MAPPING MACROECONOMIC TIME SERIES INTO WEIGHTED NETWORKS

Mircea_Gligor[1], Marcel_Ausloos[2]

[1]National College “Roman Voda”, Roman-5550, Neamt, e-mail: mrgligor@yahoo.com
[2]GRAPES, University of Liège, Sart-Tilman, B-4000, e-mail: Marcel.Ausloos@ulg.ac.be

Abstract. *The correlations between GDP/capita growth rates of 27 European countries are scanned in various moving time window sizes. The square averaged correlation coefficients are taken as the link weights for a network having the countries as vertices. The network average degree and the weight set variance are found to be monotonic functions on the time window size. The statistics of the weight distributions as well as the adjacency matrix eigensystem are discussed. A new measure of the so called country overlapping is proposed and applied to the network. The ties and clusters are better emphasized through a threshold analysis. The derived clustering structure is found to confirm intuitive or empirical aspects, like the convergence clubs i.e. have a remarkable consistency with the results reported in the actual economic literature.*

Keywords: *fluctuations, correlations, network, clusters*
PACS numbers: 89.65.Gh, 89.75.Fb, 05.45.Tp

1. INTRODUCTION

Modelling the dependences between the macroeconomic (ME) variables has to take into account circumstances that differ substantially from those encountered in the natural sciences. First, experimentation is usually not feasible and is replaced by survey research, implying that the explanatory variables cannot be manipulated and fixed by the researcher. Second, the number of possible explanatory variables is often quite large, unlike the small number of carefully chosen treatment variables frequently found in the natural sciences. Third, the ME time series are short and noisy. Most data have a yearly frequency. When social time series have been produced for a very long period, there is usually strong evidence against stationarity.

Some macroeconomic (ME) indicators are monthly and/or quarterly registered, increasing in this way the number of available data points, but some additional noise is naturally enclosed in the time series so generated (seasonal fluctuations, external and internal short range shocks, etc). This seems to be a solid argument for the fact that the main data sources, at least the ones freely available on the web, tend only to keep the annual averages/rates of growth of the ME indicators.

Let us consider, for example, a time interval of one hundred years, which is mapped onto a graphical plot of 100 data points. From the statistical physics viewpoint, 100 is a quite *small* number of data points, surely too small for speaking about the so called “thermodynamic limit”. On the other hand, from a socio-economic point of view, we can justifiably wonder if a growth, say, of 2% of any ME indicator has at the present time the same meaning as it had one century ago. One must take into account that during that time, the social, politic and economic environment was drastically changed. Moreover the methodology of data collecting and processing is today different from what it was two generations ago. Indeed, the economic world is created by people and is substantially changing from a generation to another one (sometimes also during one and the same generation). Thus, this way of statistical data aggregation turns to be controversial.

On the other hand, an increasing interest in network analysis has been registered during the last decade, particularly due to its potential unbounded area of application. Indeed, the inter-disciplinary (or rather trans-disciplinary) concept of “network” is frequently met in all scientific research areas, its covering field spanning from the computer science to the medicine and social psychology. Moreover it proves to be a reliable bridge between the natural and social sciences, so the recent interest in this field is fully justified.

Using the strong methodological arsenal of the mathematical graph theory, the physicists mainly focused on the *dynamical* evolution of networks, *i.e.* on the statistical physics of growing networks. The remarkable extension from the concept of classical random graph [1] to the one of non-equilibrium growing network [2] allows for accounting the structural properties of random complex networks in communications, biology, social sciences and economics [3, 4]. Indeed, the field of the possible applications seems to be unbounded, it spanning from the “classical” WWW and Internet structures [5, 6] to some more sophisticated social networks of scientific collaborations [7-9], paper citations [10] or collective listening habits and music genres [11].

In most approaches, the Euler graph theory legacy was preserved, especially as regards to the “Boolean” character of links: two vertices can only be either tied or not tied, thus the elements of the so-called adjacency matrix only consist of zeros and ones. However, many biological and social networks, and particularly almost all economic networks, must be characterised by *different* strengths of the links between vertices. This aspect led to the concept of “weighted network” as a natural generalisation of the graph-like approaches. Of course, various ways of attaching some weights to the edges of a fully connected network [12-14]. Some ways to relate the weights to the correlations between various properties of nodes have been proposed in the recent literature [15-18].

The correlation coefficients C_{ij} between two ME time series $\{x_i\}$ and $\{y_j\}$, $i, j = 1, \ldots, N$, is calculated in the present work according to the (Pearson’s) classical formula:

$$C_{ij}(t,T) = \frac{<x_i y_j> - <x_i><y_j>}{\sqrt{<x_i^2 - <x_i>^2><y_j^2 - <y_j>^2>}} \qquad (1)$$

Each C_{ij} is clearly a function both of the time window size T and of the initial time (*i.e.* the "position" of the constant size time window on the scanned time interval). One has to note (or recall) that the correlation coefficients are *not* additive, *i.e.* an average of correlation coefficients in a number of samples does not represent an "average correlation" in all those samples. In cases when one needs to average correlations, the C_{ij} 's first have to be converted into additive measures. For example, one may square the C_{ij} 's before averaging, to obtain the so called *coefficients of determination* (C_{ij}^2) which are additive, or one can convert the C_{ij} 's into so-called *Fisher z* values, which are also additive [19]. The former approach is used here below, so that the average correlations are calculated as:

$$\hat{C}_{ij}(T) = \left[\frac{1}{\nu}\sum_{t=k}^{k+T} C_{ij}^2(t)\right]^{1/2}, k=0,1,\ldots,N-T, \quad (2)$$

where N is the total number of points (the time span), T is the time window size used for the analysis, $\nu = N - T + 1$, and t is a discrete counter variable.

Let us consider that the M agents (countries) which the ME time series refer to, may be the vertices of a weighted network. The weight of the connection between i and j reflects the strength of correlations between the two agents and can be simply expressed as:

$$w_{ij}(T) = \hat{C}_{ij}(T) \quad (3)$$

fulfilling the obvious relations: $0 \leq w_{ij} \leq 1$; $w_{ij} = w_{ji}$ and $w_{ij} = 1$ for $i = j$.

