Assume that the assumptions for model (2) hold. If* $\alpha\lambda_T, (1-\alpha)\lambda_T \to 0$ *and* $\sqrt{T}\alpha\lambda_T, \sqrt{T}(1-\alpha)\lambda_T \to \infty$ *as* $T \to \infty$*, then with probability tending to 1, for any given* $B_{N_1N_1}$ *satisfying* $||B_{N_1N_1} - \mathcal{B}_{N_1N_1}||_F = O_p(T^{-1/2})$ *and any constant* Q*,*

$$CL(B_{N_1N_1}, 0) = \max_{||B_{-N_1-N_1}||_F \leq QT^{-1/2}} CL(B_{N_1N_1}, B_{-N_1-N_1}),$$

hence

$$P(B_{-N_1-N_1} = 0) \to 1.$$

Finally, we show that the estimator possesses the oracle property, i.e., it chooses the true model as if it were a theoretical estimator that knows the true model structure.

We define

$$F = \left[p''_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{11}), \cdots, p''_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{N_1N_1})\right]$$

as a $N_1 \times N_1$ symmetric matrix containing the second derivatives of the penalty function and

$$G = \left[p'_{\alpha\lambda_T,(1-\alpha)\lambda_T}(B_{11})sgn(B_{11}), \cdots, p'_{\alpha\lambda_T,(1-\alpha)\lambda_T}(B_{N_1N_1})sgn(B_{N_1N_1})\right]$$

as a $N_1 \times N_1$ matrix containing the first derivatives of the penalty function.

Theorem 2. *Assume that the assumptions for model (2) hold. If* $\alpha\lambda_T, (1-\alpha)\lambda_T \to 0$ *and* $\sqrt{T}\alpha\lambda_T, \sqrt{T}(1-\alpha)\lambda_T \to \infty$ *as* $T \to \infty$*, then with probability tending to 1, the root-T consistent local maximizer* $B = [B_{N_1N_1}, B_{-N_1-N_1}]$ *from Theorem 1 must satisfy*

1. *Sparsity:* $B_{-N_1-N_1} = \mathbf{0}$

2. Asymptotic normality:

$$\sqrt{T}((B_{N_1N_1} - \mathcal{B}_{N_1N_1})(I(\mathcal{B}_{N_1N_1}) + F) + G) \xrightarrow{d} N(0, I(\mathcal{B}_{N_1N_1})) \tag{10}$$

in distribution, where $I(\mathcal{B}_{N_1N_1})$ is the Fisher Information knowing that $\mathcal{B}_{-N_1-N_1} = \mathbf{0}$.

I.5 Proofs

In this section we prove the consistency and oracle property of the estimator. We prove the theorems under the assumptions made for model (2), stated in Appendix I.4. The proofs follow Fan and Li (2001), Song and Bickel (2011) and Wang et al. (2007).

We further define $O_M(\cdot)$ as big O notation for elementwise convergence within a matrix and $O_V(\cdot)$ as big O notation for elementwise convergence within a vector. Likewise we define $o_M(\cdot)$ and $o_V(\cdot)$ as small o notation for matrices and vectors. Let $vec(\cdot)$ denote the vectorizing operator to convert a matrix to a vector. Further, we denote the Fisher information matrix by $I(\cdot)$, which is assumed to be finite and positive definite.

I.6 Proof of Theorem 1

Denote by $CL(\cdot)$ the constrained likelihood and by $L(\cdot)$ the likelihood. Define $CL(B) = L(B) - T\sum_{i=1}^{N}\sum_{j=1}^{N} p_{\alpha\lambda_T,(1-\alpha)\lambda_T}(B_{ij})$. Further define $g_T = T^{-1/2} + g_{max,T}$ and $\boldsymbol{U}$ coordinates around $\mathcal{B}$. For a large constant Q, it holds that $\{\mathcal{B} + g_T\boldsymbol{U} : ||\boldsymbol{U}||_F \leq Q\}$ is the ball around $\mathcal{B}$ and we intend to show that a local maximum with maximizer $\widehat{B}$ lies in the ball. So we intend to show that on the surface of the ball, $||\boldsymbol{U}||_F = Q$, for any $\epsilon > 0$, there exists a large constant Q such that

$$P\{\sup_{||\boldsymbol{U}||_F=Q} CL(\mathcal{B} + g_T\boldsymbol{U}) < CL(\mathcal{B})\} \geq 1 - \epsilon. \tag{11}$$

The difference between the two penalized likelihoods $CL(\mathcal{B} + g_T\boldsymbol{U})$ and $CL(\mathcal{B})$ can be bounded from above by the likelihood and the penalization on $\widehat{B}$ only for the N_1^2 parameters different from 0. For the construction of the upper bound, we make use of the property $p_{\alpha\lambda_T,(1-\alpha)\lambda_T}(0) = 0$, which holds for the $((N - N_1)^2 - (N - N_1))$ parameters which are 0. In case no parameter in $\mathcal{B}$ is 0, it will be equal, otherwise larger:

$$CL(\mathcal{B}+g_T\boldsymbol{U})-CL(\mathcal{B}) \leq L(\mathcal{B}+g_T\boldsymbol{U})-L(\mathcal{B}) \tag{12}$$
$$-T\sum_{i=1}^{N_1}\sum_{j=1}^{N_1}\{p_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{ij}+g_T\boldsymbol{U}_{ij})-p_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{ij})\}.$$

Approximating by a Taylor expansion for $\mathcal{B}+g_T\boldsymbol{U}$ around $\mathcal{B}$ gives

$$\begin{aligned} L(\mathcal{B}+g_T\boldsymbol{U}) &= L(\mathcal{B})+(\mathcal{B}+g_T\boldsymbol{U}-\mathcal{B})L^{'}(\mathcal{B}) \\ &+\frac{1}{2}L^{''}(\mathcal{B})(\mathcal{B}+g_T\boldsymbol{U}-\mathcal{B})^{\top}(\mathcal{B}+g_T\boldsymbol{U}-\mathcal{B}) \\ &+o_p\{\frac{1}{2}L^{''}(\mathcal{B})(\mathcal{B}+g_T\boldsymbol{U}-\mathcal{B})^{\top}(\mathcal{B}+g_T\boldsymbol{U}-\mathcal{B})\} \end{aligned} \tag{13}$$

which leads to

$$\begin{aligned} L(\mathcal{B}+g_T\boldsymbol{U})-L(\mathcal{B}) &= g_TL^{'}(\mathcal{B})\mathrm{vec}(\boldsymbol{U})+\frac{1}{2}g_T^2\mathrm{vec}(\boldsymbol{U})^{\top}L^{''}(\mathcal{B})\mathrm{vec}(\boldsymbol{U}) \\ &+\frac{1}{2}g_T^2\mathrm{vec}(\boldsymbol{U})^{\top}L^{''}(\mathcal{B})\mathrm{vec}(\boldsymbol{U})o_p\{1\} \end{aligned} \tag{14}$$

Also,

$$\begin{aligned} T\sum_{i,j=1}^{N_1}p_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{ij}+g_T\boldsymbol{U}_{ij}) &= T\sum_{i,j=1}^{N_1}p_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{ij}) \\ &+T\sum_{i,j=1}^{N_1}(\mathcal{B}_{ij}+g_T\boldsymbol{U}_{ij}-\mathcal{B}_{ij}) \\ &\quad p_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{ij})^{'}sgn(\mathcal{B}_{ij}) \\ &+T\sum_{i,j=1}^{N_1}(\mathcal{B}_{ij}+g_T\boldsymbol{U}_{ij}-\mathcal{B}_{ij})^2 \\ &\quad p_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{ij})^{''} \\ &+T\sum_{i,j=1}^{N_1}o_p(\mathcal{B}_{ij}+g_T\boldsymbol{U}_{ij}-\mathcal{B}_{ij})^2 \\ &\quad p_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{ij})^{''} \end{aligned} \tag{15}$$

which leads to

$$T\sum_{i,j=1}^{N_1} p_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{ij}+g_T\boldsymbol{U}_{ij}) - T\sum_{i,j=1}^{N_1} p_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{ij}) \tag{16}$$

$$= T\sum_{i,j=1}^{N_1} (g_T\boldsymbol{U}_{ij}) p_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{ij})' sgn(\mathcal{B}_{ij}) \tag{17}$$

$$+ T\sum_{i,j=1}^{N_1} (g_T\boldsymbol{U}_{ij})^2 p_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{ij})''$$

$$+ T\sum_{i,j=1}^{N_1} g_T^2\boldsymbol{U}_{ij}^2 o_p(1)^2 p_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{ij})''$$

Recall that $L''(\mathcal{B}) = -TI(\mathcal{B})$.

