Rejection graph for multiple testing of elliptical model for market network

Semenov D.P., Koldanov P.A.

Abstract Market network analysis attracts a growing attention last decade. Important component of the market network is a model of stock returns distribution. Elliptically contoured distributions are popular as probability model of stock returns. The question of adequacy of this model to real market data is open. There are known results that reject such model and at the same time there are results that approve such model. Obtained results are concerned to testing some properties of elliptical model. In the paper another property of elliptical model namely property of symmetry condition of tails of 2-dimentional distribution is considered. Multiple statistical procedure for testing elliptical model for stock returns distribution is proposed. Sign symmetry conditions of tails distribution are chosen as individual hypotheses for multiple testing. Uniformly most powerful tests of Neyman structure are constructed for individual hypotheses testing. Associated stepwise multiple testing procedure is applied for the real market data. To visualize the results a rejection graph is constructed. The main result is that under some conditions tail symmetry hypothesis is not rejected if one remove a few number of hubs from the rejection graph.

Key words: elliptically contoured distributions, tail symmetry condition, multiple decision statistical procedure, family wise error rate, Holm procedure, rejection graph, stock market

Semenov D.P., Koldanov P.A.
National Research University Higher School of Economics, Laboratory of Algorithms and Technologies for Network Analysis, Nizhny Novgorod, e-mail: dimsem2010@yandex.ru, pkoldanov@hse.ru

1 Introduction

Market network model is related with two main components: probabilistic model for stock return distribution and measure of similarity between stock returns. Probabilistic models of stock return distributions are subject of intensive study in theoretical and applied finance, portfolio selection and risk management. Elliptically contoured distributions are popular as probability model of stock returns [1]. The question of adequacy of this model to real market data is open. There are known results that reject such model and at the same time there are results that approve such model using statistical approach. For example, in [2] it was shown that while Student's copulas provide a good approximation for highly correlated pairs of stocks, discrepancies appear when the correlation between pairs of stocks decreases, which makes the elliptical model non adequate to describe the joint distribution of stocks. This results were obtained by testing symmetry, symmetry of tails and some other properties of elliptically contoured distributions. But, as the authors point out, their approach differs from the usual hypotheses testing by statistical tools. In the paper [3] statistical methodology for testing symmetry condition was proposed. Distribution free multiple decision statistical procedure based on uniformly most powerful tests of the Neyman structures was constructed and it was shown that under some conditions sign symmetry hypothesis is not rejected. To describe results of application of multiple decision statistical procedure to the USA and UK stock markets the concept of rejection graph was introduced.

In this paper we consider the problem of testing symmetry of tails for 2-dimensional distributions. Note that sign symmetry is a particular case of the tail symmetry. Uniformly most powerful Neyman structure test for hypothesis of tail symmetry for any pair of stocks is constructed. Multiple decision statistical procedure for simultaneously testing such hypotheses is constructed and applied to the stock markets of different countries. Numerical experiments show that hypothesis of tail symmetry for overall stock market is rejected. At the same time it is observed that the graph of rejected individual hypotheses has a specific structure. Namely, this graph is sparse and has several hubs of high degree. Removing this hubs leads to non-rejection of hypothesis of tail symmetry.

The paper is organized as follows. In Section 2, basic definitions and problem statement are given. In Section 3, multiple statistical procedure for testing symmetry condition for tails of 2-dimentional distributions is constructed. In Section 4, constructed procedure is applied to analysis UK, USA, Germany, France, India and China stock markets. Section 6 describes the evolution of the rejection graph when some hubs are removed. In Section 6, some concluding remarks are given.

