International Journal of Basic and Applied Sciences, 11 (1) (2022) 9-16

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS

Research paper

Principal centrality measures: a comprehensive approach to the Spanish stocks market

Amauri Gutiérrez

Strategy, Marketing, Sales and Data Science, International Private Bank - Deutsche Bank
Paseo de Castellana 18, 28046, Madrid, Spain

E-mail: amaurigh@yahoo.es

Abstract

In social network analysis, for determining the relevance or significance of a node in the network, several node centrality measures are often used such as degree centrality, betwenness centrality, closeness centrality, eigenvector, subgraph and page rank centrality. In this paper we apply a principal components analysis over the traditional centrality measures for obtaining an overall single metric that combines the best attributes of the traditional centrality measures and permits to detect relevant nodes in the network. Concretely, a detailed study of the Spanish stocks market will be used for demonstrating the advantages of this approach.

Keywords: *social network analysis; centrality measures*

1. Introduction

Stocks evolution is a consecuence of a very large set of complex causes. Among other aspects, causes can vary from the economical situation of the country, the evolution of economical sectors, the strategies and results of the companies until social and customers behaviour and expectations. Sometimes, the evolution of certain stocks may produce an impact (negative or positive) in other stocks and these interactions can be measured along the time. From technical point of view, interactions between stocks can be represented in terms of correlations between them and the overall view of the entire stock market can be represented as a (network) graph, whose nodes are stocks and weighted edges whose weights are the correlation coeficient between every two nodes.
In 1999, Mantegna [7] introduced the usage of the minimal spanning tree (MST) as an important tool for filter the relevant information contained in sotcks network. Since then MST has been widely applied for the analysis of complex networks in several fields such as complex systems, politics, organizational management and, in particular in the financial industry for dealing with topics such as portfolio analysis [8] and risk assessment [10].
Evidently, not all stocks in the market have the same role or influence on the rest of stocks. One of the most important tasks researchers are interested in is the identification of those stocks with a special role or inflence. For doing this, some relative node measures have been recently introduced. Most well-known node measures are the so-called centrality measures such as degree, closeness, betweeness, eigenvector centralities, see [3, 6, 8].
The objective of this paper is to introduce a novel centrality measure. This novel centrality measure is a result of a combination of the existing centrality measures that will be shown in next sections. For demonstrating the advantages that this new centrality measure has, we analyze the network of 110 stocks traded at Madrid stock exchanges (BME and MAB). Data is gathered by using an Open Source Python package (investpy) that facilitates to retrieve financial data from Investing.com. Details on the package can be found at https://www.kaggle.com/alvarob96/financial-data-retrieval-using-investpy. The 100 selected stocks for the study have the time range from January 1, 2019 to December 31, 2021.

2. Definitions

Networks stocks analysis tipically starts [Mantegna] by applying the following procedure:
1) Calculation of the the logarithm of price return of stock $i \in 1,\ldots,N$ at time t,

$$R_i(t) = ln(P_i(t) - ln(P_i(t-1)),$$

where $P_i(t)$ is the stock closure price.
2) Calculation of the the $N \times N$ correlation matrix C, where the correlation coeficient c_{ij} between stocks j and j is

$$c_{ij} = \frac{\langle R_i \cdot R_j \rangle - \langle R_i \rangle \langle R_j \rangle}{\sqrt{(\langle R_i^2 \rangle - \langle R_i \rangle^2)(\langle R_j^2 \rangle - \langle R_j \rangle^2}}$$

where $\langle R_i \rangle$ is the average of $R_i(t)$.
3) Construction of the $N \times N$ distance matrix D, where $d_{ij} = \sqrt{2(1 - c_{ij})}$.
4) Construction of the minimum spanning tree (MST) by applying Kruskal´s algorithm.
5) Based on the MST, apply centrality measures to calculate the relative relevance of each stock.