Hence,

$$\begin{aligned} CL(\mathcal{B}+g_T\boldsymbol{U}) - CL(\mathcal{B}) \leq & bL'(\mathcal{B})^\top \mathrm{vec}(\boldsymbol{U}) \\ & - \frac{1}{2} T g_T^2 \mathrm{vec}(\boldsymbol{U})^\top I(\mathcal{B}) \mathrm{vec}(\boldsymbol{U})(1+o_p(1)) \\ & - T\sum_{i,j=1}^{N_1} g_T p_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{ij})' sgn(\mathcal{B}_{ij})\boldsymbol{U}_{ij} \\ & - T\sum_{i,j=1}^{N_1} (g_T\boldsymbol{U}_{ij})^2 p_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{ij})''(1+o_p(1)) \end{aligned} \tag{18}$$

If the right-hand side of the inequality is smaller 0, the inequality holds. Note that it holds $T^{-1/2}L(\mathcal{B})' = \mathcal{O}_V(1)$. It follows that the first term on the right-hand side is of order $\mathcal{O}_V(T^{1/2}g_T)$. The second term is of order $\mathcal{O}_p(Tg_T^2)$, and it holds $\mathcal{O}_p(T^{1/2}g_T) = \mathcal{O}_p(Tg_T^2)$. For a sufficiently large Q, the second term dominates the first term uniformly in $||\boldsymbol{U}||_F = Q$. The third and fourth term are bounded by

$$Tg_T\sum_{i,j=1}^{N_1} g_{max,T}\boldsymbol{U}_{ij} + Tg_T^2\sum_{i,j=1}^{N_1}\boldsymbol{U}_{ij}^2 \max\left(\frac{\partial^2 p_{\alpha\lambda_T,(1-\alpha)\lambda_T}}{\partial B_{ij}^2} : B_{ij}\neq 0\right)(1+o_p(1)), \tag{19}$$

and therefore are $\mathcal{O}_p(Tg_T)$ and $\mathcal{O}_p(Tg_T^2)$. Since $\max(\frac{\partial^2 p_{\alpha\lambda_T,(1-\alpha)\lambda_T}}{\partial B_{ij}^2} : B_{ij}\neq 0) \to 0$ and the term is of $\mathcal{O}_p(Tg_T^2)$, it is dominated by the second term in case of a large Q. Also the third term is dominated by the second term since it is of order $\mathcal{O}_p(Tg_Tg_{max,T})$ which is dominated by $\mathcal{O}_p(Tg_T^2)$ and Q takes on a larger effect in the second term. Therefore, the negativity of the second term ensures the right-hand side to be smaller 0 in case of a large Q. Hence (11) holds. This implies that there exists a local maximizer $\widehat{B}$ for which $||\widehat{B}-\mathcal{B}||_F = \mathcal{O}_p(g_T)$. This completes the proof of the theorem.

I.7 Proof of Lemma 1

We carry out the proof by showing that all parameters in $B_{-N_1-N_1}$ cannot be different from 0 since this would be a contradiction. One has

$$\frac{\partial CL(\widehat{B}_k)}{\partial B_{ij}} = \frac{\partial L(\widehat{B})}{\partial B_{ij}} - Tp'_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\widehat{B}_{ij})sgn(\widehat{B}_{ij}), \tag{20}$$

hence for a consistent selection of $B_{-N_1-N_1}$ all parameters have to be 0. Otherwise the first derivative of the constrained likelihood would not equal the unconstrained one, which is 0.

It is sufficient to show that $\frac{\partial CL(B)}{\partial B_{ij}} \neq 0$ if and only if $B_{ij} \neq 0$. Hence we will show that with probability tending to 1 for $T \to \infty$, for any $B_{N_1N_1}$ satisfying $B_{N_1N_1} - \mathcal{B}_{N_1N_1} = O_M(T^{-1/2})$ and for some small $\epsilon_T = QT^{-1/2}$ and $i, j = 1, \ldots, N_1$,

$$\frac{\partial CL(B)}{\partial B_{ij}} < 0 \quad \text{for } 0 < B_{ij} < \epsilon_T \tag{21}$$

$$> 0 \quad \text{for } -\epsilon_T < B_{ij} < 0 \tag{22}$$

By Taylor's expansion,

$$\frac{\partial CL(\widehat{B})}{\partial B_{ij}} = \frac{\partial L(\widehat{B}_k)}{\partial B_{ij}} - Tp'_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\widehat{B}_{ij})sgn(\widehat{B}_{ij}) \tag{23}$$

$$= \frac{\partial L(\mathcal{B})}{\partial B_{ij}} + \sum_{l_1=1}^{N_2}\sum_{l_2=1}^{N_2} \frac{\partial^2 L(\mathcal{B})}{\partial B_{ij}\partial B_{l_1l_2}}(\widehat{B}_{l_1l_2} - \mathcal{B}_{l_1l_2}) \tag{24}$$

$$+ \sum_{l_1=1}^{N_2}\sum_{l_2=1}^{N_2}\sum_{l_3=1}^{N_2}\sum_{l_4=1}^{N_2} \frac{\partial^3 L(B^*)}{\partial B_{ij}\partial B_{l_1l_2}\partial B_{l_3l_4}} \times (\widehat{B}_{l_1l_2} - \mathcal{B}_{l_1l_2})(\widehat{B}_{l_3l_4} - \mathcal{B}_{l_3l_4})$$

$$- Tp'_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\widehat{B}_{ij})sgn(\widehat{B}_{ij})$$

with B^* lying between $\widehat{B}$ and $\mathcal{B}$.

Recall that

$$T^{-1}\frac{\partial L(\mathcal{B})}{\partial B_{ij}} = O(T^{-1/2})$$

$$T^{-1}\frac{\partial^2 L(\mathcal{B})}{\partial B_{ij}\partial B_{l_1l_2}} = \mathrm{E}\left(\frac{\partial^2 L(\mathcal{B})}{\partial B_{ij}\partial B_{l_1l_2}}\right) + o(1)$$

The first term is therefore of order $O(T^{1/2})$. The second term is also of order $O(T^{1/2})$ because it consists of the Fisher information matrix and $o(1)$, where the latter is negligible because o goes to 0 faster than O. The third term is obviously faster at 0 due to the squared O_M, meaning it is bounded by $O_P(T^{-1/2})^2$, hence it goes faster to 0 than the first and second term. It follows that

$$\frac{\partial CL(B)}{\partial B_{ij}} = -Tp'_{\alpha\lambda_T,(1-\alpha)\lambda_T}(B_{ij})sgn(B_{ij}) + O(T^{1/2}) \tag{25}$$

The first term dominates, because $\sqrt{T}g_{max,T} \to \infty$. Hence the sign of B_{ij} determines the sign of $\frac{\partial CL(B)}{\partial B_{ij}}$. Hence the inequalities (21) and (22) hold, which implies that $\frac{\partial CL(B)}{\partial B_{ij}}$ can only be 0 if and only if $B_{ij} = 0$. This completes the proof.

