

Proceedings of the 6th Polish Symposium of Physics in Economy and Social Sciences (FENS2012), Gdańsk, Poland

# Network Analysis of Correlation Strength between the Most Developed Countries

J. MIŚKIEWICZ<sup>a,b</sup>

<sup>a</sup>Institute of Theoretical Physics, University of Wrocław, pl. M. Borna 9, 50-204 Wrocław, Poland

<sup>b</sup>Department of Physics and Biophysics, Wrocław University of Environmental and Life Sciences  
C.K. Norwida 25, 50-375 Wrocław

A new algorithm of the analysis of correlation among economy time series is proposed. The algorithm is based on the power law classification scheme followed by the analysis of the network on the percolation threshold. The algorithm was applied to the analysis of correlations among gross domestic product per capita time series of 19 most developed countries in the periods (1982, 2011), (1992, 2011) and (2002, 2011). The representative countries with respect to strength of correlation, convergence of time series and stability of correlation are distinguished. The results are compared with ultrametric distance matrix analysed by network on the percolation threshold.

DOI: [10.12693/APhysPolA.123.589](https://doi.org/10.12693/APhysPolA.123.589)

PACS: 89.65.Gh, 05.45.Tp

## 1. Introduction

Economy system is a perfect example of a complex system where various factors interact. Investigation of those interactions is a non-trivial task. On the other hand, knowledge about relationship between economy entities is crucial while preparing strategy and taking decisions. Analysis of correlations among significant number of entities can be divided into two steps: correlation analysis and correlation matrix analysis. The second part arises due to the fact that the correlation matrix consists of hundreds (or thousands of elements) and it would be difficult to analyse each element separately.

Since the series of works by Mantegna and collaborators [1–7] the standard approach to tackle the correlation problem in an economy system is to calculate distance matrix between entities and based on it build a minimum spanning tree (MST), which provides various information about clusterings, sectors, leaders/hubs etc. The analysis is in fact an extension of the Markowitz portfolio analysis [8], where the Pearson correlation coefficient plays a key role in portfolio optimisation procedure.

Although very successful in stock market analysis in the case of macroeconomy systems the other algorithms (with different distance metrics or network structures used) seem to be more useful [9–16]. The main disadvantages of the Mantegna algorithm result from the choice of the distance metrics. It has been shown [12, 16] that the ultrametric distance (UD) is sensitive to the presence of noise and wrongly classifies nonlinear correlations. Therefore it is very disputable whenever UD reflects fully the relationship between time series. Moreover, the second part of the Mantegna algorithm i.e. MST construction is in fact a very strong restriction influencing structure of generated network.

Recall that the MST algorithm solves the problem to find a graph which connects all the vertices, but is the smallest considering the number of links (or lowest sum

of the links weights) [17, 18]. Within the present paper an alternative algorithm of the time series correlation analysis is used. The algorithm consists of two elements: (i) correlation matrix construction, where instead of the UD a power law classification scheme (PLCS) is used and (ii) the analysis of the graph on the percolation threshold.

## 2. Time series correlation analysis

### 2.1. Power law classification scheme

The standard correlation analysis is based on two fundamental methods: the Pearson correlation coefficient [19] and regression analysis, e.g. [20]. Both methods are focused on linear correlation investigations. Recently an interesting detrended cross-correlation analysis (DCCA) was proposed [21], which allows to investigate long-range correlations with power law cross-correlation function. Sometimes “back-doors” are used to apply regression analysis to investigate nonlinear correlations by applying appropriate transformations of variable i.e. log–log transformation makes possible to apply regression analysis to verify power law dependences. However, if the nonlinear dependence of unknown form is expected, various rank methods, entropy based methods or numerous hypothesis tests [22–28] are used. Unfortunately, they are able to answer the question if the correlation exists, but cannot find the character of the correlation. Therefore, there is a need to develop a method which could investigate nonlinear dependences between time series.

In the present paper the Manhattan distance (MD) [29, 30] (known also as taxi cab distance) will be used as a starting point of the proposed correlation analysis method. MD is the cumulative sum of the absolute value of the differences between two time series and, as it has been shown [12, 16], it is more robust to the noise presence than the Pearson correlation coefficient, besides allows to analyse nonlinear dependences. The crucial feature of MD is the fact that its value depends on the length

of the time series — it is a nondecreasing function of its length.

Let us consider two time series  $A$  and  $B$  and denote their elements as  $a_i$  and  $b_i$ , respectively. The correlation between the time series will be defined as:

$$a_i = f(b_i), \quad (1)$$

where  $i \in (1, \dots, N)$ ,  $N$  is the number of the elements in considered time series  $A$  or  $B$ . It is assumed that the time series are of equal length or only the overlapping parts are considered. MD distance between  $A$  and  $B$  is defined:

$$\text{MD}(A, B) = \sum_{i=1}^N |a_i - b_i|. \quad (2)$$

As it was mentioned, MD is a nondecreasing function of the time series length therefore the time series  $A$  and  $B$  can be transformed into cumulative time series of MD

$$\text{MD}(A, B)(j) = \sum_{i=1}^j |a_i - b_i|, \quad j \in (1, 2, \dots, N). \quad (3)$$

The cumulative time series  $\text{MD}(A, B)(j)$  can be replaced by a continuous function  $\text{MD}(A, B)(x)$ , where  $x \in (1, N)$ . Moreover, the cumulative time series  $\text{MD}(A, B)(j)$  is strictly related to the surface between time series  $A$  and  $B$ . The procedure is in fact similar to the definition of the Riemann integral or various procedures of numerical integration e.g. [31, 32].