2 Basic notations and problem statement

Let N is the number of stocks on the stock market and n is the number of observations. Let $p_i(t)$ is the price of the stock i at day $t (i = 1,\dots,N; t = 1,\dots,n)$. Then return of stock i for the day t is defined as follows:

$$x_i(t) = \ln(\frac{p_i(t)}{p_i(t-1)})$$

We assume that $x_i(t)$ $(i = 1,\dots,N; t = 1,\dots,n)$ is a sample (iid) from distribution of the random variables X_i. Distribution of the random vector $X = (X_1, X_2, \dots, X_N)$ belongs to the class of elliptically contoured distributions if its density functions is [4]:

$$f(x;\mu,\Lambda) = |\Lambda|^{-\frac{1}{2}} g(x-\mu)'\Lambda^{-1}(x-\mu), \tag{1}$$

where Λ is a positive definite matrix, $\mu = (\mu_1, \mu_2, \dots, \mu_N)$ - vector of means, $g(x) \geq 0$, and

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(y^{'}y) dy_1 \dots dy_N = 1$$

The well-known distributions such as multivariate Gaussian, and multivariate Student distributions belong to this class of distributions.

In what follows we assume that $\mu = (\mu_1, \mu_2, \dots, \mu_N)$ is known vector. Without loss of generality one can choose $\mu = (0,0,...,0)$. Upper and lower tail dependency ratio for the random variables X_i, X_j are defined by [2]:

$$\tau_{uu}(p) = P(X_i > c_{1-p}|X_j > c_{1-p}), \;\; \tau_{ll}(p) = P(X_i < c_p|X_j < c_p), \tag{2}$$

where c_{1-p} is quantile of the level p of X_i. The tail symmetry condition has the form $(0 < p < 1)$:

$$\tau_{uu}(p) = \tau_{ll}(p) \tag{3}$$

For a random vector $X = (X_1, X_2, \dots, X_N)$ with elliptical distribution the tail symmetry condition has to be satisfied for any pair of stocks. Moreover, if there is a pair of random variables X_i, X_j, such that tail symmetry condition is not satisfied then the random vector $X = (X_1, X_2, \dots, X_N)$ can't have an elliptical distribution. Note that for $p = \frac{1}{2}$ tail symmetry is equivalent to sign symmetry tested in [3]:

$$P(X_i > 0, X_j > 0) = P(X_i < 0, X_j < 0) \tag{4}$$

The authors detected pairs of stocks for which sign symmetry hypotheses are rejected and studied associated rejection graph.

In the present paper we construct and apply multiple statistical procedure for simultaneous testing of hypotheses:

$$h^{i,j} : P(X_i > c_i, X_j > c_j) = P(X_i < -c_i, X_j < -c_j) \tag{5}$$

where $i, j = 1, \dots N, i \neq j$. To accept the hypothesis about elliptical distribution it is necessary to accept hypotheses (5) for any i and j $(i \neq j)$.

Our main goal is to define which hypotheses are true and which are false. This information gives an opportunity to find set of stocks such that for any pair of stocks the tail symmetry hypothesis is accepted. For this set of stocks hypothesis of multivariate elliptical distribution will not be rejected.

3 Multiple hypotheses testing procedure

3.1 Test for individual hypothesis

Consider the individual hypothesis (5) for stocks i and j. Denote by:

$$p_1 = P(X_i > c_i, X_j > c_j),\;\; p_2 = P(-c_i < X_i < c_i, X_j > c_j),$$

$$p_3 = P(X_i < -c_i, X_j > c_j),\;\; p_4 = P(X_i < -c_i, -c_j < X_j < c_j),$$

$$p_5 = P(-c_i < X_i < c_i, -c_j < X_j < c_j),\;\; p_6 = P(X_i > c_i, -c_j < X_j < c_j),$$

$$p_7 = P(X_i > c_i, X_j < -c_j),\;\; p_8 = P(-c_i < X_i < c_i, X_j < -c_j),$$

$$p_9 = P(X_i < -c_i, X_j < -c_j)$$

Then hypothesis (5) can be written as:

$$h^{i,j} : p_1 = p_9 \tag{6}$$

To construct test for testing hypothesis $h^{i,j}$ let us introduce the indicators:

$$I^1_{x_i x_j}(t) = \begin{cases} 1, if x_i(t) > c_i \text{ and } x_j(t) > -c_j \\ 0, else \end{cases}$$