One must stress at this point that the link connecting the vertices i and j does not reflect here either an underlying interaction or a physical/geographical path. Instead, the weight w_{ij} is a measure of the similarity degree between the ME fluctuations in the two countries. The term "fluctuations" refers here to the account of the annual rates of growth of the considered ME indicator.Networks are characterized by various parameters. For instance, the vertex degree is the total number of vertex connections. It may be generalised in a weighted network [13,14] as:

$$k_i = \sum_{\substack{j=1 \\ j\neq i}}^{M} w_{ij} \quad (4)$$

Thus, the average degree in the network is:

$$<k> = \frac{1}{M}\sum_{i=1}^{M}\sum_{\substack{j=1 \\ j\neq i}}^{M} w_{ij} \quad (5)$$

Another describes the number of triangles in the network indicating some correlations. In the literature, there have been several ways to evaluate assortative correlations, such as the assortativity coefficient introduced by Newman [4] that is the Pearson correlation coefficient of the degrees at either ends of an edge. Nonetheless, all of them focus on local degree-correlations between two connected nodes. Here below we will introduce an *overlapping coefficient* in order to indicate some hierarchy in clusters on the network. Yet, the first question is to find whether clusters exist. This will be done through a study of the eigenvalues and eigenvectors of the correlation weights matrix defined here above.

A question of great interest in factor analysis is to evaluate how many factors can be extracted from the eigenvalue spectrum *i.e.* how many *common* factors are underlying the correlation matrix. The Kaiser criterion [20] and the Cattell *scree test* [21] are perhaps the most widely used in this question. According to the former, one can retain only factors with eigenvalues greater than 1. In essence this is like saying that, unless a factor extracts at least as much as the equivalent of one original variable, one has to drop it. The latter test suggests finding the "place" where the smooth decrease of the eigenvalue distribution appears to drop significantly before levelling off to the right of the plot. On the left of this point, presumably, one finds the "factorially significant" eigenvalues. Both methods were found remarkably convergent in [22] when the number of common factors is not too large.

The here below investigated ME indicators are the GDP/capita annual growth rates. Indeed, the GDP/capita is expected to reflect to the largest extent what A. Smith called, over two centuries ago, "the wealth of nations". In fact, it is expected to account both for the economic development and for the people well being. The target group of countries is composed of M = 27 countries belonging to the European Union in 2008. The countries are abbreviated according to The Roots Web Surname List (RSL) [23] which uses 3 letters standardized abbreviations to designate countries and other regional locations. Given the target country group, the World Bank database [24] is here used instead of the more refereed to Penn World Tables [25] in which some data is missing for several East-European countries. In this way, the investigated time span goes from 1993 to 2008.

A general question facing researchers in many areas of inquiry is how to *organize* observed data into meaningful structures. Having computed the square averaged correlation coefficients from Eq. 2, i.e. the adjacency matrix entries w_{ij} in various time window sizes (Eq. 3), some statistical properties of the $\{w_{ij}\}$ dataset are analysed in Section 2. The cumulative distribution function $F(\{w_{ij}\})$ and the kurtosis $K(\{w_{ij}\})$ indicate a shift from a Gaussian distribution (in small size time windows) to a uniform-like one (in large time window sizes). The variance $\sigma^2(\{w_{ij}\})$ and the network average degree $<k>$ (from Eq. 5), are found to display a smooth behaviour when the time window size increases: σ^2 increases roughly linearly while $<k>$ decreases following an inverse cubic root law. The adjacency matrix eigenvalues spectrum is also studied through a time window perspective. The number of "factorially relevant" eigenvalues is emphasized. The problem of the "optimal" time window for ME correlation investigations is addressed in Appendix A.The "best" time window size is usually considered to be the one corresponding to the minimal variance of the output dataset. However, this criterion is proved not to be universal. From the Kolmogorov-Smirnov test and a chi-square test, we find that a 5-8 years time window has an "optimal" size.

From the eigenvectors corresponding to the two largest eigenvalues, a cluster-like structure of the EU-27 countries is built in Section 3, on the basis of the eigenvector components. The EU-27 network along a more geographical perspective is also plotted, by emphasizing the relative importance of the link strength (weight) through a display at different threshold values. The threshold values are chosen according to a significance level derived from the *t*-Student statistic test applied to the $[w_{ij}]$ matrix in Appendix B. The clustering scheme and the network structure are in agreement with results reported in the recent economic literature as regards to "convergence clubs". In particular, the so-called "Scandinavian", "Continental", and "East-European" clusters are identified, as well as the particular position of GBR as the single member of any "Anglo" pattern.

In Section 4 the "clustering" structure of the EU-27 countries is measured through a parameter indicating to what extent a country is "connected" to the *whole* system. Using some new coefficient, O_{ij} which takes into account not only the degrees k_i and k_j, but also the number N_{ij} of *common* neighbours of *i* and *j* vertices, whence called the "overlapping", we show some hierarchy in countries.

A conclusion is drawn in Section 5 emphasizing the main gains of mapping the GDP/capita (and possibly other ME time series) into the weighted network formalism: an increasing explanatory power, a better intuitive understanding and the possibility of using some new analytical tools, in addition to the ones existing in the actual economic literature.

2. DATA ANALYSIS

Having built the adjacency matrix $[w_{ij}]$ (Eq. 3), the first observation one can make is that its entries are functions of the time window size *T* from Eq. 2. Thus, the most important characteristics of the weighted network must be seen as depending on *T* as well. Consider first the cumulative distribution function (CDF) of the weights - elements of the adjacency matrix. The cumulative distribution of the weights is given in Fig. 1 for different time windows, including the minimal (3 years), the maximal (16 years) and two intermediate (5 respectively 10 years). It can be readily seen that the CDF shape is dramatically changing when the time window size changes. For the small size time windows (3 and 5 years) the CDF is close to a Gaussian, while for the large time windows the CDF approaches the shape expected for a uniform CDF.

These changes of the distribution shape can also be pointed out through the kurtosis (*K*) variation with the time window size (Fig. 2). For the Gaussian distribution $K_G = 0$, while for the discrete uniform distribution of *m* data (m = 300 here) it can be calculated [19] as:

$$K_U = -\frac{6}{5}\frac{m^2+1}{m^2-1} \approx -\frac{6}{5} \qquad (6)$$

It is found on Fig. 2 that the *K* value shifts between the limit K_G and K_U indeed.