Let us approximate time series  $A$  and  $B$  by a continuous function and in order to avoid absolute value notation assume that  $a_i \leq b_i$ , additionally let assume that time series  $A$  and  $B$  are correlated Eq. (1), then

$$\text{MD}(A, B)(n) \simeq \int_0^n [a(t) - f(a(t))] dt. \quad (4)$$

Considering that the integral is calculated along  $A$  time series the correlation function can be found by

$$f(n) = \frac{d\text{MD}(A, B)(n)}{dn}. \quad (5)$$

In principal, the method is capable to detect arbitrary correlations, but due to the practical constraints such as a limited length of time series, stationarity of time series, presence of noise, etc. it is useful to distinguish a class of correlation functions. In analogy to the algorithm classification complexity [33, 34] a class of correlation function is introduced. Since one of the main aims of the method is to measure “strength” of correlation, the power laws functions are chosen as a basis of the method. Moreover, from the practical point of view power law functions are relatively easy to fit. Additionally, they are frequently reported in the analysis of financial data e.g. [35–37].

Considering the above PLCS is defined as follows:

1. Transform time series  $A$  and  $B$  into a cumulative MD time series (Eq. (3)).
2. Present the calculated cumulative MD time series as a function of the time series length in logarithmic scale.

3. Fit the linear function to the data and find the slope coefficient  $\alpha$  corresponding to the integrated correlation function.
4. Let the class of the correlation function be labelled as  $\gamma(A, B) = \alpha - 1$  (see Eq. (5)).
5. Additionally, calculate the quality of the fit, measured by the statistical significance probability  $\beta$ . It provides information about the stability of the correlation in the considered time interval.

## 2.2. Ultrametric distance

The UD [3] can be considered as a standard in analysis of relationship among shares [38–41], therefore is used to compare with the results of PLCS. For the convenience of the reader the main definitions and basic features of UD will be recalled here. The Pearson correlation coefficient is defined as:

$$P_{(t,T)}(A, B) = \frac{\langle AB \rangle_{(t,T)} - \langle A \rangle_{(t,T)} \langle B \rangle_{(t,T)}}{\sqrt{(\langle A^2 \rangle_{(t,T)} - \langle A \rangle_{(t,T)}^2)(\langle B^2 \rangle_{(t,T)} - \langle B \rangle_{(t,T)}^2)}}. \quad (6)$$

The UD in literature is used in two forms

$$\text{DU}_{\text{MS}}(A, B)_{(t,T)} = \sqrt{2[1 - P_{(t,T)}(A, B)]}, \quad (7)$$

as introduced by Mantegna [3] or

$$\text{DU}_{\text{AM}}(A, B)_{(t,T)} = \sqrt{\frac{1}{2}[1 - P_{(t,T)}(A, B)]}, \quad (8)$$

introduced in [15]. The difference between Eq. (7) and Eq. (8) lies in normalization. The Mantegna formulation map distance between time series into the interval  $(0, 2)$ , while the definition Eq. (8) maps the distance between time series into the interval  $(0, 1)$ , where zero corresponds to the linearly correlated time series, one refers to anti-correlated time series and  $\frac{\sqrt{2}}{2}$  to the case where linear correlation was not found. In the following Eq. (8) will be used and denoted by UD.

## 2.3. Network analysis

The distance or, in general, correlation analysis in the case of investigations of group of stocks, bonds, shares or economy indices results in distance/correlation matrix, which consists of hundreds of data. The detailed analysis of such a matrix is beyond standard investigations. Moreover, usually researches are not interested in the correlation itself but rather what kind of structure of dependences do they form. Therefore, the correlation/distance matrix is usually analysed by constructing and analysing a chosen network structure. In the case of correlation analysis based on UD the most popular structure is the MST. It should be noticed that the network structure can influence strongly the results of the analysis. Therefore, in the present analysis a network on the percolation threshold is chosen. This network is constructed by removing links from the smallest (or strongest) correlations. Links are removed until the next removed link

leads to breakdown of the network into disconnected sub-networks. Finally, the structure of network on percolation threshold (NPT) is investigated.