$$I^2_{x_i x_j}(t) = \begin{cases} 1, if -c_i < x_i(t) < c_i \text{ and } x_j(t) > c_j \\ 0, else \end{cases}$$

$$I^3_{x_i x_j}(t) = \begin{cases} 1, if -c_i > x_i(t) \text{ and } x_j(t) > c_j \\ 0, else \end{cases}$$

$$I^4_{x_i x_j}(t) = \begin{cases} 1, if -c_i > x_i(t) \text{ and } -c_j < x_j(t) < c_j \\ 0, else \end{cases}$$

$$I^5_{x_i x_j}(t) = \begin{cases} 1, if -c_i < x_i(t) < c_i \text{ and } -c_j < x_j(t) < c_j \\ 0, else \end{cases}$$

$$I^6_{x_i x_j}(t) = \begin{cases} 1, if x_i(t) > c_i \text{ and } -c_j < x_j(t) < c_j \\ 0, else \end{cases}$$

$$I^7_{x_ix_j}(t) = \begin{cases} 1, if x_i(t) > c_i \text{ and } x_j(t) < -c_j \\ 0, else \end{cases}$$

$$I^8_{x_ix_j}(t) = \begin{cases} 1, if -c_i < x_i(t) < c_i \text{ and } x_j(t) < -c_j \\ 0, else \end{cases}$$

$$I^9_{x_ix_j}(t) = \begin{cases} 1, if -c_i > x_i(t) \text{ and } x_j(t) < -c_j \\ 0, else \end{cases}$$

Let:

$$T_k = I^k_{x_ix_j} = \sum_{t=1}^{n} I^k_{x_ix_j}(t), \;\; k = 1, 2, \ldots, 9$$

The joint distribution of statistics $T_i (i = 1, \ldots, 9)$ has the form:

$$P(T_1 = k_1, T_2 = k_2, \ldots, T_9 = k_9) = \frac{n!}{k_1!k_2!\ldots k_9!} p_1^{k_1} p_2^{k_2} \ldots p_9^{k_9},$$

where $k_1 + k_2 + \ldots + k_9 = n$. In exponential form the joint distribution can be written as:

$$P(T_1 = k_1, T_2 = k_2, \ldots, T_9 = k_9) = \frac{n!}{k_1!k_2!\ldots k_9!} exp(\ln(p_1^{k_1}) + \ln(p_2^{k_2}) + \ldots + \ln(p_9^{k_9})) =$$

$$= \frac{n!}{k_1!k_2!\ldots k_9!} exp(\ln(p_1^{k_1}) + \ln(p_9^{k_9}) + \ln(p_9^{k_1}) - \ln(p_9^{k_1}) + \ln(p_2^{k_2}) + \ln(p_3^{k_3}) +$$

$$+ \ln(p_4^{k_4}) + \ln(p_5^{k_5}) + \ln(p_6^{k_6}) + \ln(p_7^{k_7}) + \ln(p_8^{k_8}) + \ln(p_8^{k_1}) - \ln(p_8^{k_1}) + \ln(p_8^{k_2}) - \ln(p_8^{k_2}) +$$

$$+ \ln(p_8^{k_3}) - \ln(p_8^{k_3}) + \ldots + \ln(p_8^{k_9}) - \ln(p_8^{k_9})) =$$

$$= \frac{n!}{k_1!k_2!\ldots k_9!} exp(k_1 \ln \frac{p_1}{p_9} + (k_1 + k_9) \ln \frac{p_9}{p_8} + k_2 \ln \frac{p_2}{p_8} + k_3 \ln \frac{p_3}{p_8} + \ldots + k_7 \ln \frac{p_7}{p_8} + nlnp_8) =$$

$$= \frac{n!}{k_1!\ldots k_7!k_9!} \frac{(1 - p_1 - \ldots - p_7 - p_9)^n}{(n - k_1 - \ldots - k_7 - k_9)!} exp(k_1 \ln \frac{p_1}{p_9} + (k_1 + k_9) \ln \frac{p_9}{p_8} + k_2 \ln \frac{p_2}{p_8} + \ldots + k_7 \ln \frac{p_7}{p_8})$$