I.8 Proof of Theorem 2

From Lemma 1 there follows 1. It can be easily shown that there exists an $\widehat{B}_{N_1N_1}$ in Theorem 1 that is a root-T consistent local maximizer of $CL((B_{N_1N_1}, \mathbf{0}))$ that satisfies the likelihood equations

$$\left.\frac{\partial CL(B)}{\partial B_{ij}}\right|_{B=[B_{N_1N_1},B_{-N_1-N_1}]} = 0 \qquad \text{for } i = 1,\cdots,N_1; j = 1,\cdots,N_1 \tag{26}$$

Recall that $B_{N_1N_1}$ is a consistent estimator,

$$\begin{aligned}
&\left.\frac{\partial L(B)}{\partial B_{ij}}\right|_{B=[B_{N_1N_1},B_{-N_1-N_1}]} - Tp'_{\alpha\lambda_T,(1-\alpha)\lambda_T}(B_{ij})sgn(B_{ij}) && (27)\\
&= \frac{\partial L(\mathcal{B})}{\partial B_{ij}} + \sum_{l_1=1}^{N_1}\sum_{l_2=1}^{N_1}\left(\frac{\partial^2 L(\mathcal{B})}{\partial B_{ij}\partial B_{l_1l_2}} + o_P(1)\right)(B_{ij} - \mathcal{B}_{ij}) && (28)\\
&- T\Big(p'_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{ij})sgn(\mathcal{B}_{ij})\\
&+(p''_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{ij}) + o_P(1))(B_{ij} - \mathcal{B}_{ij})\Big).
\end{aligned}$$

Setting the first derivative equal to 0 and rearranging terms gives

$$\begin{aligned}
(B_{ij} - \mathcal{B}_{ij}) &= -\frac{\frac{\partial L(\mathcal{B})}{\partial B_{ij}} - Tp'_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{ij})sgn(\mathcal{B}_{ij})}{H - TK}\\
&= -\frac{\frac{1}{T}\frac{\partial L(\mathcal{B})}{\partial B_{ij}} - p'_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{ij})sgn(\mathcal{B}_{ij})}{\frac{1}{T}H - K},
\end{aligned}$$

whereas $H = \sum_{l_1=1}^{N_1} \sum_{l_2=1}^{N_1} \left(\frac{\partial^2 L(\mathcal{B})}{\partial B_{ij} \partial B_{l_1 l_2}} + o_P(1) \right)$ and $K = p''_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{ij}) + o_P(1)$.

The nominator converges in distribution by the Central Limit Theorem to

$$\frac{1}{T} \frac{\partial L(\mathcal{B})}{\partial B_{ij}} - p'_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{ij}) sgn(\mathcal{B}_{ij}) \xrightarrow{d} N(0, \frac{I(\mathcal{B}_{N_1 N_1})_{ij}}{T}) - G_{ij} \tag{29}$$

By Slutsky's Theorem, the denominator goes to

$$\begin{aligned} &- \frac{1}{T} \sum_{l_1=1}^{N_1} \sum_{l_2=1}^{N_1} \left(\frac{\partial^2 L(\mathcal{B})}{\partial B_{ij} \partial B_{l_1 l_2}} + o_P(1) \right) \\ &+ (p''_{\alpha\lambda_T,(1-\alpha)\lambda_T}(\mathcal{B}_{ij}) + o_P(1)) \to I(\mathcal{B}_{N_1 N_1})_{ij} + F_{ij} \end{aligned} \tag{30}$$

Combining the two results and writing this in matrix form gives

$$\begin{aligned} (B_{N_1 N_1} - \mathcal{B}_{N_1 N_1}) &\xrightarrow{d} N(0, \frac{I(\mathcal{B}_{N_1 N_1})}{T} (I(\mathcal{B}_{N_1 N_1}) + F)^{-2}) - G(I(\mathcal{B}_{N_1 N_1}) + F)^{-1} \\ (B_{N_1 N_1} - \mathcal{B}_{N_1 N_1}) + G(I(\mathcal{B}_{N_1 N_1}) + F)^{-1} &\xrightarrow{d} N(0, \frac{I(\mathcal{B}_{N_1 N_1})}{T} (I(\mathcal{B}_{N_1 N_1}) + F)^{-2}) \\ \sqrt{T}((B_{N_1 N_1} - \mathcal{B}_{N_1 N_1})(I(\mathcal{B}_{N_1 N_1}) + F) + G) &\xrightarrow{d} N(0, I(\mathcal{B}_{N_1 N_1})) \end{aligned}$$

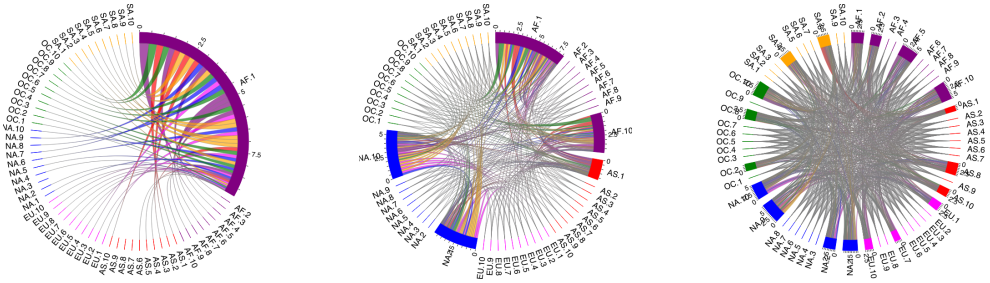
Hence by applying Slutsky's Theorem and the Central Limit Theorem, we find

$$\sqrt{T}((B_{N_1 N_1} - \mathcal{B}_{N_1 N_1})(I(\mathcal{B}_{N_1 N_1}) + F) + G) \xrightarrow{d} N(0, I(\mathcal{B}_{N_1 N_1})) \tag{31}$$

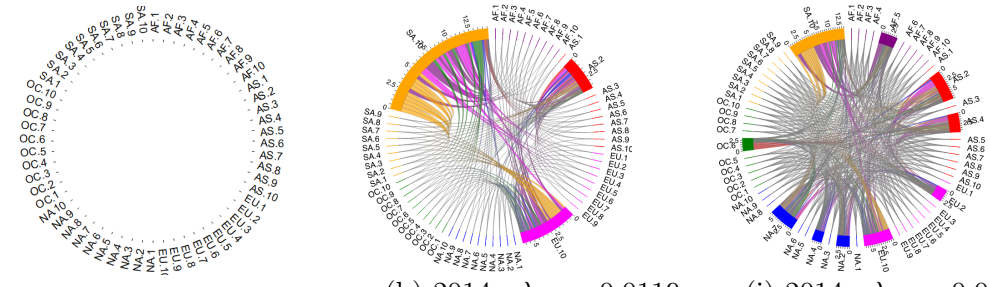
This completes the proof.

Figure 12: Adjacency matrices for 3 λ-values in analysis with 10 groups and evaluation period length of 180 days based upon BIC.