### 3. Correlations of developed countries

The analysis of correlations was performed for the group of developed countries (in alphabetic order): Australia (AU), Austria (AT), Belgium (BE), Canada (CA), Denmark (DN), France (FR), Germany (DE), Great Britain (GB), Greece (GR), Holland (NL), Italy (IT), Ireland (IR), Japan (JP), Luxembourg (LU), Poland (PL), Portugal (PT), Spain (SE), Sweden (SW), USA (US). The countries were characterised by the gross domestic product (GDP) per capita time series. The time series were downloaded from the Mathematica data base and covered the interval (1970, 2011). The correlation analysis was conducted for the short (10 years: (2002, 2011)), medium (20 years: (1992, 2011)) and long (30 years: (1982, 2011)) time span. PLCS algorithm is based on MD therefore correlation strength and stability commutes i.e.  $\gamma(A, B) = \gamma(B, A)$  and  $\beta(A, B) = \beta(B, A)$ . Therefore, the correlation and stability matrix consists of 171 unique values.

#### 3.1. Key results of PLCS

The detailed analysis of a correlation matrix is a laborious process, therefore it is analysed by constructing the chosen network structure. However, for the sake of clarity some representative cases will be discussed in detail.

##### 3.1.1. The negative correlation strength

The negative correlation strength  $\gamma$  corresponds to the situation where MD between analysed time series is decreasing so the time series are converging. The example of negative correlation strength are presented in Table I. The plots of GDP/c and cumulative MD in logarithmic scale with fitted linear function to the data points at large  $n$  are presented in the case of the countries with the lowest correlation strength — Denmark and Ireland (Fig. 1). The PLCS negative correlation strength corresponds to

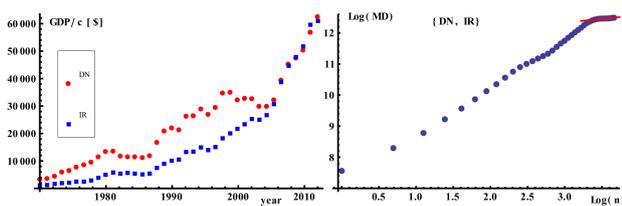


Fig. 1. Left: the GDP per capita of Denmark and Ireland in US [\$] in years (1970, 2011). Right: cumulative MD between Denmark and Ireland as a function of the time series length in logarithmic scale. A linear function (red line) is fitted to ten data points at large  $n$ .

the evolution of the GDP/c of DN and IR. In fact, since 1998 GDP/c of those countries are converging and since 2005 they are overlapping. The UD of pairs presented

in Table I takes values in the range (0.05; 0.19) which might be considered as an indication of possible linear correlation.

TABLE I

Correlation strength correlation stability and UD of the first three pairs of countries with the most negative  $\gamma$  in recent ten (2002, 2011), twenty (1992, 2011) and thirty years (1982, 2011).

10 years (2002, 2011)			
Pair	$\gamma$	$\beta$	UD
(DN, IR)	-0.76	$4.3 \times 10^{-5}$	0.10
(SW, GB)	-0.53	$1.2 \times 10^{-6}$	0.10
(NL, SW)	-0.53	$1.9 \times 10^{-8}$	0.06
20 years (1992, 2011)			
(DN, IR)	-0.76	$1.0 \times 10^{-6}$	0.14
(SW, GB)	-0.53	$1.7 \times 10^{-4}$	0.19
(NL, SW)	-0.53	$3.3 \times 10^{-9}$	0.14
30 years (1982, 2011)			
(IT, AU)	-0.18	$8.4 \times 10^{-21}$	0.16
(AT, DE)	-0.16	$1.0 \times 10^{-11}$	0.05
(IT, CA)	-0.13	$4.0 \times 10^{-20}$	0.14

##### 3.1.2. The correlation strength close to zero and one

The correlation strength equal to zero or one corresponds to the situation where the time series are linearly correlated. The first one, where correlation strength is close to zero, is in fact a trivial correlation where the time series are overlapping. The pairs for which the correlation strength was close to zero are presented in Table II. The plot examples of cumulative MD with the fitted linear functions in the case of correlation strength is close to zero or one ((GB, CA) and (GR, PT), respectively) are presented in Fig. 2. In such a case UD might classify the dependence as a linear correlation, although in the case of correlation between (GR, PT) the results might change if different time interval is chosen. This example illustrates the possibility of two trends which can compensate.

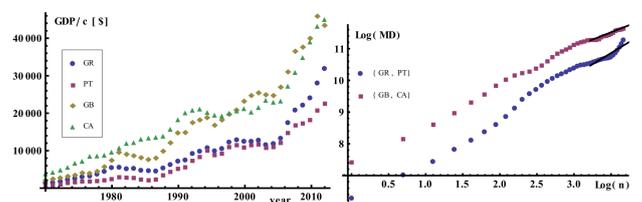


Fig. 2. Left: the GDP per capita of Greece, Portugal, Great Britain and Canada in US [\$] in years (1970, 2011). Right: cumulative MD between (Greece, Portugal) and (Great Britain, Canada) as a function of the time series length in logarithmic scale. Linear functions (black lines) are fitted to ten data points at large  $n$  for every pair.