According to [5] the UMP test for testing individual hypothesis has the form:

$$\phi_{i,j} = \begin{cases} 0, \text{ if } d_1((k_1 + k_9), k_2, \ldots, k_7) < k_1 < d_2((k_1 + k_9), k_2, \ldots, k_7) \\ 1, \text{else} \end{cases}$$

where d_1 and d_2 are defined from:

$$P(d_1 > k_1 \text{ or } k_1 > d_2)/h^{i,j}) =$$

$$= P(T_1 = k_1 < d_1 \text{ or } T_1 = k_1 > d_2/T_1 + T_9 = k_1 + k_9, T_2 = k_2, \ldots, T_7 = k_7) = \alpha,$$

where α is a given significance level. One has:

$$P(T_1=k_1/T_1+T_9=k_1+k_9,T_2=k_2,\ldots,T_7=k_7)=$$

$$=\frac{P(T_1=k_1,T_1+T_9=k_1+k_9,T_2=k_2,\ldots,T_7=k_7)}{P(T_1+T_9=k_1+k_9,T_2=k_2,\ldots,T_7=k_7)}$$

where

$$P(T_1+T_9=k,T_2=k_2,\ldots,T_7=k_7)=\frac{n!}{k!k_2!k_3!\ldots k_7!}p_2^{k_2}p_3^{k_3}\ldots p_7^{k_7}(p_1+p_9)^k$$

and

$$P(T_1=k_1,T_1+T_9=k,T_2=k_2,\ldots,T_7=k_7)=\frac{n!}{k_2!k_3!\ldots k_7!}p_2^{k_2}p_3^{k_3}\ldots p_7^{k_7}\frac{p_1^{k_1}p_9^{k_9}}{k_1!k_9!}$$

Therefore, conditional distribution has the form:

$$P(T_1=k_1/T_1+T_9=k_1+k_9,T_2=k_2,\ldots,T_7=k_7)=C_k^{k_1}(\frac{p_1}{p_1+p_9})^{k_1}(\frac{p_9}{p_1+p_9})^{k-k_1}$$

Then the test $\phi_{i,j}$ can be written as:

$$\phi_{i,j}=\begin{cases}0, \text{ if } d_1(k)<k_1<d_2(k)\\ 1, \text{ else}\end{cases},$$

where $k_1+k_9=k$. If hypothesis (5) is true, then $p_1=p_9$ and one has:

$$P(T_1=k_1/T_1+T_9=k_1+k_9,T_2=k_2,\ldots,T_7=k_7)=C_k^{k_1}(\frac{1}{2})^{k_1}(\frac{1}{2})^{k-k_1}$$

Finally $d_1(k)$ and $d_2(k)$ are defined by:

$$d_1(k)=max(C:(\frac{1}{2})^k\sum_{i=0}^{c}C_k^i\leq\frac{\alpha}{2})$$

$$d_2(k)=min(C:(\frac{1}{2})^k\sum_{i=c}^{k}C_k^i\leq\frac{\alpha}{2})$$

3.2 Holm procedure

The Holm step-down procedure is applied for simultaneous testing individual hypotheses $h_{i,j}$. This procedure consists of at most $M=C_N^2$ steps. At each step either of individual hypothesis $h_{i,j}$ is rejected or all remaining hypotheses are accepted. Let α be a Family-Wise Error Rate (FWER) of the multiple testing procedure and $q_{i,j}$ be a p-value of the individual test for testing hypothesis $h_{i,j}$. The procedure is constructed as follows:

- **Step 1:** If

$$\min_{i,j=1,..,N} q_{i,j} \geq \frac{\alpha}{M} \quad (7)$$

then accept all hypotheses $h_{i,j}, i,j = 1,..,N$
else if $\min_{i,j=1,..,N} q_{i,j} = q_{i_1,j_1}$ then reject hypothesis h_{i_1,j_1} and go to step 2.