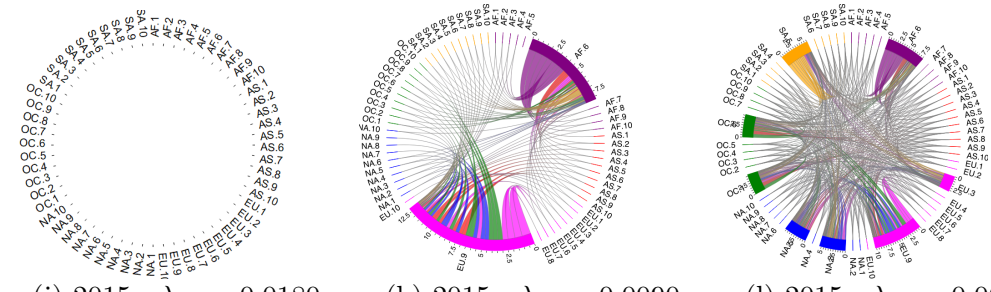
(a) 2012 $\alpha\lambda_{14} = 0.0184$ (b) 2012 $\alpha\lambda_{15} = 0.0092$ (c) 2012 $\alpha\lambda_{16} = 0.0046$



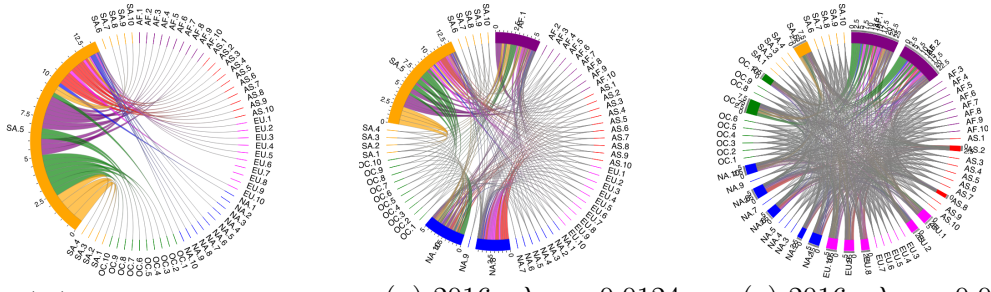
(d) 2013 $\alpha\lambda_{14} = 0.0208$ (e) 2013 $\alpha\lambda_{15} = 0.0104$ (f) 2013 $\alpha\lambda_{16} = 0.0052$



(g) 2014 $\alpha\lambda_{14} = 0.0238$ (h) 2014 $\alpha\lambda_{15} = 0.0119$ (i) 2014 $\alpha\lambda_{16} = 0.0060$



(j) 2015 $\alpha\lambda_{14} = 0.0180$ (k) 2015 $\alpha\lambda_{15} = 0.0090$ (l) 2015 $\alpha\lambda_{16} = 0.0045$



(m) 2016 $\alpha\lambda_{14} = 0.0249$ (n) 2016 $\alpha\lambda_{15} = 0.0124$ (o) 2016 $\alpha\lambda_{16} = 0.0062$

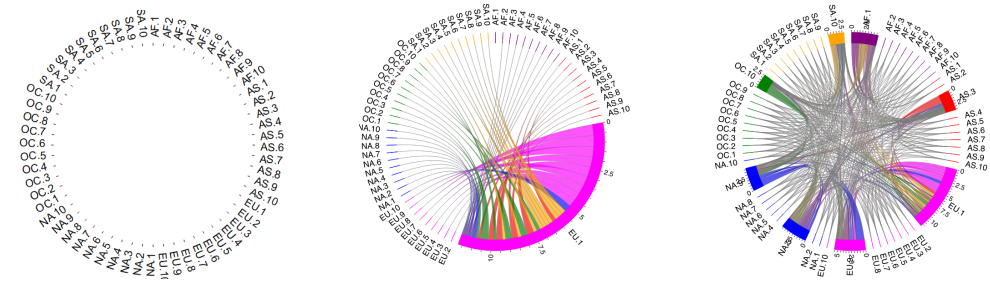
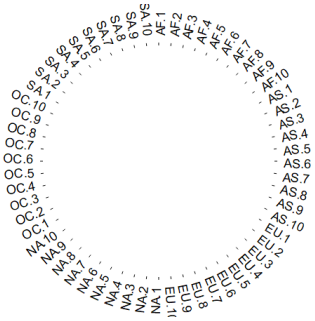
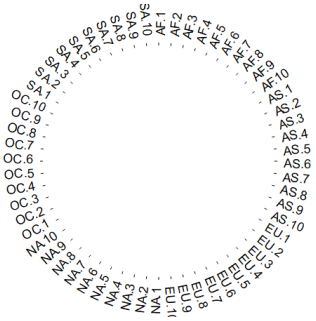


Figure 13: Adjacency matrices for 3 λ-values in analysis with 10 groups and evaluation period length of 180 days based upon BIC.