Commonly, main node centrality measures used in previous step 5) are degree, closeness, betweeness and eigenvector centrality. Definitions are as follows:
a) Degree centrality $DC(i)$ for any stock i,

$$DC(i) = \frac{\sum_{j=1}^{N} A_{ij}}{N-1},$$

where $A_{ij} = 1$ if there is a link in the MST between stock i and stock j, and 0 otherwise.
b) Closeness centrality $CC(i)$ for any stock i, is the ratio of the sum of the number of links in the path from i to j for all $j \neq i$, and the number of links in the MST, which is $n-1$, concretely

$$CC(i) = \frac{\sum d(i,j)}{n-1},$$

where $d(i,j)$ is the distance (length of shortest path) between nodes i and j.
c) Betweeness centrality $BC(i)$, for any stock i, is the ratio of the number of path passing through i between two different nodes and the number of all possible paths from j to k for all j and k where $j \neq i$ and $k \neq i$, more specifically,

$$BC(i) = \sum_{j \neq i \neq k} \frac{\sigma_{jk}(i)}{\sigma_{jk}},$$

where σ_{jk} is the total number of shortest paths from node j to node k, and $\sigma_{jk}(i)$ is the number of those paths that pass through i.
d) Eigenvector centrality $EC(i)$, for any stock i,

$$EC(i) = \frac{1}{\lambda_{max}} \sum_{j=1}^{N} A_{ij} \cdot e_j,$$

where $(e_1, e_2, \ldots, e_N)^t$ is the eigenvector associated to the largest eigenvalue λ_{max} of the adjacency matrix A.
e) Subgraph centrality $SC(i)$, for any stock i,

$$SC(i) = \sum_{k=0}^{\infty} \frac{\mu_k(i)}{k!},$$

where $\mu_k(i)$ is the number of closed walks of length k starting and ending node v in the graph, ie, the local spectral moment or order k.
f) Page rank $PR(i)$, for any stock i,

$$PR(i) = \sum_{j \in \Gamma(i)} \frac{PR(j)}{\deg(j)},$$

where $\Gamma(i)$ is the set of nodes connected to i, and $deg(j)$ is the degree of the node j.
Carefully analyzing previous definitions, it can be derived that degree centrality indicates the connectivity of nodes. It give us the information on how many other nodes are connected with a particular node. Closeness centrality measures how close a node is to all other nodes in terms of correlations. Closeness can also be regarded as a measure of how long the information can be spread from a given node to other reachable nodes. Betweenness centrality indicates the extent to which a node lies in relative position with respect to the others This measure indicates the potentiality of node to influence the others. And finally, eigenvector centrality indicates how an important node is connected to other important nodes (stocks).
Previous centrality measures explain some, but not all aspects of the influence or relevance of a node within the entire network. Individually considered, these measures are not be able to explain how relevant is a particular node. In order to mitigate this fact, we introduce a novel centrality measure, ie, *principal centrality measure*, that combines previous centrality measures. Combination is made by applying principal component analysis procedure and technical definition is as follows. g) r-Principal centrality $PC_r(i)$, for any stock i. Let X be an $N \times 6$ matrix, in which the elements of each row i, are

$$(DC(i), CC(i), BC(i), EC(i), SC(i), PR(i)),$$

ie, the traditional centrality measures of node (stock) i, $1 \leq i \leq N$. Based on principal components analysis (PCA) procedure, we define the *r-principal centrality measure* of the node i, $PC_r(i)$, $1 \leq r \leq 6$, as

$$PC_r(i) = \left(\sum_{j=1}^{r} z_{ij}^2 \right)^{\frac{1}{2}}$$

where z_{ij} is the element at (i,j) coordinate of the transformed matrix Z in the new (PCA) vectorial space, such that $Z = X \cdot B^T$. In this case, B is an orthonormal 6×4 matrix, whose k-th column is the k-th eigenvector b_k of the matrix S, where S is the covariance matrix of the columns of X.

3. Methodology

Using a Python program, we have obtained the minimum spanning tree (MST) representing the Spanish stocks network. See Figure 1.
In the graph, stocks are represented by nodes and edges represent the correlation betwen two sotcks. In the figure, each stock is labeled with the tick symbol of the stock. Minimum spanning tree illustrates the relationship between stocks that are highly correlated. although there is no single dominant node (stock), it can be seen some interesting facts. There are some stocks clusters diminated by REP in the top of the figure, other one dominated by FER in the bottom left of the figure and another one dominated by SAN in the central right of the figure. In general, researchers identify such dominating nodes by using traditional centrality measures.