- ...
- **Step K:** Let $I = (i_1, j_1), (i_2, j_2), \ldots, (i_{K-1}, j_{K-1})$ be the set of indexes of previously rejected hypotheses. If

$$\min_{(i,j)\notin I} q_{i,j} \geq \frac{\alpha}{M-K+1} \quad (8)$$

then accept all hypotheses $h_{i,j}$, $(i,j) \notin I$,
else if $\min_{(i,j)\notin I} q_{i,j} = q_{i_K,j_K}$ then reject hypothesis h_{i_K,j_K} and go to step (K+1).

- ...
- **Step M:** Let $I = (i_1, j_1), (i_2, j_2), \ldots, (i_{M-1}, j_{M-1})$ be the set of indexes of previously rejected hypotheses. Let $(i_M, j_M) \notin I$. If

$$q_{i_M,j_M} \geq \alpha \quad (9)$$

then accept hypothesis h_{i_M,j_M}, else reject hypothesis h_{i_M,j_M} (reject all hypotheses).

4 Practical application

The procedure for testing individual hypothesis (5) was applied to analyze data from the stock markets of US, UK, France, Germany, India and China for the period from January 1, 2006 to December 31, 2016. For each country, the greatest by sales stocks were selected, which were present on the market whole period. The number of stocks selected is $N = 100$, the sample size is $n = 250$ (1 calendar year). Constants c_i and c_j are chosen as an estimations of quantiles of order p from marginal distributions of stocks i and j respectively. For each year, for each country, symmetry conditions for tail distributions were tested and hubs from the rejection graph were obtained. These hubs are removed from consideration and the procedure is repeated until all individual hypotheses are accepted. At each iteration the Holm procedure is used.

The obtained results are shown in the tables below. Significance level for Holm procedure was chosen equals $0,05$. In each table element $(i,j) = k$ means that with quantile level p (element $(i,1)$) of X_i at year that are placed in $(1,j)$ it is necessary to delete k stocks to satisfy tail symmetry property. For example, for USA market, year = 2006, and $p = 0,05$ it is necessary to delete 22 stocks to obtain set of stocks such that for any pair of stocks the symmetry property is satisfied (Holm procedure is applied).

Table 1 USA Market

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
$p = 0,05$	22	14	17	8	9	17	18	22	23	19	25
$p = 0,1$	29	32	26	16	20	25	22	37	33	25	40
$p = 0,25$	32	25	43	32	27	32	28	36	38	39	55
$p = 0,5$	31	32	38	26	30	50	30	34	41	36	57

Table 2 UK Market

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
$p = 0,05$	21	13	5	11	10	8	21	27	24	23	25
$p = 0,1$	28	16	8	25	17	18	34	35	38	32	29
$p = 0,25$	46	34	17	36	42	27	41	44	47	47	45
$p = 0,5$	24	22	22	49	46	20	39	41	38	34	31

Table 3 Germany Market

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
$p = 0,05$	9	8	13	9	16	6	3	19	17	12	11
$p = 0,1$	16	22	27	24	19	16	17	34	30	21	17
$p = 0,25$	32	31	42	33	27	29	21	45	41	28	31
$p = 0,5$	29	23	57	40	32	34	23	64	43	20	22

Table 4 France Market

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
$p = 0,05$	22	13	13	9	21	7	14	19	21	18	21
$p = 0,1$	34	20	23	21	39	17	30	25	32	25	30
$p = 0,25$	45	31	30	29	49	21	35	35	36	35	35
$p = 0,5$	57	25	33	23	40	21	20	25	32	36	38

Table 5 China Market

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
$p = 0,05$	13	17	12	28	24	14	13	23	38	14	12
$p = 0,1$	20	34	21	51	33	25	20	31	61	28	22
$p = 0,25$	37	38	32	60	41	37	48	48	72	56	46
$p = 0,5$	49	50	0	51	41	45	41	45	76	61	57