Taking into account the above results one may conclude that the distribution of the adjacency matrix entries w_{ij} ($i > j$) becomes flatter and flatter when the time window increases, shifting from the Gaussian-like shape to the uniform-like distribution. Some statistical tests are done in Appendix A.

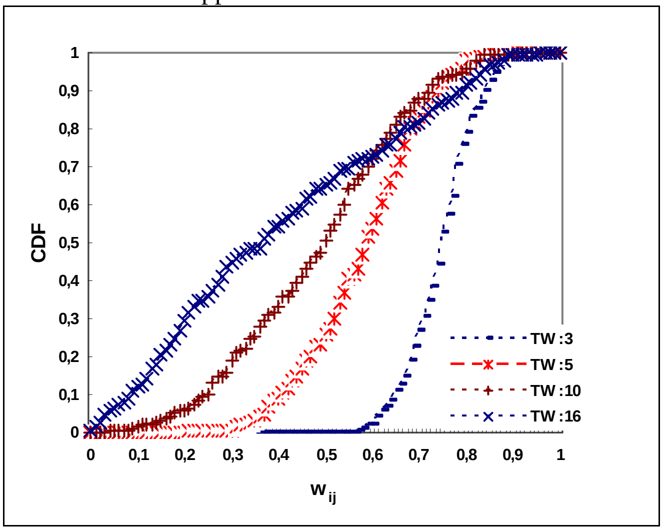


Fig. 1 The CDF of the weights set $\{w_{ij}\}$ for four different time window sizes

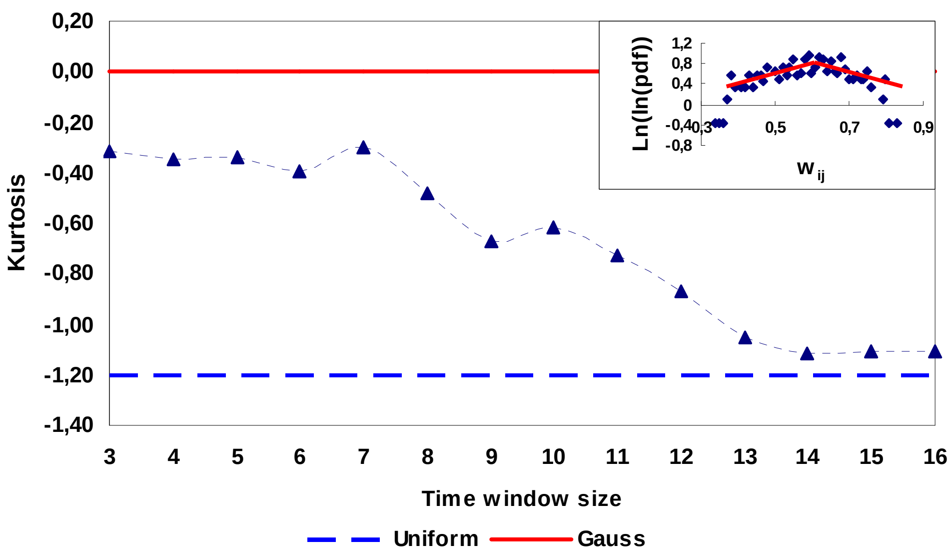


Fig. 2 The kurtosis coefficient of the weights set $\{w_{ij}\}$ versus the time window size. Inset: the double logarithm of the w_{ij}'s probability density function for 5 years time window size. The thick line has a ±2 slope, corresponding to the Gaussian distribution.

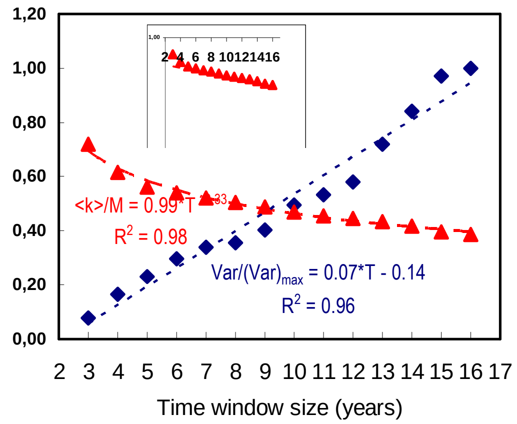


Fig. 3 The average degree $<k>/M$ and the variance of the weights set $\{w_{ij}\}$ in the EU-27 network versus the time window size. Variance is normalised to its maximal value. Inset: $<k>/M$ versus T in log-plot for emphasizing the inverse cubic root law.

The average degree of the EU-27 weighted network is plotted in Fig. 3 for all possible time window sizes with which the time span 1993-2008 can be scanned. One can see that the decreasing of $<k>$ with the time window size T is well fitted by a power law: $<k> \sim 1 / T^{1/3}$.

The somewhat unexpected behaviour, i.e., the monotonic (roughly linear) increase of the variance with the time window size can be understood through the change in CDF shape. A possible explanation of this behaviour is based on the number of common factors underlying the correlation coefficients.

As the adjacency matrix of the EU-27 weighted network is in fact a squared-averaged correlation matrix of the GDP/capita growths, it is natural to ask for the interpretation of its eigensystem.

The six largest eigenvalues are plotted in Fig. 4 for each possible moving time window size scanning the time span 1993-2008. As mentioned in the Introduction, the Kaiser criterion suggests to evaluate the number of common factors taking into account the eigenvalues having a percent contribution to the total variance at least $1/M$ = 1/27 (the continuous horizontal line in Fig. 4)

At first sight one can see that the first eigenvalue contribution to the total variance monotonically decreases when the time window size increases, while the other eigenvalues contribution becomes more and more significant. Moreover, the cumulated contribution of the first two largest eigenvalues decreases from 80% for T = 3 years, to 64% for T = 16 years. Therefore, for the small size time windows (3-5 years) two common factors may be accounted for, while in the largest time windows the number of common factors increases to six.