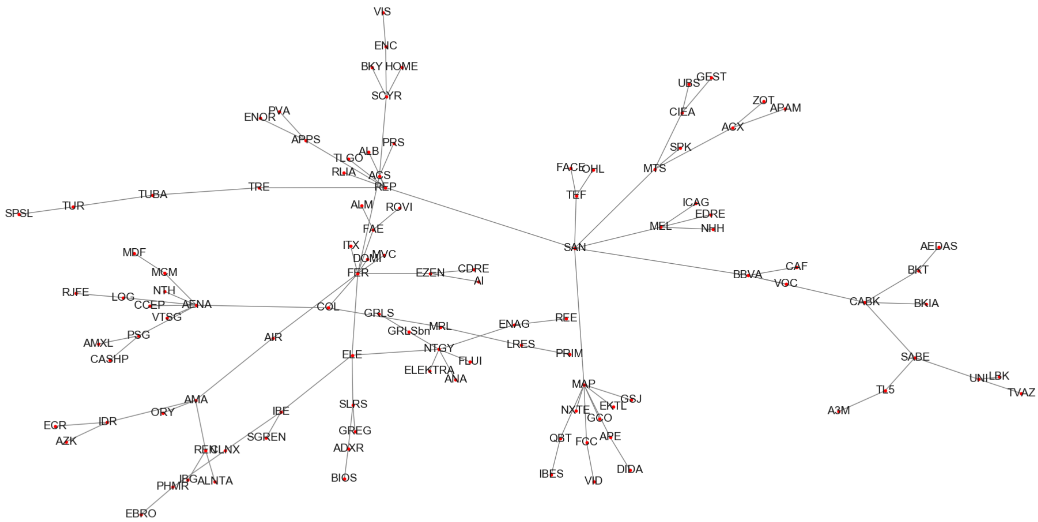

Figure 1: Minimum spanning tree of the Spanish stocks market

Based on the MST, the calculations of the six (degree, closeness, betweenness, eigenvector, subgraph and pagerank) traditional centrality measures for a sample of 20 stocks are depcted on Table 1.

	DC	CC	BC	EC	SC	PR
MAP	0.094983	0.228604	0.232174	0.527848	11.081115	0.040754
FER	0.091593	0.254899	0.638746	0.098405	11.468918	0.037861
SAN	0.079833	0.263381	0.689433	0.424127	12.479398	0.033878
REP	0.073787	0.223742	0.158290	0.260941	7.930440	0.032753
SCYR	0.056142	0.183561	0.081186	0.022640	5.010546	0.025417
ACS	0.053417	0.217125	0.175473	0.046769	5.667266	0.022670
AENA	0.052731	0.262261	0.532216	0.056356	5.411559	0.023394
CIEA	0.044652	0.185633	0.061211	0.051862	3.897221	0.021221
CABK	0.041228	0.228822	0.119845	0.108888	5.434651	0.021850
ELE	0.040411	0.217372	0.277491	0.040107	4.588264	0.016889
AIR	0.039706	0.215793	0.138531	0.033781	4.519573	0.018088
BBVA	0.039217	0.233907	0.061211	0.092320	4.372032	0.019580
MEL	0.036228	0.265997	0.525558	0.143137	4.435598	0.017299
FAE	0.035741	0.200817	0.041022	0.041907	3.468739	0.016020
APPS	0.035653	0.185156	0.041022	0.101356	3.299822	0.016467
ANA	0.032686	0.163216	0.080541	0.004117	3.083775	0.015148
REE	0.030542	0.184613	0.100086	0.013681	3.262307	0.013729
AMA	0.030153	0.183173	0.041022	0.010483	3.099572	0.015443
ENAG	0.030142	0.160157	0.041022	0.004259	3.029730	0.015348
ACX	0.029720	0.192559	0.041022	0.032952	3.043851	0.015351

Table 1: Centrality measures

1) **Degree centrality:** As we mentioned before, degree centrality indicates the connectivity of nodes. It give us the information on how many other nodes are connected with a particular node. Based on this measure, see Table 2, MAP (Mapfre) has the highest number of edges connected with other stocks, and in consequence its highest value 0.0949. After MAP, FER (Ferrovial) is the second stock with higher value with 0.091. The next five stocks with higher score are SAN (Banco Santander) 0.079, REP (Repsol) 0.0737, SCYR (Sacyr) 0.0561, ACS (Actividades de Construcciones y Servicios) 0.0534 and AENA 0.0527.