Table 6 India Market

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
$p = 0,05$	3	11	9	11	18	12	12	13	21	13	10
$p = 0,1$	6	20	16	11	23	23	28	19	26	17	15
$p = 0,25$	16	26	13	27	36	38	35	29	32	22	23
$p = 0,5$	74	9	6	8	6	10	10	9	7	8	9

5 The evolution of the rejection graph.

The Figure 1-4 below (Appendix 2) demonstrate the changes in structure of the rejection graph in dependence of the number of stocks removed. It is shown clearly that after the removal of hubs, the number of rejected hypotheses gradually decreases and eventually tends to zero. Particularly, for 100 vertices the max vertex degree of graph is 43. By deleting hubs from this graph and remaining, for example, 90 vertices - the max vertex degree of graph decreases (this means that the number of pairs of stocks, for which the symmetry property is not satisfied, decreases). And when 30 vertices are deleted, the graph with the max vertex degree equals 2 is obtained: for node that has maximum degree there are only 2 stocks for which the symmetry property is not satisfied.

6 Discussion

In this paper elliptical model as a model of multivariate distribution of stocks returns was considered. Statistical procedure for multiple hypotheses testing of tails symmetry is constructed. This procedure were applied for the data from US, UK, China, India, Germany and France markets. The sets of stocks for any pair of which tail symmetry is satisfied have been obtained. It is shown that for small vakue of quantiles it is necessary to remove less stocks to satisfy tail symmetry condition for all pair of stocks. There are years where it necessary to remove small number of stocks (for example, 3 for India (2006 year) or 3 for Germany (2012 year)) to satisfy tail symmetry condition for all pair of stocks.

Separate discussion can be devoted to constants c_i and c_j choice. In the paper this constants are chosen as an estimations of quantiles of order p from marginal distributions of stocks i and j respectively. But there are another ways how to choose them. This is the problem for further investigations.

Acknowledgements The work is partially supported by RFHR grant 15-32-01052.

References

1. F.K. Gupta, T.Varga and T. Bondar (2013). Elliptically contoured Models in Statistics and Portfolio Theory. Springer.
2. Chicheportiche, Remy and Bouchaud, Jean-Philippe, The Joint Distribution of Stock Returns is Not Elliptical (September 6, 2010). International Journal of Theoretical and Applied Finance, Vol. 15, No. 3, 2012.
3. Petr A. Koldanov, Nina Lozgacheva (2016). Multiple testing of sign symmetry for stock return distributions. International Journal of Theoretical and Applied Finance. December, 2016.
4. Anderson T.W. An introduction to multivariate statistical analysis. Wiley-Interscience, New-York, 3-d edition, 2003.

5. L. Lehmann and J. P. Romano (2005) Testing Statistical Hypotheses, third edition. New-York: Springer.
6. S. Holm (1979) A simple sequentially rejective multiple test procedure, Scandinavian Journal of Statistics 6(2), 65-70.

Appendix 1. List of stocks with tail symmetry

For US market (2016, quantile $p = 0,1$) the set of stocks for any pair of which the tail symmetry hypothesis is not rejected is as follows (60 tickers):

'AVY','EXR','AON','FRT','NDAQ','LVLT','LEN','ARNC','TDC','CLX','HST', 'CRM','SYK','GPN','TAP','GPC','RIG','MMC','EA','NLSN','MCHP','SLB', 'AAPL','NUE','GM','COH','MKC','ICE','CSRA','AMZN','ACN','LKQ','MA', 'CXO','CTL','HON','PNW','AEP','BMY','DVA','XOM','KSU','CBG','AME', 'GOOGL','ETN','SPLS','BAC','PXD','MTD','O','MTB','JCI','NTRS','RF', 'KHC','COF','PGR','TXN','ROK'

Appendix 2. The evolution of the rejection graph.