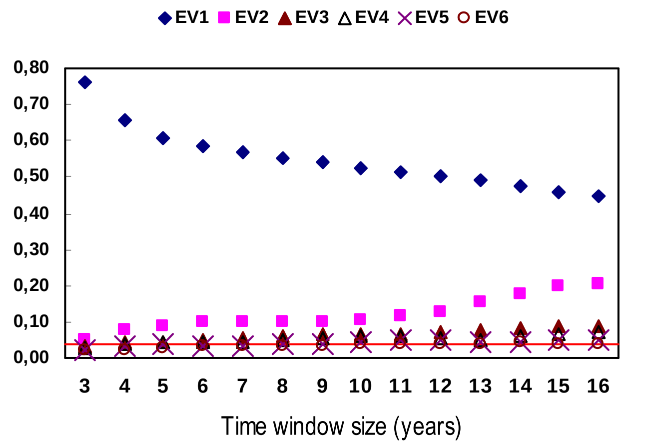


Fig. 4 The six largest eigenvalues (EV) of the adjacency matrix $[w_{ij}]$ for the EU-27 weighted network versus the moving time window size. The eigenvalues are normalized the correlation matrix size (M = 27), thus the vertical axis may be read as a fractional contribution to the total variance.

This finding reinforces the results of the previous subsection on the change in shape of the CDF. When the number of common factors is small, the correlation coefficients are grouped around the mean value, leading to the Gaussian-like distribution shape. On the contrary, when many common factors (economic, social, political) are accounted for, the correlations between the GDP/capita rates of growth tend to cover the whole interval between 0 and 1 (in absolute values), and a uniform-like distribution emerges.

3. THE CLUSTERED WEIGHTED NETWORK OF EU-27 COUNTRIES

Since the eigenvectors corresponding to the largest eigenvalues of the correlation matrix are usually expected to be those carrying the most useful information, a cluster-like structure of the EU-27 countries is built in Fig. 5 on the basis of the structure of the first two eigenvectors.

One can easily see (Fig. 5) that a multi-polar structure exists: the "Continental" group (l.h.s., up) and the "Scandinavian" group (l.h.s., middle) are somewhat apart from each other. An extreme position is taken by GBR (r.h.s., down) which appears as the single member of any "Anglo" pattern, since the other OECD representatives of a (supposed to be) anglo-convergence club, *e.g.* U.S.A., Canada and Australia [26, 27]), are missing from our study. Another interesting aspect pointed out by Moran in [27] is also found here, i.e. IRL has a non appartenence to the "Anglo" cluster, but is rather in the "Scandinavian" group and close to the "Continental" one.

For the first time, i.e. as a complementary addition to previous investigations on the subject, we observe an emerging East European convergence club (r.h.s., down), tying to Scandinavian and Continental group through a HUN – POL line. Observe some "Mediterranean" clustering as well.

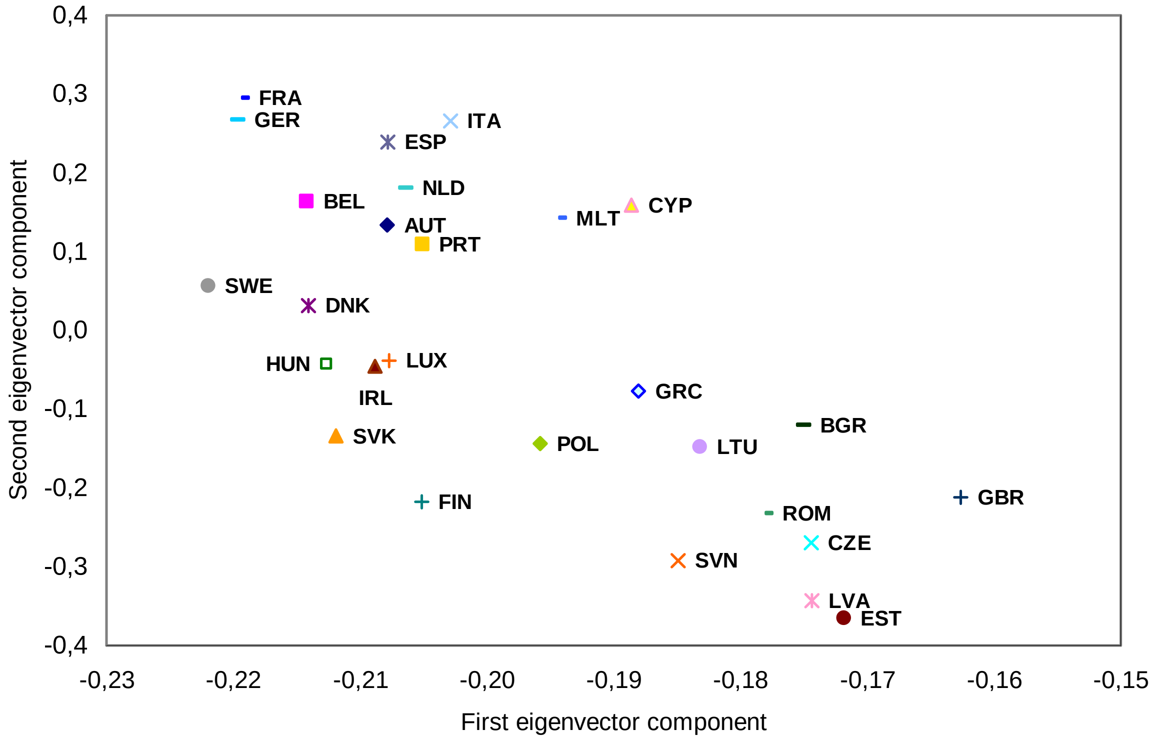
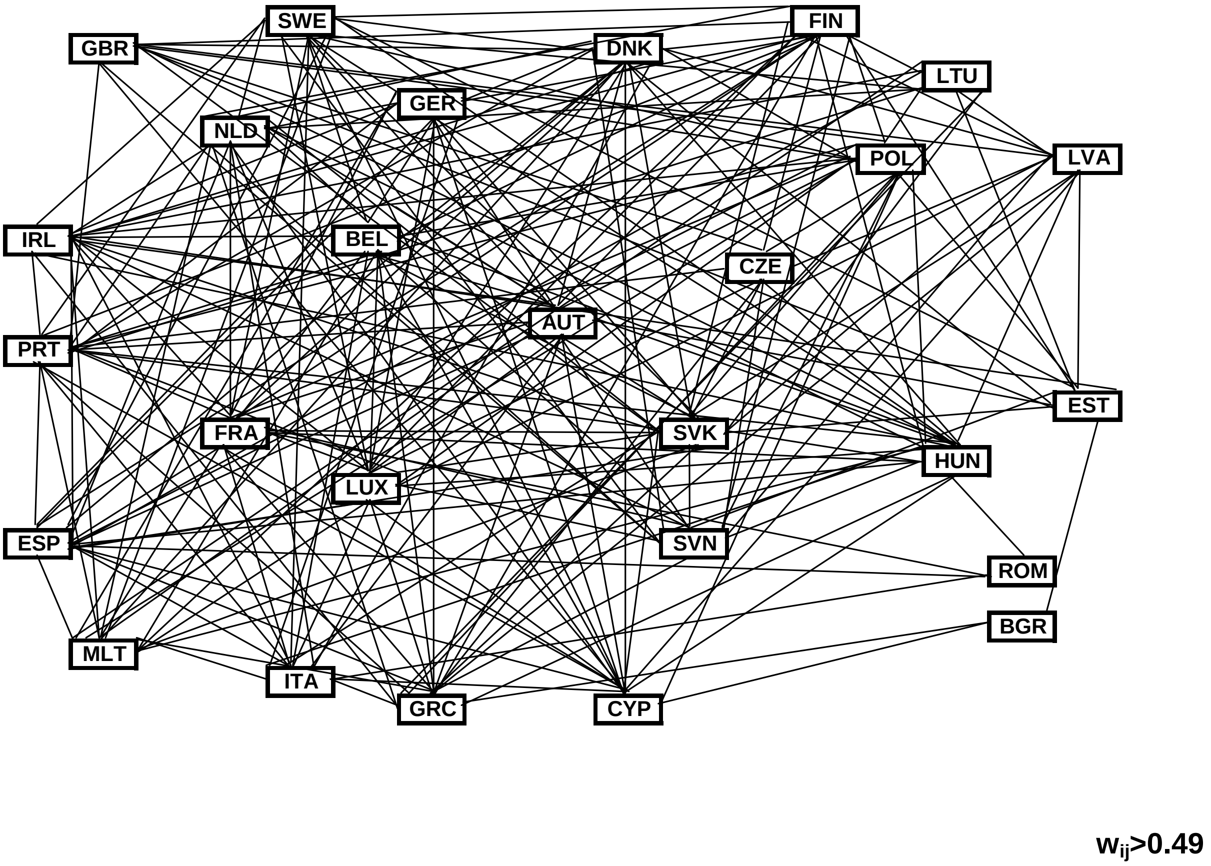