Stocks	*DC*
MAP	0.094983
FER	0.091593
SAN	0.079833
REP	0.073787
SCYR	0.056142
ACS	0.053417
AENA	0.052731
CIEA	0.044652
CABK	0.041228
ELE	0.040411

Table 2: Degree centrality *DC*

2) **Closeness centrality:** It measures how close a node is to all other nodes in terms of correlations. Closeness can also be regarded as a measure of how long the information can be spread from a given node to other reachable nodes. Based on this measure, see Table 3, MEL (Hoteles Meliá) has the highest scoring with 0.266. After MEL, ICAG (International Airlines Group) is the stock with the second highest score with 0.264. The next five stocks with higher closeness centrality score are SAN (Banco Santander) 0.263, AENA 0.262, FER (Ferrovial) 0.254, BBVA 0.233 and CABK (Caixabank) 0.228.

Stocks	*CC*
MEL	0.265997
ICAG	0.264793
SAN	0.263381
AENA	0.262261
FER	0.254899
BBVA	0.233907
CABK	0.228822
MAP	0.228604
REP	0.223742
MTS	0.221875

Table 3: Closeness centrality *CC*

3) **Betweenness centrality:** It indicates the extent to which a node lies in relative position with respect to the others This measure indicates the potentiality of node to influence the others. Based on this metric, see Table 4, SAN (Banco Santander) has the highest scoring with 0.689. After SAN, FER (Ferrovial) is the stock with the second highest score with 0.638. The next five stocks with higher betweenness centrality score are AENA 0.532, MEL (Hoteles Meliá) 0.525, ICAG (International Airlines Group) 0.504, ELE (Grupo Elektra) 0.277, MAP (Mapfre) 0.232.

Stocks	*BC*
SAN	0.689433
FER	0.638746
AENA	0.532216
MEL	0.525558
ICAG	0.504725
ELE	0.277491
MAP	0.232174
ACS	0.175473
REP	0.158290
AIR	0.138531

Table 4: Betweenness centrality *BC*

4) **Eigenvector centrality:** It indicates how an important node is connected to other important nodes (stocks). Hence, the stocks with higher scoring are also connected to other high-scoring stocks. Based on this metric, see Table 5, MAP (Mapfre) has the highest scoring with 0.527. Next five stocks with highest scores are SAN (Banco Santander) 0.424, REP (Repsol) 0.26, MVC (Metrovacesa) 0.202, FACE 0.189 and TUBA (Tubacel) 0.238.

Stocks	EC
MAP	0.527848
SAN	0.424127
REP	0.260941
MVC	0.202932
FACE	0.189226
TUBA	0.184399
NXTE	0.181685
FCC	0.180263
GSJ	0.164941
EZEN	0.160614

Table 5: Eigenvector centrality EC

5) **Subgraph centrality:** It indicates how an important node is connected to other important nodes (stocks). Hence, the stocks with higher scoring are also connected to other high-scoring stocks. Based on this metric, see Table 6, SAN (Banco Santander) has the highest scoring with 12.47. Next five stocks with highest scores are FER (Ferrovial) 11.46, MAP (Mapfre) 11.08, REP (Repsol) 7.93, ACS (Actividades de Construcciones y Servicios) 5.66 and CABK (Caixabank) 5.43.

Stocks	SC
SAN	12.479398
FER	11.468918
MAP	11.081115
REP	7.930440
ACS	5.667266
CABK	5.434651
AENA	5.411559
SCYR	5.010546
ELE	4.588264
AIR	4.519573

Table 6: Subgraph centrality SC

6) **Pagerank centrality:** It also indicates how an important node is connected to other important nodes. Based on this metric, see Table 7, SAN (Banco Santander) has the highest scoring with 12.47. Next five stocks with highest scores are FER (Ferrovial) 11.46, MAP (Mapfre) 11.08, REP (Repsol) 7.93, ACS (Actividades de Construcciones y Servicios) 5.66 and CABK (Caixabank) 5.43.