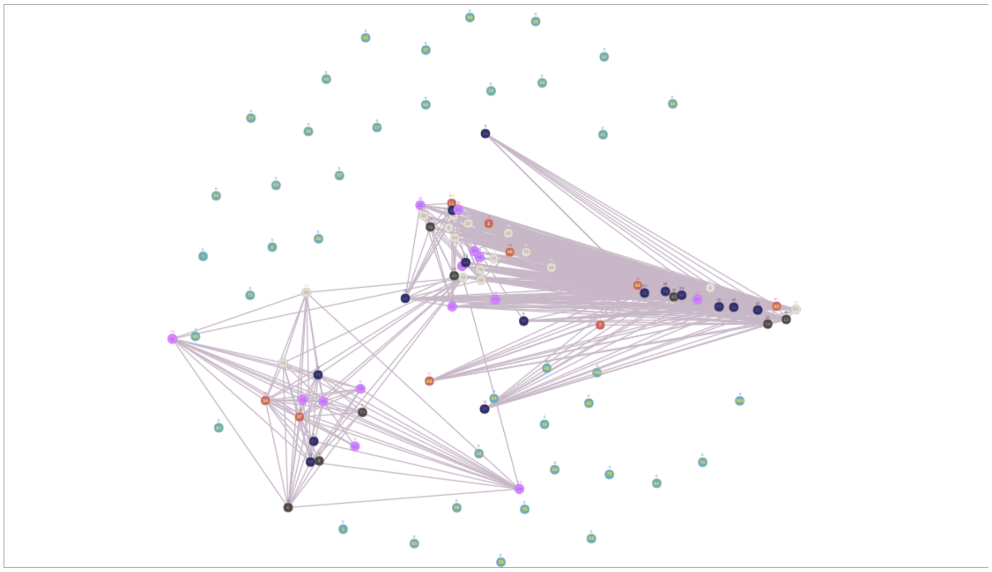

Fig. 1 Germany, 2016. Quantile $p = 0,25$. 100 vertices. Max degree of graph is 43

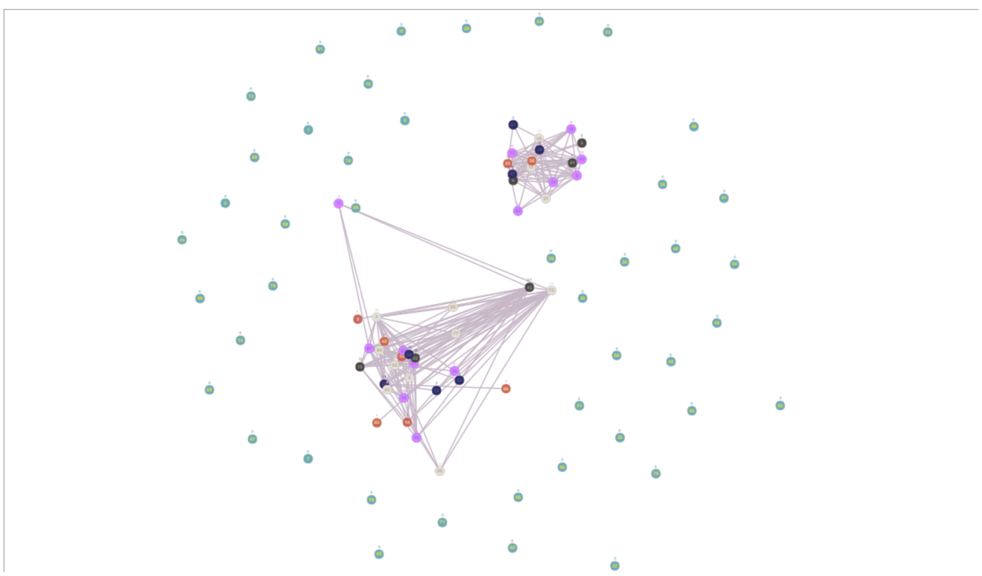

Fig. 2 Germany, 2016. Quantile $p = 0,25$. 90 vertices. Max degree of graph is 27