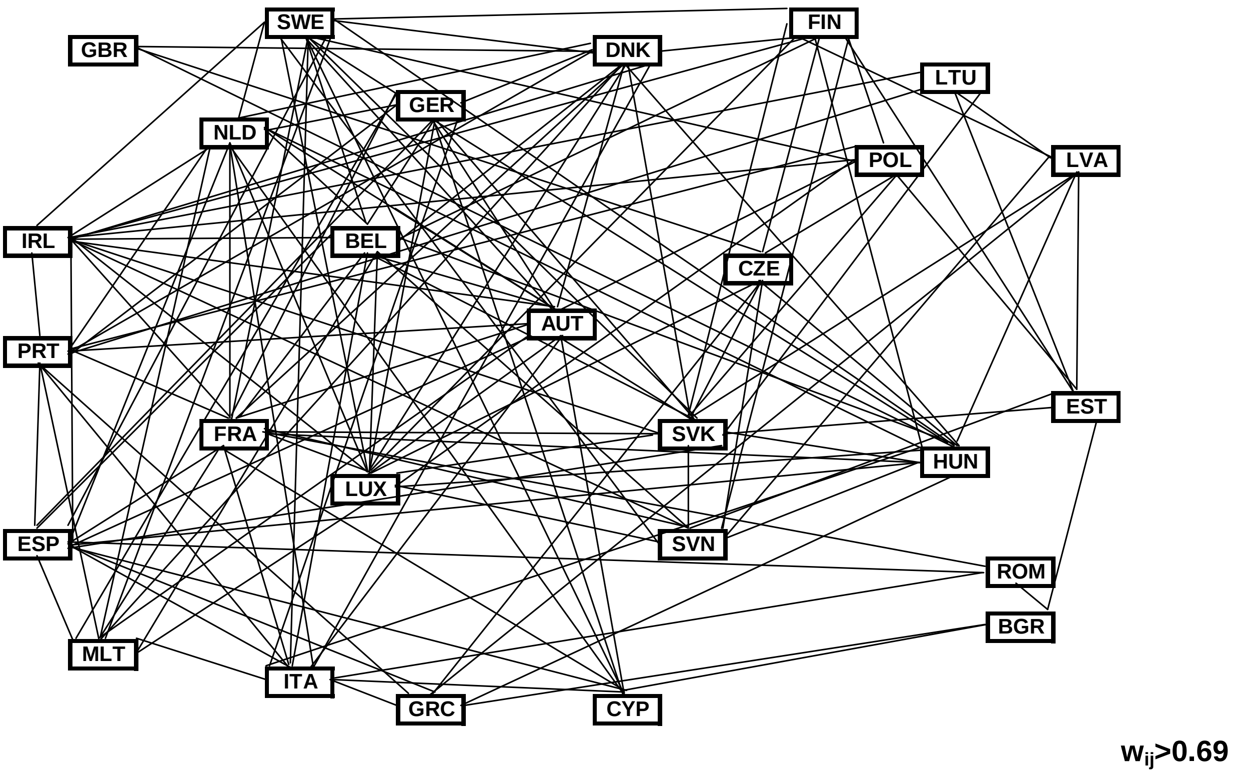
Fig. 5 The cluster-like structure of the EU-27 countries according to the GDP/capita rates of growth. The country coordinates are the corresponding eigenvector components of the EU-27 weighted network adjacency matrix $[w_{ij}]$.