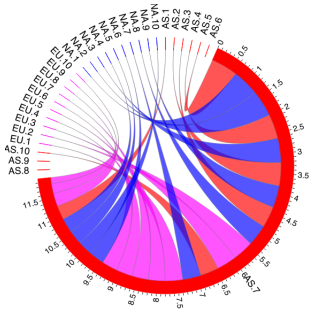
(a) 2017 $\alpha\lambda_{14} = 0.0327$



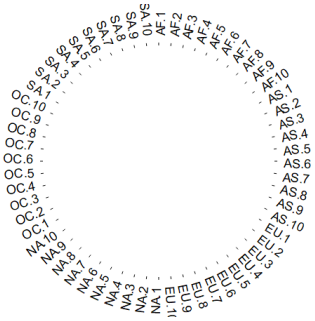
(b) 2017 $\alpha\lambda_{15} = 0.0163$



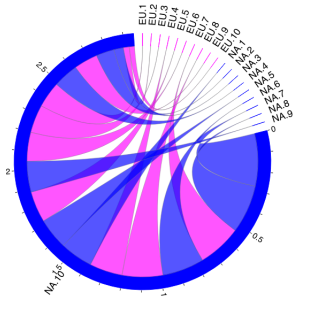
(c) 2017 $\alpha\lambda_{16} = 0.0082$



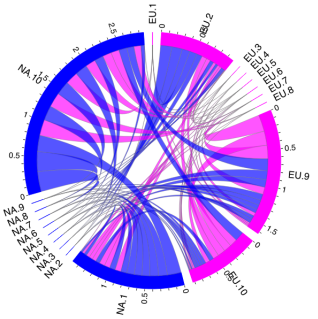
(d) 2018 $\alpha\lambda_{14} = 0.0264$



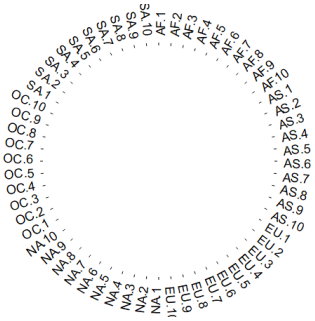
(e) 2018 $\alpha\lambda_{15} = 0.0132$



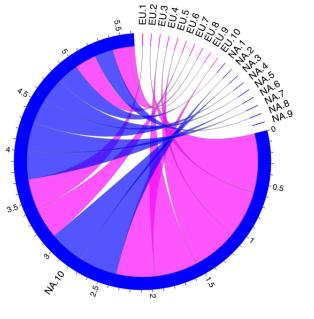
(f) 2018 $\alpha\lambda_{16} = 0.0066$



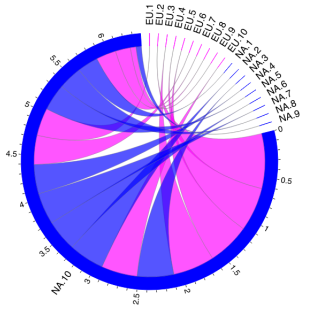
(g) 2019 $\alpha\lambda_{14} = 0.0316$



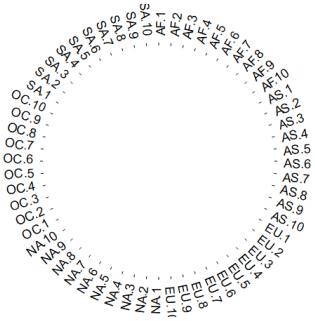
(h) 2019 $\alpha\lambda_{15} = 0.0158$



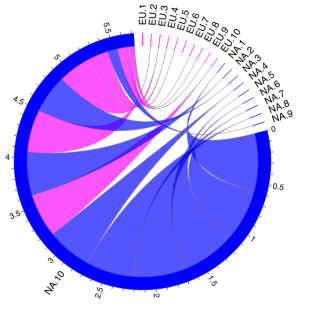
(i) 2019 $\alpha\lambda_{16} = 0.0079$



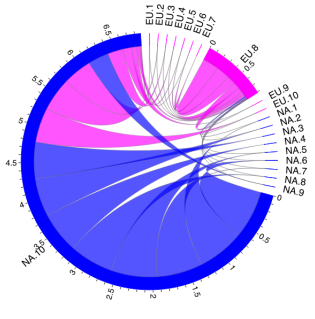
(j) 2020 $\alpha\lambda_{14} = 0.0319$



(k) 2020 $\alpha\lambda_{15} = 0.0159$



(l) 2020 $\alpha\lambda_{16} = 0.0080$



I.9 Additional tables

Table 19: AIC, HQ and BIC evaluation criteria for a VAR model selection procedure derived on the yearly data.

	Selected Lag	Lag 1	Lag 2	Lag 3
AIC	3.00	-160.54	-158.81	-164.97
HQ	1.00	-143.02	-123.77	-112.41
BIC	1.00	-116.73	-71.20	-33.55
AIC	3.00	-191.41	-188.76	-192.03
HQ	1.00	-176.02	-157.99	-145.87
BIC	1.00	-152.70	-111.36	-75.92
AIC	3.00	-179.11	-177.92	-182.62
HQ	1.00	-163.48	-146.68	-135.76
BIC	1.00	-139.84	-99.39	-64.82
AIC	3.00	-214.00	-215.86	-235.79
HQ	1.00	-194.84	-177.54	-178.31
BIC	1.00	-166.28	-120.41	-92.62
AIC	3.00	-213.13	-211.23	-215.20
HQ	1.00	-197.44	-179.85	-168.12
BIC	1.00	-173.70	-132.37	-96.90
AIC	3.00	-212.87	-214.18	-220.66
HQ	1.00	-197.48	-183.41	-174.51
BIC	1.00	-174.17	-136.78	-104.56
AIC	1.00	-223.76	-218.99	-219.49
HQ	1.00	-208.37	-188.22	-173.33
BIC	1.00	-185.05	-141.59	-103.38
AIC	1.00	-233.69	-229.96	-230.28
HQ	1.00	-218.34	-199.25	-184.23
BIC	1.00	-195.07	-152.71	-114.42

Table 20: BIC and AIC for out-of-sample performance and the number of parameters and identified groups for 3-grouping dataset whereas a group is counted as active if 25% of the parameters are different from 0. The results are displayed for the models found in the respective years for DINAR, LASSO, SCAD and BGR. The first 4 columns show the results evaluated on 90 days out-of-sample, the second 4 columns the results evaluated on 180 days out-of-sample.

		90 days				180 days			
		BIC	AIC	para	groups	BIC	AIC	para	groups
DINAR	2012	-1262	-1345	35	1	-2319	-2425	35	1
	2013	-1924	-1969	18	0	-2510	-2568	18	0
	2014	-1895	-1940	18	0	-4038	-4195	52	2
	2015	-1472	-1517	18	0	-2989	-3046	18	0
	2016	-3182	-3227	18	0	-5482	-5539	18	0
	2017	-1269	-1374	43	2	-1930	-2064	43	2
	2018	-1178	-1235	23	1	-2163	-2236	23	1
	2019	-1361	-1404	17	1	-2342	-2397	17	1
	2020	-1413	-1453	17	1	-2620	-2672	17	1
LASSO	2012	-1235	-1280	18	0	-2219	-2277	18	0
	2013	-1924	-1969	18	0	-2510	-2568	18	0
	2014	-1895	-1940	18	0	-3995	-4052	18	0
	2015	-1472	-1517	18	0	-2989	-3046	18	0
	2016	-3182	-3227	18	0	-5482	-5539	18	0
	2017	-1235	-1273	15	0	-1673	-1721	15	0
	2018	-1155	-1182	12	0	-2094	-2148	17	0
	2019	-1271	-1331	24	2	-2211	-2333	38	4
	2020	-1267	-1290	9	0	-2425	-2584	50	7
SCAD	2012	-1261	-1341	32	1	-2219	-2277	18	0
	2013	-1924	-1969	18	0	-2510	-2568	18	0
	2014	-1895	-1940	18	0	-4013	-4118	33	1
	2015	-1472	-1517	18	0	-2989	-3046	18	0
	2016	-3224	-3284	24	0	-5638	-5714	24	0
	2017	-1235	-1273	15	0	-1673	-1721	15	0
	2018	-1165	-1212	19	2	-2164	-2225	19	2
	2019	-1329	-1402	29	3	-2322	-2415	29	3
	2020	-1407	-1447	16	1	-2596	-2685	28	3
BGR	2012	62	-745	324	18	-739	-1770	324	18
	2013	-648	-1458	324	18	-1067	-2102	324	18
	2014	-667	-1474	324	18	-2706	-3737	324	18
	2015	-166	-976	324	18	-1490	-2525	324	18
	2016	-1784	-2594	324	18	-3883	-4918	324	18
	2017	-482	-1044	225	15	-1031	-1749	225	15
	2018	-584	-944	144	12	-1553	-2013	144	12
	2019	-802	-1005	81	9	-1548	-1807	81	9
	2020	-961	-1163	81	9	-1859	-2118	81	9

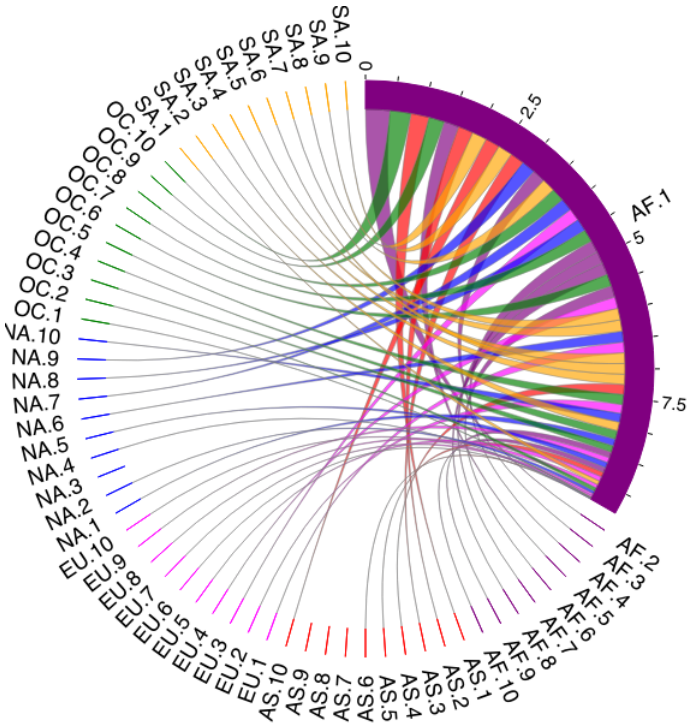
Table 21: BIC and AIC for out-of-sample performance and the number of parameters and identified groups for 3-grouping dataset whereas a group is counted as active if 25% of the parameters are different from 0. The results are displayed for the models found in the respective years for DINAR, LASSO, SCAD and BGR. The first 4 columns show the results evaluated on 270 days out-of-sample, the second 4 columns the results evaluated on the entire following year out-of-sample.