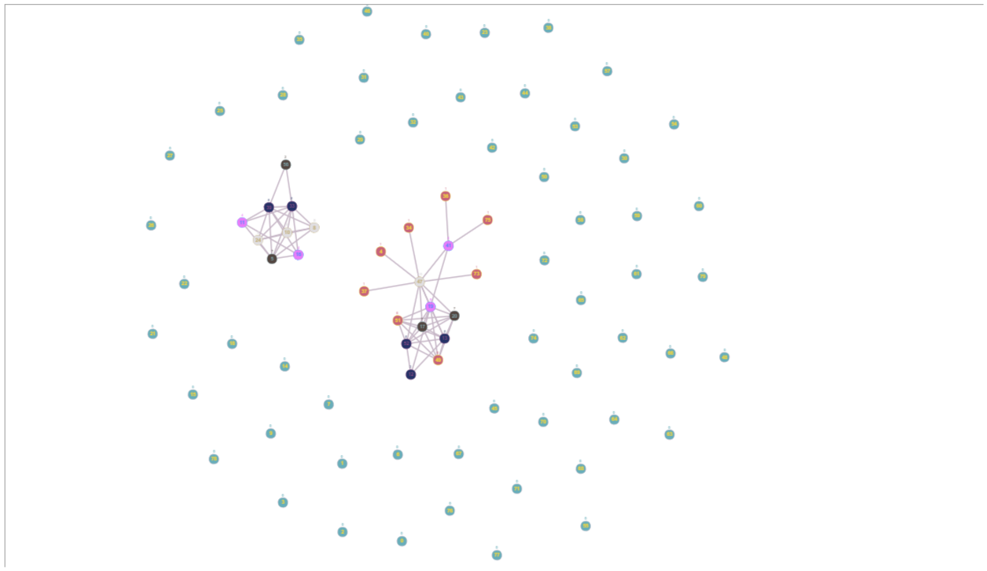

Fig. 3 Germany, 2016. Quantile $p = 0,25$. 80 vertices. Max degree of graph is 10

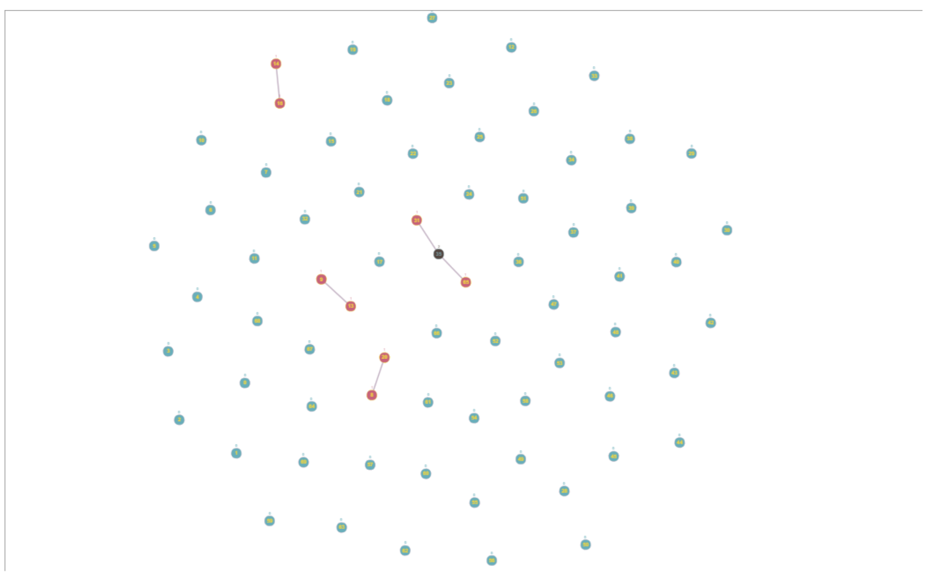

Fig. 4 Germany, 2016. Quantile $p = 0,25$. 70 vertices. Max degree of graph is 2