The clustering scheme in Fig. 5 is in agreement with results reported in the recent economic literature as regards the so-called "convergence clubs" across the Western Europe, *i.e.* groups of economies that present a homogeneous pattern and converge towards a common steady state [26-30]. In particular, in [26] it has been showed that Sweden, (Norway) and Denmark, registered a similar level of income mobility while in [27] three distinct patterns of development and income distribution, indeed called "Continental", "Anglo" and "Scandinavian", have been found by examining a group of 17 OECD economies during the two decades before 2000. In the same idea, "a high degree of heterogeneity in preferences for redistribution across four clusters of different systems of social protection of OECD countries" namely Scandinavian, Continental, Anglo-Saxon and Mediterranean has been reported in [30].

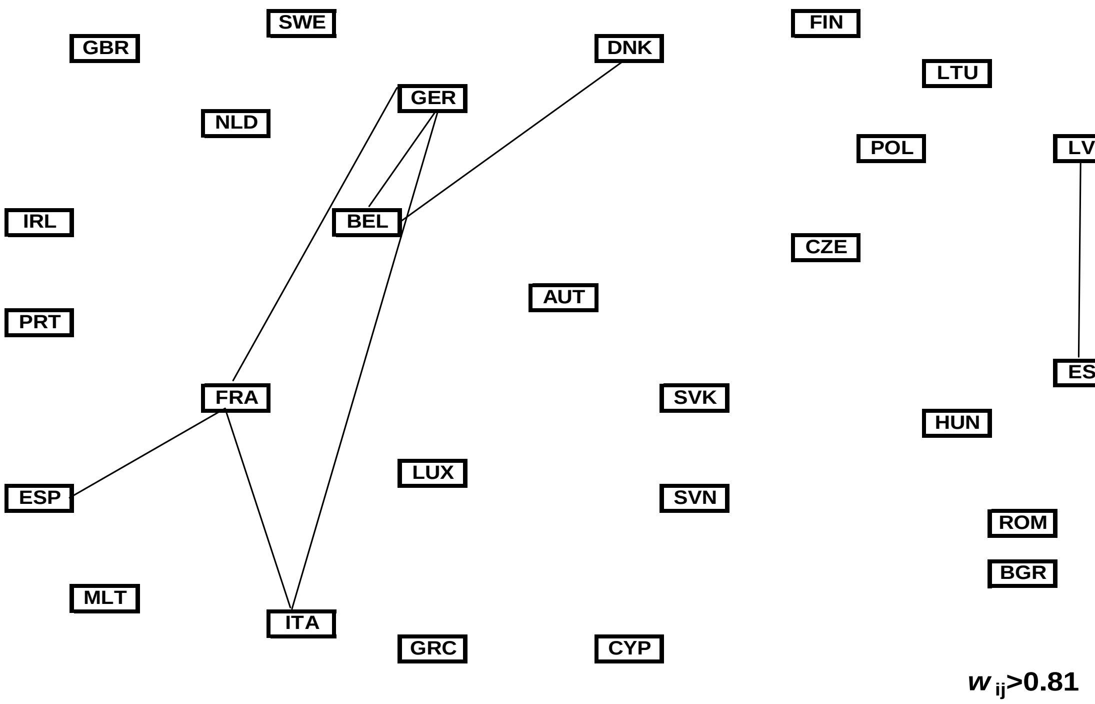
Finally, the adjacency matrix $[w_{ij}]$ can be used to plot the EU-27 network along a more geographical perspective. The network is, obviously, fully connected; it is of interest to observe the relative importance of the link strength (weights) through a display at different threshold values. In the subsequent figures, only the links having the corresponding weights greater than a certain threshold value, w_α are taken into account. This threshold value is *a priori* chosen according to a significance level derived from the *t*-Student statistic test applied to the $[w_{ij}]$ matrix (Appendix B). The resulting networks are plotted in Figs. 6a, 6b, 6c for the $T = 5$ years moving time window size.


SWE
FIN
GBR
DNK
LTU
GER
NLD
POL
LVA
IRL
BEL
CZE
AUT
PRT
EST
FRA
SVK
HUN
LUX
ESP
SVN
ROM
BGR
MLT
ITA
GRC
CYP
w_{ij}>0.49

(a)


SWE
FIN
GBR
DNK
LTU
GER
NLD
POL
LVA
IRL
BEL
CZE
AUT
PRT
EST
FRA
SVK
HUN
LUX
ESP
SVN
ROM
BGR
MLT
ITA
GRC
CYP
w_{ij}>0.69

(b)



(c)

Fig. 6 The EU-27 weighted network for three different weight thresholds: (a) w_T=0.49; (b) w_T=0.69; (a) w_T=0.81.

Fig. 6a includes all four "convergence clubs" above discussed. The single element "Anglo" club is well isolated, as seen already in Fig. 6b; the "Scandinavian" and "East European" clusters become isolated for a higher threshold, as seen in Fig. 6c. It is also remarkable to observe the decreasing number of long-range links when going from Fig. 6a to Fig. 6b and further to Fig. 6c. Even if the actual geographic, investment and trade inter-country ties are not explicitly considered in our study, the degree of similarity of the country GDP/capita *fluctuations* well supports the evidence of the so-called "regionalization" [28, 29].

4. THE COUNTRY OVERLAPPING HIERARCHY

The previous results lead to consider the hierarchical or "clustering" structure of the EU-27 countries. For the purpose of describing this aspect we introduce a quantity which is able to measure to what extent a country is "connected" to the whole system. The idea, first hereby applied to a non-weighted network, is to construct a country hierarchy using some new coefficient, O_{ij} which takes into account not only the degrees k_i and k_j, but also the number N_{ij} of the *common* neighbours of i and j vertices,. This coefficient O_{ij} is here called "overlapping" of i and j vertices (in spite of the fact that this term has already been assigned various meanings in the network literature).

Firstly, for a non-weighted network consisting of M vertices, O_{ij} must satisfy the following properties:

(1) $O_{ij} = 0 \Leftrightarrow N_{ij} = 0$ (fully disconnected, or "tree-like" network)

(2) $O_{ij} = 1$, $\forall i \neq j$ in a fully connected network, where $N_{ij} = M - 2$; $k_i = k_j = M - 1$;

(3) $0 < O_{ij} < 1$, otherwise;

(4) $O_{ij} \sim N_{ij}$ and $O_{ij} \sim < k_{ij} > \equiv (k_i + k_j)/2$.