		270 days				one year			
		BIC	AIC	para	groups	BIC	AIC	para	groups
DINAR	2012	-3748	-3866	35	1	-5291	-5420	35	1
	2013	-3673	-3738	18	0	-4504	-4574	18	0
	2014	-4347	-4412	18	0	-4420	-4485	18	0
	2015	-4350	-4415	18	0	-5841	-5911	18	0
	2016	-12378	-12443	18	0	-13678	-13748	18	0
	2017	-2672	-2823	43	2	-3554	-3718	43	2
	2018	-3414	-3497	23	1	-4913	-5003	23	1
	2019	-3357	-3418	17	1	-4420	-4486	17	1
	2020	-3752	-3809	17	1	-5205	-5267	17	1
LASSO	2012	-3580	-3645	18	0	-5064	-5134	18	0
	2013	-3672	-3737	18	0	-4504	-4573	18	0
	2014	-4348	-4412	18	0	-4420	-4485	18	0
	2015	-4350	-4414	18	0	-5841	-5910	18	0
	2016	-12378	-12443	18	0	-13677	-13748	18	0
	2017	-2287	-2636	97	10	-3069	-3447	97	10
	2018	-3287	-3348	17	0	-4774	-4989	56	10
	2019	-3232	-3419	52	7	-4344	-4547	52	7
	2020	-3574	-3790	60	9	-5054	-5288	60	9
SCAD	2012	-3656	-3771	32	1	-5149	-5274	32	1
	2013	-3672	-3737	18	0	-4550	-4651	26	1
	2014	-4348	-4412	18	0	-4420	-4485	18	0
	2015	-4350	-4414	18	0	-5841	-5910	18	0
	2016	-12488	-12574	24	0	-13738	-13831	24	0
	2017	-2355	-2513	44	4	-3098	-3269	44	4
	2018	-3469	-3588	33	3	-4986	-5114	33	3
	2019	-3424	-3529	29	3	-4601	-4715	29	3
	2020	-3798	-3898	28	3	-5320	-5429	28	3
BGR	2012	-2050	-3212	324	18	-3485	-4744	324	18
	2013	-2155	-3321	324	18	-3024	-4281	324	18
	2014	-2884	-4046	324	18	-2947	-4114	324	18
	2015	-2815	-3981	324	18	-4225	-5480	324	18
	2016	-10622	-11784	324	18	-11750	-13009	324	18
	2017	-1708	-2518	225	15	-2427	-3303	225	15
	2018	-2850	-3368	144	12	-4364	-4926	144	12
	2019	-2377	-2669	81	9	-3266	-3582	81	9
	2020	-2814	-3101	81	9	-4015	-4327	81	9

I.10 Chord diagrams

Figure 14: Adjacency matrices and serial dependence parameter in analysis with 10 groups and evaluation period length of 180 days.

(a) Adjacency matrices for the 10 groups in 2012



(b) Adjacency matrices for the 10 groups in 2014

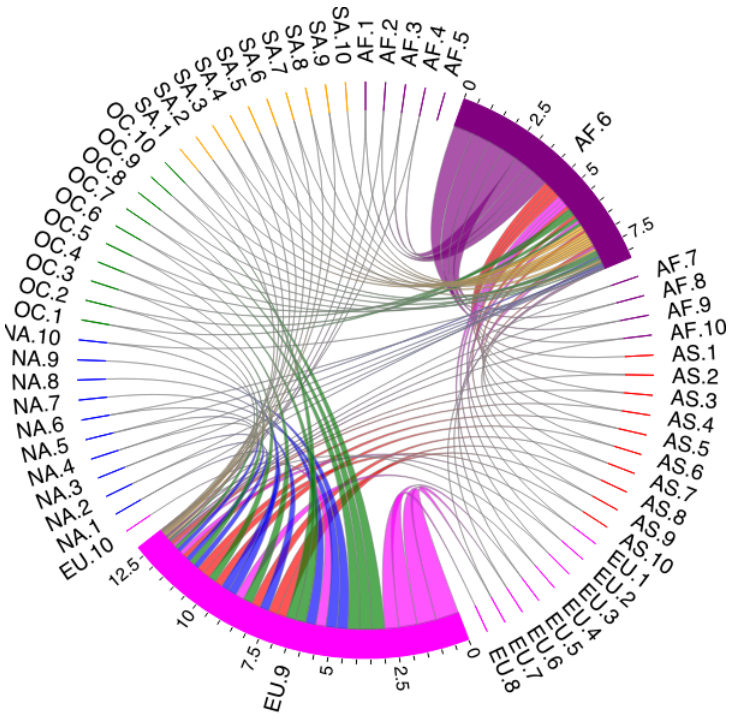
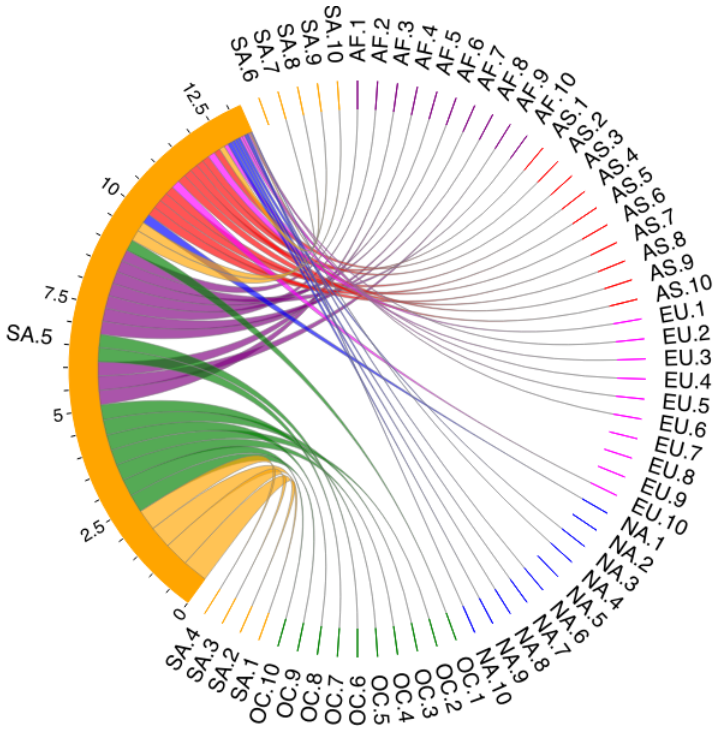


Figure 15: Adjacency matrices and serial dependence parameter in analysis with 10 groups and evaluation period length of 180 days.