TABLE II

Correlation strength, correlation stability and UD of the pairs of countries with the linearly correlated GDP/c time series in recent ten (2002, 2011), twenty (1992, 2011) and thirty years (1982, 2011).

10 years (2002, 2011)							
Pair	$\gamma$	$\beta$	UD	Pair	$\gamma$	$\beta$	UD
(GB, CA)	-0.009	$3.3 \times 10^{-9}$	0.10	(GR, PT)	1.0	$2.3 \times 10^{-5}$	0.03
(GB, US)	0.028	$1.5 \times 10^{-6}$	0.12	(BE, US)	1.0	$1.4 \times 10^{-6}$	0.16
20 years (1992, 2011)							
(SW, GB)	-0.014	$1.6 \times 10^{-11}$	0.19	(BE, PL)	1.0	$3.5 \times 10^{-11}$	0.15
(IR, CA)	0.008	$3.4 \times 10^{-11}$	0.14	(DE, AU)	1.0	$3.9 \times 10^{-9}$	0.19
30 years (1982, 2011)							
(BE, AT)	-0.003	$3.1 \times 10^{-27}$	0.04	(IT, PT)	1.0	$2.2 \times 10^{-35}$	0.10
(NL, GB)	0.007	$9.3 \times 10^{-26}$	0.09	(LU, PT)	1.0	$1.1 \times 10^{-28}$	0.08

The results of PLCS algorithm and UD are particularly equivocal for GDP/c time series of Belgium and Austria for the recent 30 years, where  $\gamma = -0.003$  and  $UD = 0.04$ . On the other hand, very interesting is the case of two pairs Great Britain–Canada and Great Britain–USA. These countries are known from the close mutual cooperation. The correlation strength  $\gamma \approx 0$  indicates linear correlation of those time series, however UD value is greater than zero ( $UD(GB, CA) = 0.10$ ,  $UD(GB, US) = 0.12$ ), which can be also interpreted as a presence of some noise. Moreover, this supposition is supported by low stability of these correlations, which is equal to  $3.3 \times 10^{-9}$  and  $1.5 \times 10^{-6}$ , respectively. Notice that  $\beta$  of correlation strength for the considered 171 pairs of countries in the period (2002, 2011) is varying from  $2.2 \times 10^{-19}$  to  $3.2 \times 10^{-4}$ . Comparing the stability of correlation for the pairs (GB, CA) and (GB, US) and the results of the whole group it must be concluded that  $\beta$  lies close to low stability cases.

The second case of the correlation strength equal to one is the linear correlation. It is very interesting that PLCS point out clear, linear correlation between GDP/c of Greece and Portugal, particularly in the view of recent crisis in Greece and difficult situation in Portugal. On the other hand, linear correlation is observed between Poland and Belgium with high stability of the correlation ( $\beta = 3.5 \times 10^{-11}$ ), which suggests that there is a chance for a good prognosis for Poland.

### 3.1.3. High value of correlation strength

The positive value of correlation strength indicates that the analysed time series are divergent. The highest strength of correlation was observed in the case of Ireland and Spain. The GDP/c time series and cumulative MD time series in logarithmic scale are presented in Fig. 3. The fit of linear function was extended to the interval (1995, 2011), which corresponds to the divergent part of GDP/c time series (Fig. 3, left). It is worth noticing that UD of this pair of countries is equal to 0.07 (Table III), which might suggest existence of linear correlation between these time series.

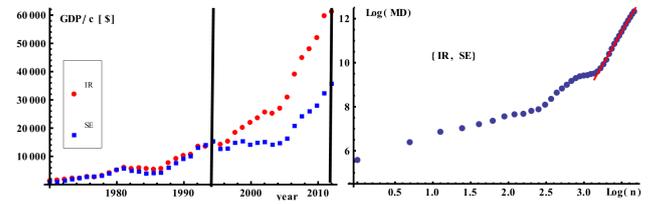


Fig. 3. Left: the GDP per capita of Ireland and Sweden in US [\$] in years (1970, 2011). The vertical lines denote the interval for which the linear function was fitted. Right: cumulative MD between Ireland and Sweden as a function of the time series length in logarithmic scale. A linear function (red line) is fitted to fifteen data points at large  $n$ .

TABLE III

The correlation strength, stability and UD of the last three pairs of countries with the highest  $\gamma$  in recent ten (2002, 2011), twenty (1992, 2011) and thirty years (1982, 2011).

10 years (2002, 2011)			
Pair	$\gamma$	$\beta$	UD
(IR, SE)	4.8	$2.5 \times 10^{-15}$	0.07
(GR, IR)	3.7	$4.4 \times 10^{-17}$	0.08
(FR, NL)	3.4	$6.7 \times 10^{-13}$	0.04
20 years (1992, 2011)			
(LU, SW)	4.0	$5.4 \times 10^{-11}$	0.14
(IR, SE)	3.6	$2.5 \times 10^{-15}$	0.13
(GR, IR)	3.1	$4.4 \times 10^{-17}$	0.11
30 years (1982, 2011)			
(GR, IR)	2.7	$7.3 \times 10^{-30}$	0.09
(LU, SW)	2.4	$2.2 \times 10^{-17}$	0.12
(IR, SE)	2.2	$1.3 \times 10^{-18}$	0.12

Comparing PLCS and UD results it must be concluded that PLCS gives more adequate description of the system. Of course, the detailed analysis should include analysis of various economy/political/social factors which is beyond the scope of this paper. However, it is worth

noticing that correlation understood in a broader than only linear sense takes place in this case.