A quantity satisfying all these conditions (1)-(4) can be defined as:

$$O_{ij} = \frac{N_{ij}(k_i + k_j)}{2(M-1)(M-2)} \ , \ i \neq j. \qquad (7)$$

For a weighted network,Eq.7may be generalised as:

$$O_{ij} = \frac{1}{2(M-1)(M-2)} \sum_{\substack{l=1 \\ l \neq i,j}}^{M} (w_{il} + w_{jl}) \left(\sum_{\substack{p=1 \\ p \neq i}}^{M} w_{ip} + \sum_{\substack{q=1 \\ q \neq j}}^{M} w_{jq} \right), i \neq j. \qquad (8)$$

One can easily see that $0 < O_{ij} < 1$, and $O_{ij} = 1$ only for all $w_{ij} = 1$, *i.e.* fully connected non-weighted network. However, for a weighted network, O_{ij} can never be zero.

Each overlapping coefficient is thus computed for each EU-27 country using the adjacency matrix defined in Eq. 3. A country average overlapping index $<O_i>$ can be next assigned to each country, dividing the sum of its overlapping coefficients by the number of neighbours:

$$< O_i > = \frac{1}{M-1} \sum_{j=1}^{M} O_{ij} \qquad (9)$$

The results are shown in Table 1.

Table 1 The country average overlapping index of each EU-27 country

SWE	**0.38**	**NLD**	**0.35**	**CYP**	**0.32**
DNK	**0.37**	**AUT**	**0.35**	**SVN**	**0.32**
GER	**0.37**	**FIN**	**0.35**	**CZE**	**0.31**
FRA	**0.37**	**POL**	**0.35**	**ROM**	**0.31**
HUN	**0.37**	**ESP**	**0.35**	**BGR**	**0.31**
SVK	**0.37**	**PRT**	**0.35**	**LTU**	**0.31**
BEL	**0.36**	**ITA**	**0.34**	**LVA**	**0.31**
IRL	**0.36**	**MLT**	**0.33**	**EST**	**0.30**
LUX	**0.36**	**GRC**	**0.33**	**GBR**	**0.29**

The highest values of the average overlapping index correspond to the countries belonging to the "Continental" and "Scandinavian" groups, while the lowest values correspond to several countries forming an East European cluster. Again, the separate position of GBR as a single representative of the "Anglo" pattern, with respect to European economies is emphasized, the former being in fact a cluster by itself.

One has to note a remarkable similarity between the country ranking over the first eigenvector component (Fig.5) and the ranking over the country average overlapping index (Table 1). This similarity proves the ability of the hereby introduced $<O_i>$ index to supply a correct description of the country weighted network.

5. CONCLUSIONS

The present paper has shown the possibility of mapping a macroeconomic time series, namely the GDP/capita rates of growth into a weighted network. The considered vertices are the 27 countries belonging to the EU community in 2008 and the weights assigned to the links are the correlation coefficients. An averaging has been performed over the squared values obtained when a constant size time window is moved with a constant time step over the scanned time interval (1993-2008).

Usually, the correlation coefficients are computed in various time windows, with a given size; the first problem brought into discussion here above and outlined in Appendix A has been the role played by the time window size. In particular, the variance of the weights dataset has been found to be a monotonically increasing function of the time window size. This unusual result reflects the weight distribution shifting from a Gaussian to a uniform-like shape when the time window size used for data analysis increases. This transition has been explained when analysing the eigenvalue spectrum of the adjacency matrix in various time windows: as the time windows size increases, more and more common factors must be taken into account as underlying the adjacency matrix, so that the correlation coefficients (absolute) values cover almost uniformly the interval between 0 and 1.

Finally, we have to point out that the mapping of the GDP/capita and other macroeconomic time series into a weighted network structure allows a direct visualisation of the inter-country connections from at least three different viewpoints: (a) as relative distances in the bi- or multi-dimensional space of the adjacency matrix eigenvalues; (b) as statistical significant edges in the graph plot; (c) as relative positions in the country averaged overlapping coefficients based hierarchy. In all these three ways, the derived clustering structure is found to have a remarkable consistency with the results reported in the actual economic literature. In the future, other network multi-vertex characteristics (clustering, minimal path, centrality, etc.) may be expected to be studied in order to show whether they play an important role in a better understanding of economic connections.

APPENDIX A: IS THERE AN OPTIMAL TIME WINDOW?

A Kolmogorov-Smirnov test has been performed over the CDF (Fig.1) corresponding to every time window size [31]. The *p*-values are found to be small (0.12 – 0.16) for 3 and 4 years as well as for 9, 10 and 11 years time window sizes; some large *p*-values are obtained for the range 5 – 8 years (0.34-0.48) and drop to 0.01 and 0.00 for 12-16 time window size. As generally accepted ([31], [32]), the null hypothesis is rejected when *p*-values are smaller than 0.10. Thus, one can conclude that the hypothesis of Gaussian distribution is rejected for the time windows larger than 11 years.

The χ^2 statistical test has been performed in contrast against the hypothesis of a uniform distribution. The standard confidence intervals are found to be less than 1% for the 3 and 5 years time window sizes, while for the 10 and 16 years time window sizes they are found to be at 85% and 90% respectively.

As regards the problem of the "optimal" time window, one must firstly recall that usually, the "best" time window size is considered to be the one corresponding to the minimal variance of the output dataset. However, this criterion is not universal. From the Kolmogorov-Smirnov test we have derived that for the 5-8 years time window sizes the corresponding distributions are "more Gaussian" than for the small time windows of 3 and 4 years. Moreover, in the 3 and 4 years window sizes the same statistical test points out to a relatively large number of "outliers" which may be seen as spurious correlations.

In view of the above considerations, we conclude that the "optimal" time window sizes are situated in the interval 5-8 years. That is why some particular results in the Sections 3-5 are derived from a constant size time window of 5 years, for which the distribution of $\{w_{ij}\}$ set is Gaussian, at least in its central part (see Fig. 3, inset).