Stocks	PR
MAP	0.040754
FER	0.037861
SAN	0.033878
REP	0.032753
SCYR	0.025417
AENA	0.023394
ACS	0.022670
CABK	0.021850
CIEA	0.021221
BBVA	0.019580

Table 7: Pagerank centrality PR

In Fig. 2 centrality-centrality correlations in the weighted representation of the network of the Spanish stocks market. In particular, diagonal entries from the figure correspond to the distribution of the corresponding centrality measure.

4. Overall Centrality Measure: 3-Principal Centrality

Although in previous section we have introduced the general definition of the principal centrality measures, PC_r, $1 \leq r \leq 6$, in this study we have focused on PC_3. The main reason is a combination of simplicity (the less components we use, the simplest model) and an adequate level of explanation of the variance of the problem. Eigenvalues of the covariance matrix are

$$\{0.74569, 0.10628, 0.10293, 0.03868, 0.005716, 0.00068306\}.$$

Thus, with only the three first componenes we are able to explain the 95.4% of total variance, which seems to be perfectly adequate. There are three centrality measures that account in PC1: degree, subgraph and pagerank centralities. PC2 mainly accounts for the closeness centrality, and PC3 only accounts for the betweenness centrality.

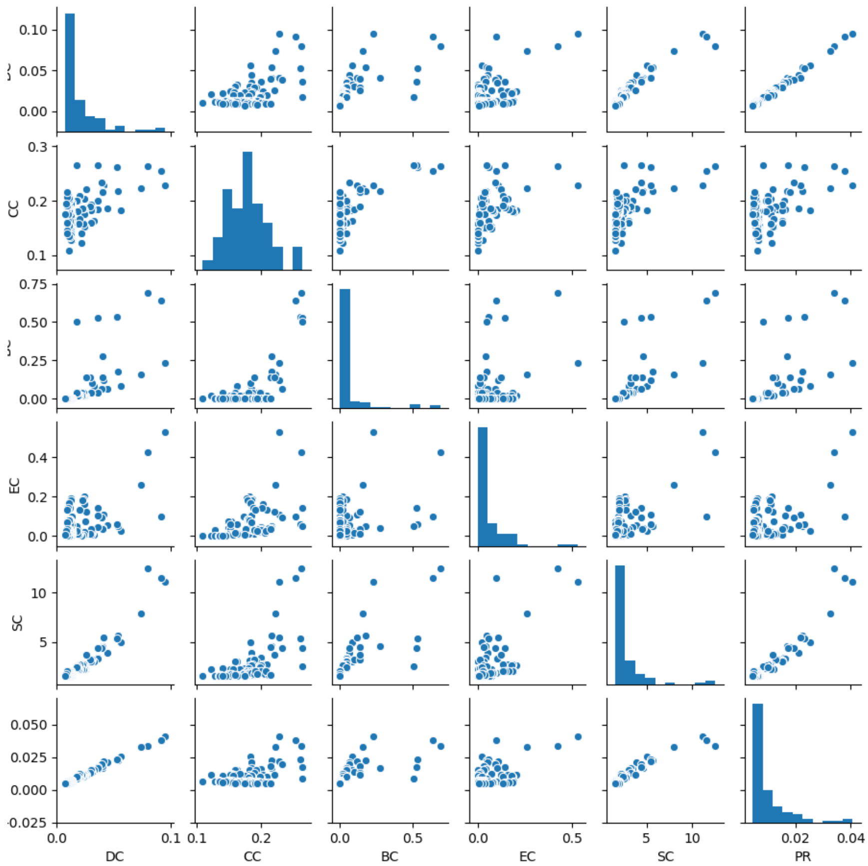

Figure 2: Scatterplots for the centrality-centrality correlations in the weighted representation of the network of the Spanish stocks market. The diagonal entries from the figure correspond to the distribution of the corresponding centrality measure. DC is the degree centralities; CC the closeness centralities; BC, the betweenness centrality; EC the eigenvector centralities; SC, the subgraph centrality; PR the page rank centralities.

For orthogonalizing the principal components and obtaining their structural interpretation, a Varimax rotation has been carried out. Fig 3 plots the first two components after Varimax rotation.

In Table 8 stocks with higher PC_3 scores are described. Based on the metric, SAN (Banco Santander) is the stock with the highest score with 9.63. The second highest score is MAP (Mapfre) 9.54. Next three stocks according the score are FER(FErrovial) with 8.73, REP (Repsol) 5.84 and AENA 5.25.