Let take a closer look in the GDP/c time series of Ireland and Spain (Fig. 3) in the interval (1995, 2011). Two stages of development can be distinguished in this period: (i) (1995, 2004) and (ii) (2004, 2011), while in the second subinterval the GDP/c of both countries are growing rapidly. Despite the change in 2004 the correlation remains stable and the cumulative MD is growing according to the well established power law ( $\beta = 2.5 \times 10^{-15}$ ).

3.1.4. Stability of correlation

The PLCS algorithm besides correlation strength provides also the information about stability of the correlation given by a significance level — the probability at which the null hypothesis can be rejected.

The examples of the three most and less stable pairs are presented in Table IV. The very interesting case is the pair Poland–USA. This pair is present in the first three pairs with the most stable correlation in all of the considered intervals.

TABLE IV

The correlation strength, stability and UD of the pairs of countries with the highest and the lowest correlation stability in recent ten (2002, 2011), twenty (1992, 2011) and thirty years (1982, 2011).

High correlation stability				Low correlation stability			
10 years (2002, 2011)							
Pair	$\gamma$	$\beta$	UD	Pair	$\gamma$	$\beta$	UD
(PL, US)	0.9	$2.2 \times 10^{-19}$	0.18	(BE, DE)	0.2	$8.1 \times 10^{-5}$	0.04
(DN, AU)	1.1	$1.5 \times 10^{-17}$	0.04	(IT, AU)	0.3	$1.7 \times 10^{-4}$	0.06
(GR, IR)	3.7	$4.4 \times 10^{-17}$	0.08	(NL, GB)	-0.5	$3.2 \times 10^{-4}$	0.20
20 years (1992, 2011)							
(PL, US)	1.9	$2.2 \times 10^{-19}$	0.19	(BE, DE)	1.2	$8.2 \times 10^{-5}$	0.09
(DN, AU)	2.1	$0.5 \times 10^{-17}$	0.10	(IT, AU)	1.3	$1.7 \times 10^{-4}$	0.12
(GR, IR)	4.7	$4.4 \times 10^{-17}$	0.11	(NL, GB)	0.5	$3.3 \times 10^{-4}$	0.13
30 years (1982, 2011)							
(PT, US)	0.7	$1.5 \times 10^{-50}$	0.12	(FR, NL)	0.8	$1.8 \times 10^{-16}$	0.07
(PL, US)	0.8	$5.5 \times 10^{-49}$	0.24	(LU, JP)	0.6	$1.0 \times 10^{-14}$	0.33
(GR, US)	0.8	$9.1 \times 10^{-47}$	0.19	(AT, DE)	-0.2	$1.0 \times 10^{-11}$	0.05

The GDP/c and the cumulative MD with fitted linear function of (PL, US) are presented in Fig. 4. The UD of the pair (PL, USA) is equal to 0.18, 0.19, 0.24 for the periods ten, twenty and thirty years, respectively. The UD results are far from the linear correlation and it is very disputable whenever they can be interpreted as any correlation. PLCS, contrary to UD points out that nonlinear correlation exists and places the correlation among the most stable correlations. This case illustrates once more PLCS features. In the analysis of correlation between time series of significantly different values global trend is more important than fluctuations. Application of MD in PLCS places small fluctuations at the right position and cover them by the global trend. This perfectly agrees with first impression of the GDP/c graph (Fig. 4).

4. Structure of the networks on the percolation threshold

The analysis of the correlation matrix is performed by construction of the NPT.

1. The algorithm of the NPT begins with sorting the links between countries.

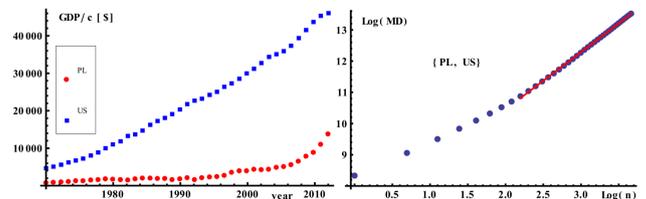


Fig. 4. The high stability of correlation on the example of Poland and USA. Left: GDP/c time series of Poland and USA in US [\$] in years (1970, 2011). Right: cumulative MD between Poland and USA as a function of the time series length in logarithmic scale. A linear function (red line) is fitted to thirty data points at large  $n$ .

2. In the next step of NPT algorithm last link of the chosen sorting is removed.
3. The integrity of the graph is verified.
4. If the graph is connected the second step of the algorithm is repeated. If the graph is not connected then the recently removed link is added to the graph and the algorithm is finished.