APPENDIX B: THE T-STUDENT'S TEST APPLIED TO THE $[w_{ij}]$ MATRIX

The linear relationship between two variables can be tested using *t*-statistics [32] by computing:

$$t = w_{ij}\sqrt{\frac{n-2}{1-w_{ij}^2}}, \qquad (10)$$

where $n - 2$ is the number of degrees of freedom. The correlation (weight) w_{ij} is considered to be statistically significant if the computed t value is greater than a critical value t_α *read* from the t-Student's distribution table for the α level of significance.

From Eq. (6), if $w_{ij} = w_\alpha$ and $t = t_\alpha$ one gets:

$$w_\alpha = \frac{t_\alpha}{\sqrt{t_\alpha^2 + n - 2}} \qquad (11)$$

Taking n = 5 (the number of statistical data used for computing each correlation coefficient in the 5 years time window size), from the t-Student distribution tables we find the critical values $t_{\alpha 1}$ = 0.98; $t_{\alpha 2}$ = 1.64 and $t_{\alpha 3}$ = 2.35 for the levels of significance α_1 = 0.4; α_2 = 0.2 and α_3 = 0.1; (or, equivalently, 60%, 80% and, respectively, 90% confidence intervals). The corresponding threshold values are $w_{\alpha 1}$ = 0.49; $w_{\alpha 2}$ = 0.69 and $w_{\alpha 3}$ = 0.81, they are therefore used as threshold for the display (Fig. 6a, 6b, 6c).

6. REFERENCES

[1] Erdös P. & Rényi A. (1959). *On random graphs*. Publications Mathematicae, 6, 290.

[2] Albert R. & Barabási A.-L. (2002). *Statistical Mechanics of complex networks*. Review of Modern Physics 74, 47-97.

[3] Dorogovtsev S.N & Mendes J.F.F. (2003). *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press, Oxford.

[4] Newman M.E.J. (2003). *The structure and function of complex networks*. SIAM Review 45, 167-256.

[5] Huberman B.A. & Adamic L.A. (1999). *Growth dynamics of the World-Wide Web*. Nature 401, 131.

[6] Pastor-Satorras R & Vespignani A. (2004). *Evolution and Structure of the Internet: A Statistical Physics Approach*. Cambridge University Press, Cambridge.

[7] Barabási A.-L., Jeog H., Néda Z., Ravasz E., Schubert A., Vicsek T. (2002). *Evolution of the social network of scientific collaborations*. Physica A 311, 590-614.

[8] Newman, M.E.J. (2001). *Scientific collaboration networks: I. Network construction and fundamental results*. Physical Review E 64, 016131; *Scientific collaboration networks: II. Shortest paths, weighted networks, and centrality*. Physical Review E 64, 016132

[9] Lambiotte R. & Ausloos M. (2005). *N-body decomposition of bipartite networks*. Physical Review E 72, 066117.

[10] Redner S. (1998). *How popular is your paper? An empirical study of citation distribution*. European Physical Journal B, 4**,** 131.

[11] Lambiotte R. & Ausloos M. (2005). *Uncovering collective listening habits and music genres in bipartite networks*. Physical Review E 72, 066107.

[12] Yook S.H., Jeong H., Barabási A.-L., Tu Y. (2001). *Weighted evolving networks*. Physical Review Letters **86**, 5835.

[13] Barrat A., Barthelemy M., Pastor-Satorras R., Vespignani A. (2004). *The architecture of complex weighted networks*. Proc. Natl. Acad. Sci. USA 101, 3747–3752.

[14] Newman M.E.J. (2004). *Analysis of weighted networks*. Physical Review E 70, 056131.

[15] Gligor, M. & Ausloos, M. (2007). *Cluster structure of EU-15 countries derived from the correlation matrix analysis of macroeconomic index fluctuations*. European Physical Journal B 57, 139-146.

[16] Gligor, M. & Ausloos, M (2008). *Convergence and Cluster Structures in EU Area according to Fluctuations in Macroeconomic Area*. Journal of Economic Integration 23, 297-330.

[17] Ausloos, M. & Gligor, M (2008). *Cluster Expansion Method for Evolving Weighted Networks Having Vector-like Nodes*. Acta Physica Polonica A 114(3), 491-499.

[18] Gligor, M. & Ausloos, M (2008) *Clusters in weighted macroeconomic networks: the EU case*. European Physical Journal B 63, 533-539.

[19] Hays W.L. (1988). *Statistics*. New York: CBS College Publishing.

[20] Kaiser H.F. (1960). *The application of electronic computers to factor analysis*. Educational and Psychological Measurement 20, 141-151.

[21] Cattell R.B. (1966). *The scree test for the number of factor*. Multivariate Behavioral Research, **1**, 245-276.

[22] Browne M.W. (1968). *A comparison of factor analytic techniques*. Psychometrika 33, 267-334.

[23] http://helpdesk.rootsweb.com/codes/

[24] http://devdata.worldbank.org/query/default.htm

[25] http://pwt.econ.upenn.edu/

[26] Aaberge R., Bjorklund A., Jantti M., Palme M., Pedersen P.J., Smith N., Wennemo T. (2002). *Income Inequality and Income Mobility in the Scandinavian Countries Compared to the United States*. Review of Income and Wealth, 48, 443-469.

[27] Moran T.P. (2005). *Bootstrapping the LIS: Statistical Inference with the Gini Index and Patterns of Inequality in the Global North*. Paper for "International Conference in Memory of Two Eminent Social Scientists: C. Gini and M. O. Lorenz", Siena, Italy, 23-26 May, 2005.

[28] Durlauf S.N. & Quah D.T. (1999). *The new empirics of economic growth*. In *Handbook of Macroeconomics*. Elsevier, North Holland, 231–304.

[29] Mora T. (2005). *Evidencing European regional convergence clubs with optimal grouping criteria*. Applied Economics Letters 12, 937-940.

[30] Angelini E.C. & Farina F. (2005). *The size of redistribution in OECD countries: does it influence wave inequality?* Paper for "International Conference in Memory of Two Eminent Social Scientists: C. Gini and M. O. Lorenz", Siena, Italy, 23-26 May, 2005.

[31] The online software used: http://www.physics. csbsju.edu/ stats/KS-test.html

[32] Box G.E.P., Hunter W.G., Hunter S. J. (1978). *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*. Wiley, New York.