From that previous tables, we realize that the top higher-scoring stock for PC_3, SAN also appear as a higher-scoring stock in two out of six traditional measures. In concrete, is the leader in: betweenness and subgraph centrality. SAN also appears in the top 3 stocks for all centrality measures. Other stocks MAP, FER and REP also appeared as higherscoring stocks in all traditional centrality measures.

Stocks	PC_3
SAN	9.626606
MAP	9.546022
FER	8.735701
REP	5.841887
AENA	5.257887
MEL	4.670781
ICAG	4.103056
ACS	3.235957
SCYR	3.053861
CABK	2.830023

Table 8: Principal centrality PC_3

In Table 7, according to principal centrality PC_3 of stocks, most relevant sectors are shown. It can be verified that results are very well balanced, in the sense that top PC_3 higher-scoring stocks belong to different sectors.

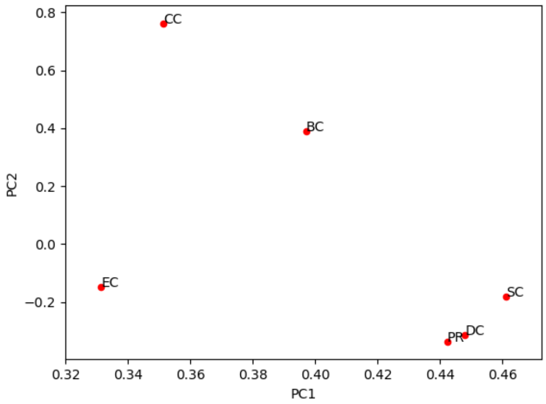

Figure 3: Plot of the different centrality measures studied in the space of the two principal components $PC1$ ans $PC2$ found by using the factor analysis. Observe the clustering of the degree, subgraph and page rank centralities in one cluster as well as that of closeness, betweenness and eigenvector centralities very far from each other, so in different clusters

Sector	Stock	Company name
Finance	SAN	Banco Santander
Insurance	MAP	Mapfre
Infraestructures	FER	Ferrovial
Energy	REP	Repsol
Transport	AENA	AENA

Table 9: Principal centrality PC_3: most relevant sectors

Other important aspect that can be verified is that stock SAN is also the most correlated stock with the IBEX index, which is the main reference stock market index of the Spanish stock market. Concretely correlation between SAN and IBEX is $0,84$ and their historic behaviour can be seen in Fig 4.