The main advantage of the network analysis is the ability to identify particularly interesting nodes. In economics, such points are leaders, which influence strongly other, sensitive entities, which reflect processes in related countries and well established relationships. The desired knowledge can be gained through by construction of NPT with appropriate sorting.

Considering the PLCS results there are three types of useful sorting: from the smallest correlation strength to the biggest, from the biggest correlation strength to the smallest and from the lowest value of correlation strength to the biggest. The first sorting corresponds to the situation where converging time series are preferred, the second sorting is particularly useful when strong non-linear relationship is suspected for which even small variations in one time series results in strong fluctuation in the second. The last sorting prefers the most stable correlations, so is the most probable. Therefore there can be constructed three possible networks: convergent (NPT-CP), strength (NPT-SP) and stability preferential (NPT-S).

The examples of NPT based PLCS analysis of GDP/c time series in the time interval (2002, 2011) are presented in Figs. 5–8. Due to the limitation of the space graphs of NPT constructed for twenty and thirty-year periods are available from the author on request. The main hub of NPT-CP network (Fig. 5) is Luxembourg, which is connected with 12 countries. Since this is the convergent preferential network, Luxembourg can be considered as a representative European country. This network has one more interesting feature, which cannot be observed in trees — cliques. Three cliques can be distinguished in Fig. 5: LU-DE-AT, LU-BE-NL, LU-NL-FR. This grouping reflects strong economical, cultural and even linguistic relationship among these countries.

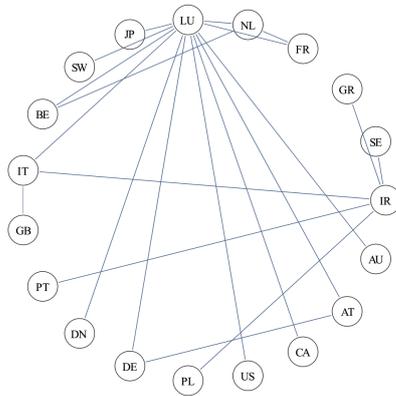


Fig. 5. Convergent preferential NPT constructed on the PLCS correlation strength matrix for the time period (2002, 2011).

The NPT-SP (Fig. 6) is in some sense a complementary network to NPT-CP, since mean degree of the node is 11.8 (NPT-SP) comparing to 2.2 (NPT-CP), so the network is close to the complete graph. In this case Luxembourg is the less connected country and is linked only

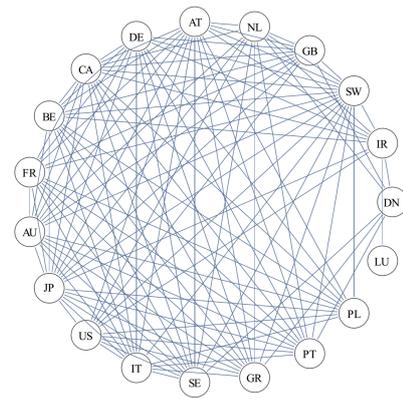


Fig. 6. Strength preferential NPT based on the PLCS correlation strength matrix for GDP/c time series of chosen countries in years (2002, 2011).

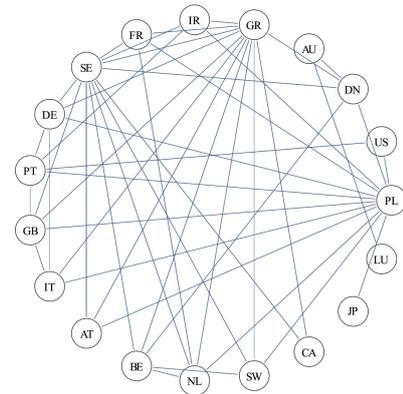


Fig. 7. Stability preferential NPT based on the PLCS correlation stability matrix for GDP/c time series of chosen countries in years (2002, 2011).

to Ireland, while the countries with the highest number of links are: Japan (16), Sweden (16), Canada (15) and Austria (15) — the number of links is given in brackets. It is noticeable that mentioned countries are the developed countries, which are globally influential. This

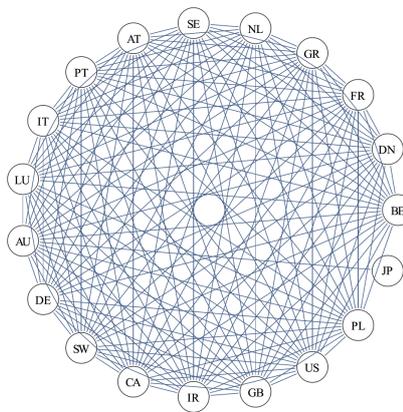


Fig. 8. NPT based on the UD matrix for GDP/c time series of chosen countries in years (2002, 2011).

finding agree with the main purpose of NPT-SP, which should point out the most influential entities.