(a) Adjacency matrices for the 10 groups in 2015



(b) Adjacency matrices for the 10 groups in 2016

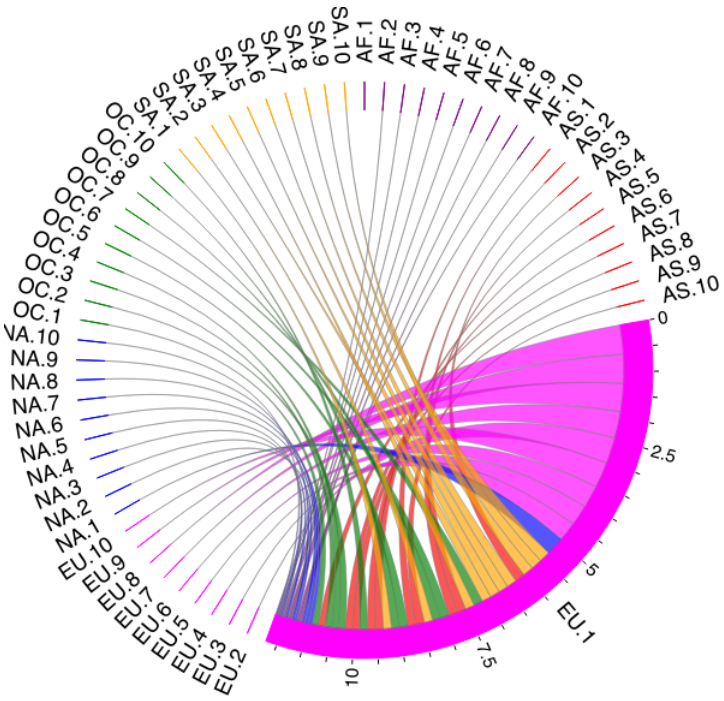
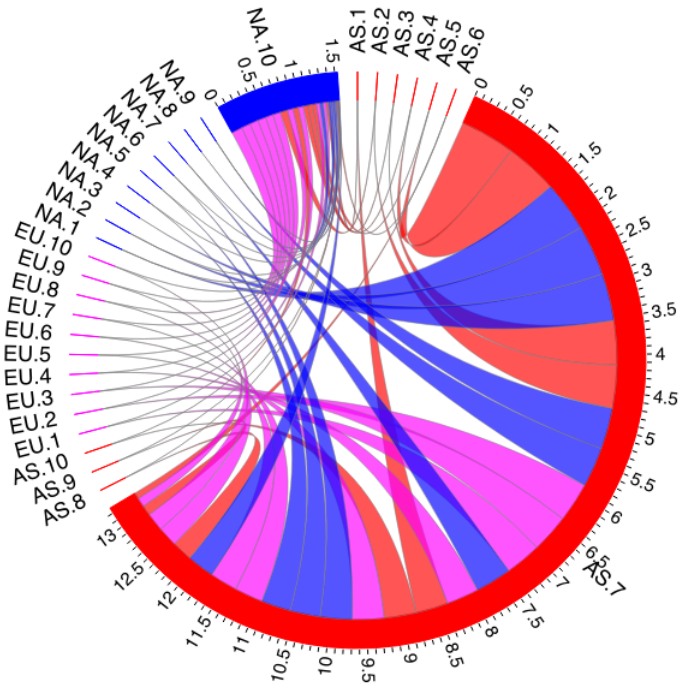


Figure 16: Adjacency matrices and serial dependence parameter in analysis with 10 groups and evaluation period length of 180 days.

(a) Adjacency matrices for the 10 groups in 2017



(b) Adjacency matrices for the 10 groups in 2018

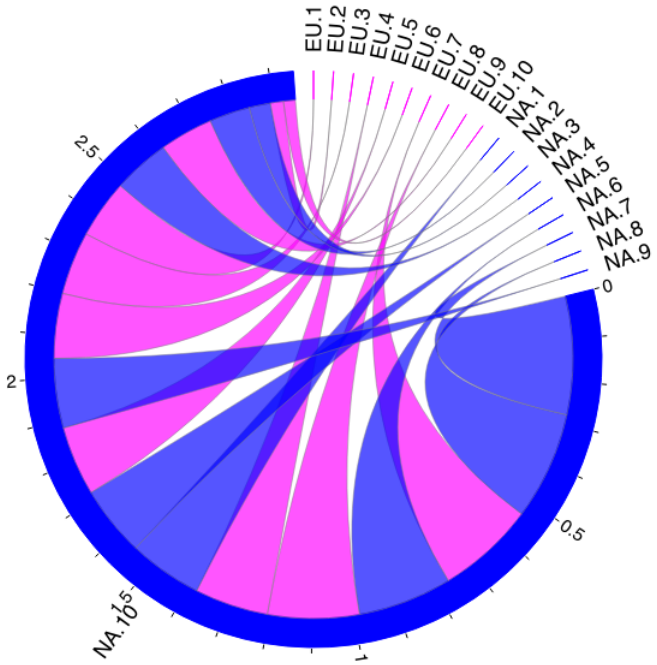
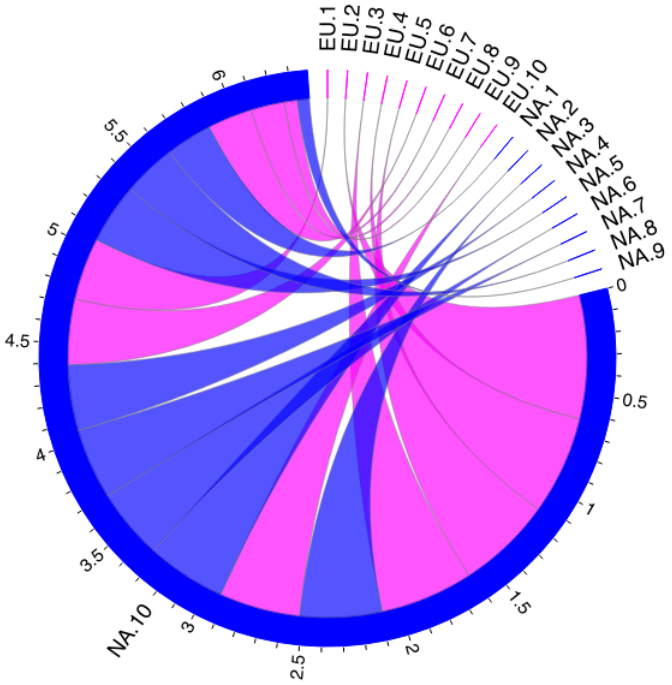


Figure 17: Adjacency matrices and serial dependence parameter in analysis with 10 groups and evaluation period length of 180 days.

(a) Adjacency matrices for the 10 groups in 2019



(b) Adjacency matrices for the 10 groups in 2020

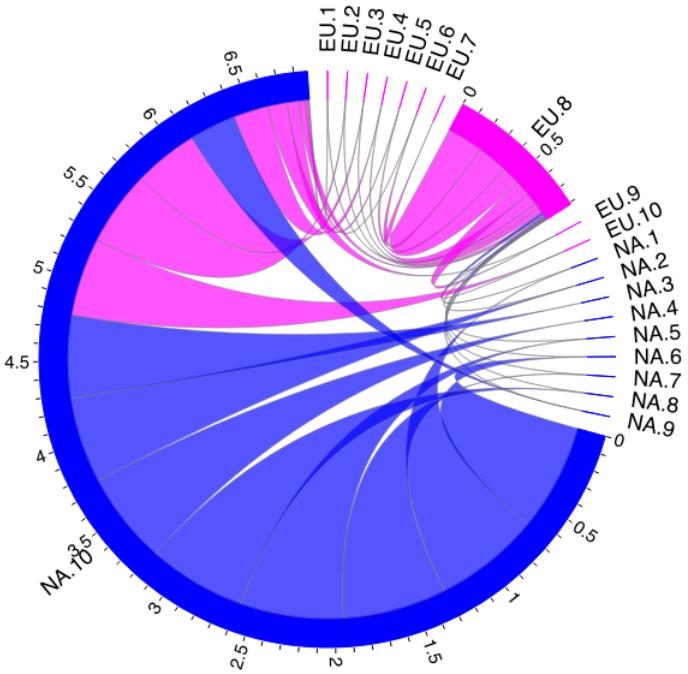
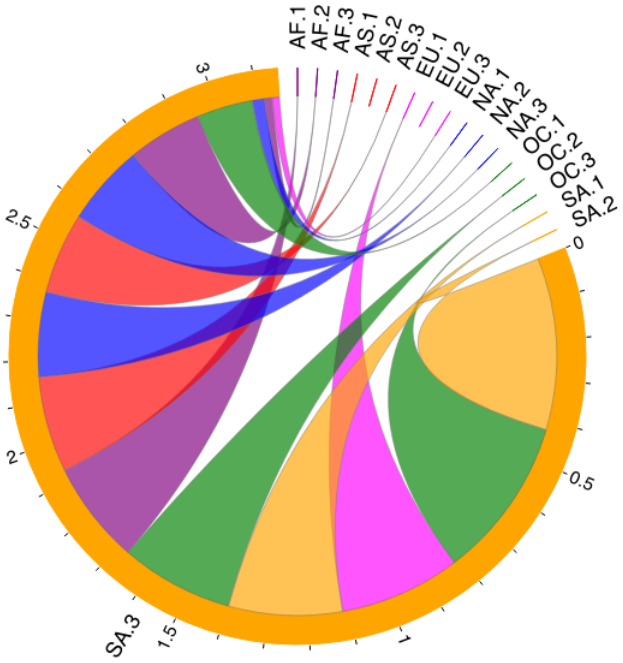


Figure 18: Adjacency matrices and serial dependence parameter in analysis with 3 groups and evaluation period length of 180 days.