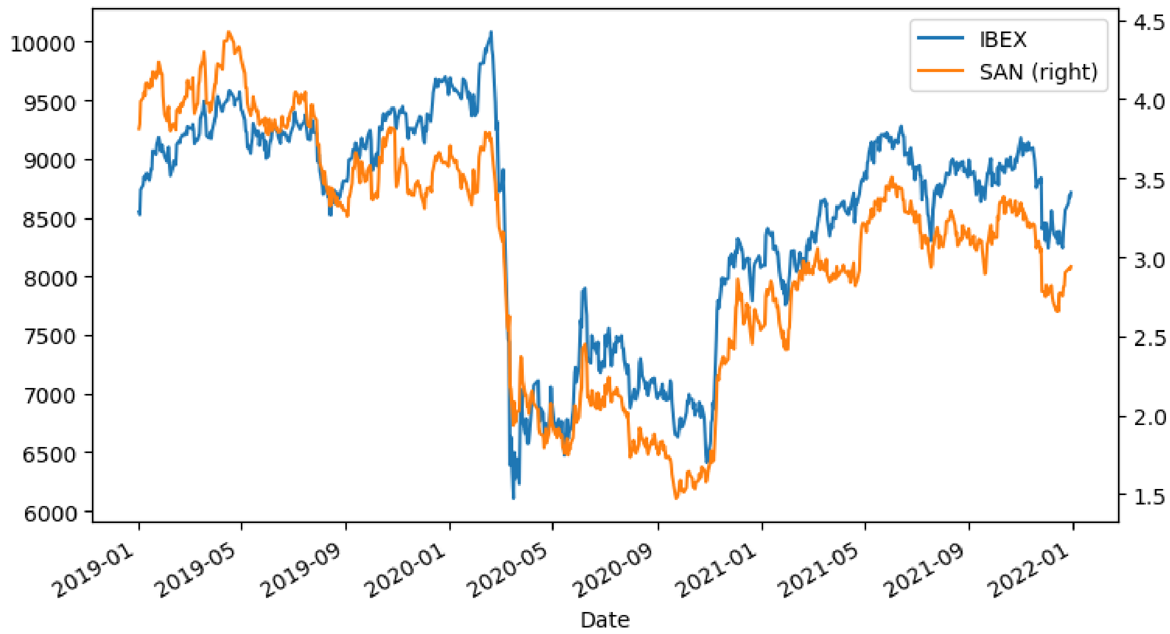

Figure 4: Historic evolution of SAN and IBEX

5. Conclusion

In this paper, the behaviour of the Spanish stocks market is analyzed by using network analysis approach. Since traditional centrality metrics have different interpretations and give us partial understanding of the nodes (stocks) relevance or significance in the network, we have applied principal components analysis for properly combining traditional centrality measures and obtaining an overall single network measure that permits us to have a consolidated vision of the role or influence of the stocks in the Spanish stocks network.

SAN, MAP, FER, REP, AENA are the five PC_3-highest-scoring stocks in the Spanish market. They also appear as highest-scoring stocks in all traditional centrality measures. Topologically, these stocks are connected to each other through a path that is centrally located in the MST. They also have a relevant number of connections with other stocks and are able to control the distribution of information and to spread it rapidly to other stocks in the stocks network. In general terms, they can be considered as the more relevant stocks in the Spanish stocks market, in terms of influence, information control and spread on it.

Acknowledgment

The author would like to thank to Prof. Dr. Ernesto Estrada for his helpful comments during the preparation of this article and the anonymous referee for his/her constructive comments that helped improve the quality of the paper.

References

[1] G. Bonanno, G. Caldarelli, F. Lillo and R. N. Mantegna, Topology of correlation-based minimal spanning trees in real and model market, Physical Review E, vol. 68, pp. 046130-046133, 2003.
[2] E. Estrada,, and J. A. Rodriguez Velazquez, Subgraph centrality in complex networks, Physical Review E 71 (2005) 056103.
[3] Estrada, E., and Örjan Bodin, Using network centrality measures to manage landscape connectivity, Ecological Applications, 18(7), (2008) pp. 1810–1825.
[4] L. C. Freeman, Centrality in networks: I. Conceptual clarification, Social Networks, vol. 1,(1979) pp. 251-239.
[5] L. C. Freeman, A set of measures of centrality based on betweenness. Sociometry, 40(1), (1977) 35-41.
[6] I. T. Jolliffe, Principal Component Analysis, 2nd ed. Springer, New York, 2002.
[7] J.B. Kruskal, On the shortest spanning subtree of a graph and the travelling salesman problem. American Mathematical Society, vol. 7, no. 1, (1956) pp. 48-50.
[8] G.S. Lee, Maman A. Djauhari, An Overall Centrality Measure: The Case of U.S Stock Market. International Journal of Basic and Applied Science, Vol 12, (2012).
[9] R.N. Mantegna, Hierarchical Structure in Financial Markets, European Physical Journal B, 11, (1999) 193-197.
[10] J. P. Onnela, A. Chakraborti, K. Kaski, J. Kert'esz and A. Kanto, Dynamics of market correlations: Taxonomy and portfolio analysis, Physical Review E, vol. 68, (2003) pp. 056110(1)-056110(12).
[11] S. Sharif, M. A. Djauhari, A Proposed Centrality Measure: The Case of Stocks Traded at Bursa Malaysia. Modern Applied Science; Vol. 6, No. 10 (2012).
[12] V. Tola, F. Lillo, M. Gallegati and R. N. Mantegna, Cluster analysis for portfolio optimization, Journal of Economic Dynamic and Control, vol. 32, (2008) pp. 235-258.
[13] M. Tumminello, T. Aste, T. D. Matteo and R. N. Mantegna, A tool for filtering information in complex systems," Proceedings of the National Academy of Sciences, vol. 102, (2005) pp. 10421-10426.
[14] D.B. West, Introduction to Graph Theory (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall. 2000.