The third type of the network — NPT-S is constructed in order to point out the nodes, which form stable relationship with the rest of the network. In the period (2002, 2011) the nodes with the highest number of links were: Greece (12), Poland (12) and Spain (11). The common factor of these hubs is that they have relatively low GDP/c so the key factor is the main trend while the fluctuations of GDP/c are higher order corrections in cumulative MD time series. On the other hand, nodes with particularly stable relationship with the other countries might serve as a good starting point for preparing prognosis for the whole region. Unfortunately, Greece and Spain are recently under a strong financial pressure which casts a shadow on the presented results.

The NPT based on the UD of GDP/c in recent 10 years is in fact almost complete graph, Fig. 8, where Belgium is connected to 18 countries, Japan is connected only to Australia therefore the remaining countries are linked to 17 nodes.

The conclusions of ten-years period NPT analysis are also valid for twenty- and thirty-years periods. NPT-CP constructed for the period (1992, 2011) point out Luxembourg as the node with highest number of links (14 countries connected). The other considered countries have significantly smaller degree (between 4 and 1). The following cliques are observed FR-NL-DE, PL-PT-IR, GR-SE-IR, LU-IT-GB, AT-LU-JP and IT-LU-JP. Comparing the set of cliques observed in twenty-years period with those seen in ten-years interval it can be noticed that the number of cliques is doubled. Moreover, the clusters can be divided into two groups: high GDP/c level (FR-NL-DE, LU-IT-GB, AT-LU-JP, IT-LU-JP) and lower GDP/c level (PL-PT-IR, GR-SE-IR).

NPT-SP for the twenty-year period is a network with very high number of links (mean node degree is equal to 12.1). The countries with the highest number of links are Sweden (node degree: 16), USA (node degree: 15) and Portugal (node degree: 15). The less connected country is Luxembourg — similarly to the case of NPT-SP in ten-years analysis. The stability preferential NPT based on PLCS twenty-years period analysis mark out Luxembourg and Poland as the nodes with the highest degree (17 nodes attached). Combining the conclusions of NPT-S and NPT-CT analysis for the twenty-years period it should be noticed that Luxembourg plays a special role for the chosen group of countries — it is the biggest hub in convergent and stability preferential network, which indicates this country as a central point for the prognosis and economical analysis of the region.

NPT based on UD matrix for twenty-year interval (1992, 2011) is very similar to the ten-year analysis (Fig. 8) with the difference: Germany is the country with the maximum degree (18 links), Japan is the lowest degree node (one link), other countries have 17 links.

In the longest considered time interval (1982, 2011) the highest degree nodes of NPT-CP are Japan and Lux-

embourg, which are connected to 9 other nodes. In this network there is present the four-nodes clique: GR-IR-SE-JP, besides that there are ten three-nodes cliques: GB-JP-IT, PL-JP-IT, PT-PL-IR, CA-LU-DN, CA-US-LU, AU-US-LU, DN-NL-LU, NL-BE-LU, FR-DE-LU, IT-GR-JP. It is interesting that the Benelux countries form a clique at the thirty-years period, so this is rather long-term relationship than short term cooperation or coalition.

Analysing the strength preferential NPT-ST in the interval (1982, 2011) one can observe that Sweden is the country with the highest links attached (12 links). Considering the fact that Sweden is a highest degree node for NPT-SP for ten, twenty and thirty-years periods, this country should be considered as an indicator of trend in GDP/c evolution for the most influential countries. The last PLCS based NPT is the stability preferential NPT-S for the period (1982, 2011). In this case Poland plays the role of the main hub having 17 nodes attached. An interesting observation is that Poland belongs to the group of the main hubs for all of the considered intervals. Of course it is worth further economical analysis, but at the beginning of '80 Poland underwent economical and political transformation and is adopting its economy and law to the developed country standards, therefore its GDP/c time series are "following" those of developed countries. The UD based NPT network in the longest considered here interval is in fact difficult to analyse since there is no well established leaders or a group of the key hubs. On the network there are nine nodes which have 17 links, five nodes with 16 nodes attached and three countries with 15 country connected. Only Luxembourg and Japan have 12 and 6 linked countries, respectively.

## 5. Conclusions

The analysis of correlation is a often key problem in scientific analysis particularly in economics. In the presented paper a new algorithm of correlation analysis is proposed. The algorithm is based on the power law classification scheme and the analysis of network of percolation threshold. The algorithm was applied to the analysis of correlation among GDP/c time series of the 19 most developed countries. It has been shown that PLCS is capable to recognise other than linear correlations, particularly points out converging and diverging time series. It provides also information on the stability of correlations, which is particularly important while preparing prognosis or simply justifying quality of the discovered correlation. Moreover, when applied to analysis of linear correlations the method is more robust to noise than the UD.

Finally, the PLCS outcome was used to construct following networks: NPT-CP, NPT-SP and NPT-S. The network analysis gives the opportunity to point out the entity which is representative for the group, with respect to analysed feature. In the case of the considered countries the representative country with respect to the strength of correlations, so the most influential one is

Sweden, the most representative in the view of the similarities among GDP/c time series appeared Luxembourg and finally the most stable correlation so the best choice for the starting point for preparing prognosis was Poland.