(a) Adjacency matrices for the 3 groups in 2012



(b) Adjacency matrices for the 3 groups in 2014

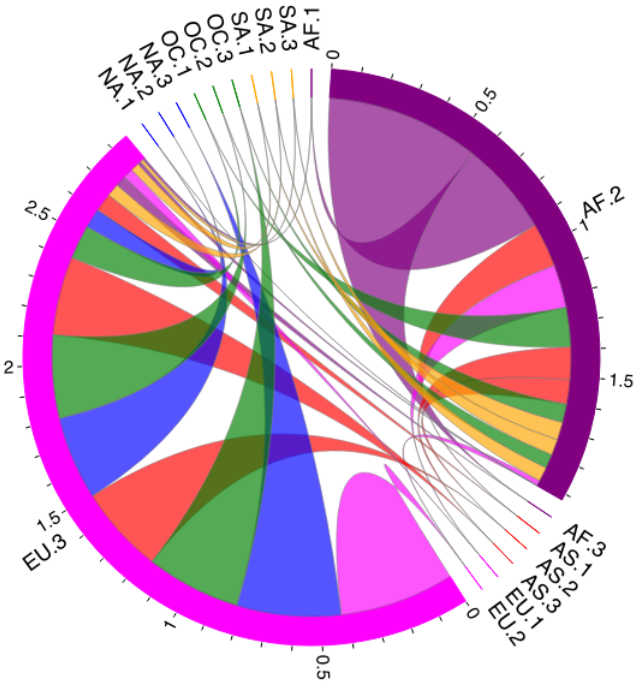
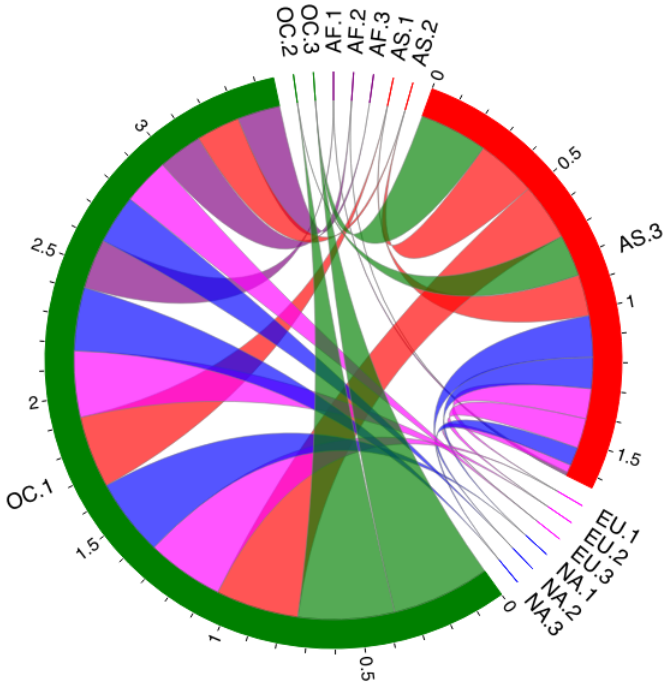


Figure 19: Adjacency matrices and serial dependence parameter in analysis with 3 groups and evaluation period length of 180 days.

(a) Adjacency matrices for the 3 groups in 2017



(b) Adjacency matrices for the 3 groups in 2018

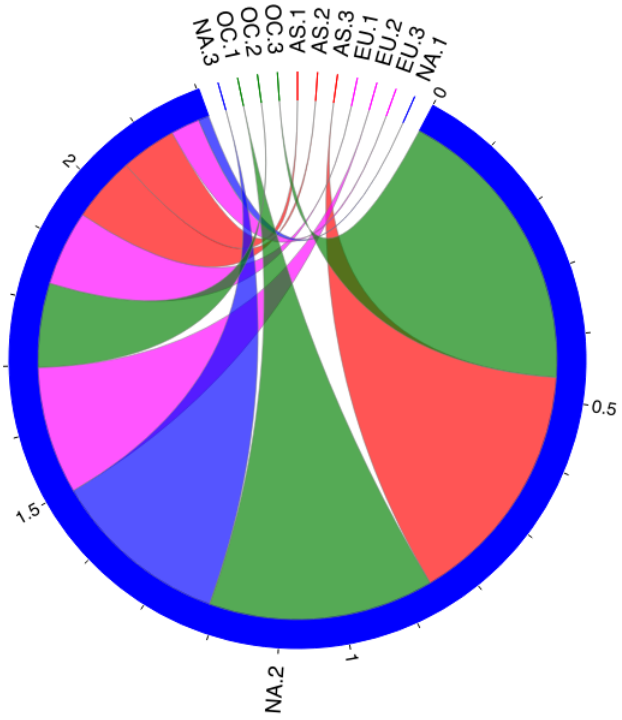
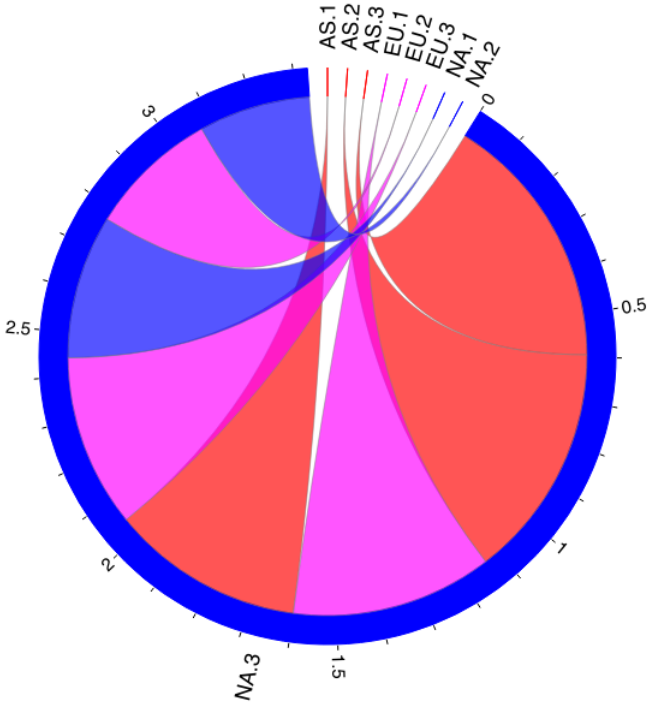


Figure 20: Adjacency matrices and serial dependence parameter in analysis with 3 groups and evaluation period length of 180 days.

(a) Adjacency matrices for the 3 groups in 2019



(b) Adjacency matrices for the 3 groups in 2020