### References

- [1] R.N. Mantegna, *Eur. Phys. J. B* **11**, 193 (1999).
- [2] R.N. Mantegna, *Comput. Phys. Commun.* **121-122**, 153 (1999).
- [3] R.N. Mantegna, E. Stanley, *Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge University Press, Cambridge 1999.
- [4] G. Bonanno, G. Caldarelli, F. Lillo, S. Miccichè, N. Vandewalle, R.N. Mantegna, *Eur. Phys. J. B* **38**, 363 (2004).
- [5] G. Bonanno, F. Lillo, R.N. Mantegna, *Physica A* **299**, 16 (2001).
- [6] G. Bonanno, N. Vandewalle, R.N. Mantegna, *Phys. Rev. E* **62**, R7615 (2000).
- [7] T. Di Matteo, T. Aste, R.N. Mantegna, *Physica A* **339**, 181 (2004).
- [8] H. Markowitz, *J. Financ.* **7**, 77 (1952).
- [9] M. Ausloos, J. Miśkiewicz, *Int. J. Bifurcat. Chaos* **20**, 381 (2010).
- [10] J. Miśkiewicz, *Physica A* **389**, 1677 (2010).
- [11] J. Miśkiewicz, *Physica A* **387**, 6595 (2008).
- [12] J. Miśkiewicz, *Acta Phys. Pol. A* **121**, B89 (2012).
- [13] J. Miśkiewicz, M. Ausloos, *Braz. J. Phys.* **39**, 388 (2009).
- [14] J. Miśkiewicz, M. Ausloos, *Physica A* **389**, 797 (2010).
- [15] J. Miśkiewicz, M. Ausloos, *Physica A* **387**, 6584 (2008).
- [16] J. Miśkiewicz, *Statistical Tools for Finance and Insurance*, Springer Verlag, Berlin 2011, Ch. Distance matrix method for network structure analysis, p. 251.
- [17] R.L. Graham, P. Hell, *Annals of the History of Computing* **7**, 43 (1985).
- [18] A.M. Frieze, *Discrete Appl. Math.* **10**, 47 (1985).
- [19] J.L. Rodgers, W.A. Nicewander, *Am. Statistic.* **42**, 59 (1988).
- [20] M. Sobczyk, *Statistic: practical and theoretical aspects*, Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, Lublin 2006, (in Polish).
- [21] B. Podobnik, H. Stanley, *Phys. Rev. Lett.* **100**, (2008).
- [22] D.D. Wackerly, W. Mendenhall, R.L. Scheaffer, *Mathematical Statistics with Applications*, Cengage Learning, Belmont (CA) 2008.
- [23] A.K. Sharma, *Text Book Of Correlations And Regression*, Discovery Publishing House, Delhi 2005.
- [24] R.S. Tsay, *Analysis of Financial Time Series*, Wiley, New Jersey 2010.
- [25] B. Mirkin, *Core Concepts in Data Analysis: Summarization, Correlation and Visualization*, Springer, London 2011.
- [26] H. Kantz, T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, Cambridge 2003.
- [27] F.O. Redelico, A.N. Proto, P. Clippe, M. Ausloos, *J. Phys., Conf. Series* **221**, 012001 (2010).
- [28] F.O. Redelico, A.N. Proto, M. Ausloos, *Physica A* **388**, 3527 (2009).
- [29] R. Moeckel, B. Murray, *Physica D* **102**, 187 (1997).
- [30] E.F. Krause, *Taxicab Geometry: An Adventure in Non-Euclidean Geometry*, Courier Dover Publications, 1987.
- [31] E. Isaacson, H.B. Keller, *Analysis of Numerical Methods*, Courier Dover Publications, New York 1994.
- [32] C. Woodford, C. Phillips, *Numerical Methods with Worked Examples: Matlab Edition*, Springer, Netherlands 2011.
- [33] W. Homenda, *Algorithms, computational complexity, calculability limits*, Centrum Studiów Zaawansowanych Politechniki Warszawskiej, Warszawa 2009, (in Polish).
- [34] C.H. Papadimitriou, *Computational Complexity*, WNT, Warszawa 2002 (in Polish).
- [35] L. Todorova, B. Vogt, *Physica A* **390**, 4433 (2011).
- [36] W. Li, F. Wang, S. Havlin, H.E. Stanley, *Phys. Rev. E* **84**, 046112 (2011).
- [37] J. Ludescher, C. Tsallis, A. Bunde, *EPL* **95**, 68002 (2011).
- [38] U. Skórnik-Pokarowska, A. Orłowski, *Physica A* **344**, 81 (2004).
- [39] M.J. Naylor, L.C. Rose, B.J. Moyle, *Physica A* **382**, 199 (2007).
- [40] T. Mizuno, H. Takayasu, M. Takayasu, *Physica A* **364**, 336 (2006).
- [41] J. Brida, W. Risso, *Computat. Econom.* **35**, 85 (2010).