Network Indicators: a new generation of measures? Exploratory review and illustration based on ESS data*

Elsa Fontainha[1], Edviges Coelho[2]
[1]ISEG Technical University of Lisbon, e-mail: elmano@iseg.utl.pt
[2]Statistics Portugal (INE), e-mail: edviges.coelho@ine.pt

Abstract

The main goal of this paper is to illustrate how to construct innovative indicators which contribute to a better knowledge of cohesion and convergence across Europe and the identification of the roles (e.g. centrality, reciprocity) performed by the several entities (countries, regions, institutions, individuals, enterprises, etc.) in European networks mapped by demographic, economic, financial and communication flows or links.
Adopting network analysis, derived from graph theory, and exploring data from the European Statistical System (ESS) related with flows of people, capital, goods or information, network indexes are computed (e.g. density, centrality). The empirical illustrations include networks which differ in size, scope, aggregation level, and nature (population migration, student and teacher temporary mobility, direct foreign investment).
The new indicators proposed here can be produced on a regular basis and are summary measures which will complement and improve the current official statistics.

Keywords: intra-EU flows, network analysis, economic indicators

1. Introduction

Every single minute there are millions and millions of flows crossing Europe. There are flows of capital, people, information, knowledge, goods, and services. Sets of these flows can be used to determine networks, and they can be represented through graphs or matrices and characterized by several network measures.
Some of these flows are registered by statistical offices on a regular basis. They can be interpreted as inter or intra links among diverse entities and aggregation levels (e.g. countries, individuals, regions, enterprises).
Regular statistical data and indicators are mainly focused on the attributes or characteristics of the entities, for example population mortality, regional income disparities, the ECB interest rate, or EU27 exports.
Attribute Data Analysis differs from Relational or Network Data Analysis because the latter focuses on the relation or flow among entities and computes measures (here designated as network indicators) which evaluate the density or cohesion of the

* All views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the Statistics Portugal (INE) or ISEG Technical University of Lisbon.

network, the reciprocity between the network entities or the centrality of the egos (nodes) inside the network. Both analyses - attribute and relational - are complementary and can be combined. Relational information can be represented by graphs as well as by the associated adjacency or association matrices.[1]

Socio and economic network analysis has already proved useful in diverse study domains: consensus and social influence, sociology of science, markets, diffusion and adoption of innovations, social support, and the world's political and economic system. A brief review of the framework of socioeconomic networks was recently presented by Vega-Redondo (2007) and the state of art of social and economic network analysis is summarized by Jackson (2008). The process of technological diffusion and the importance of informal social networks in labor markets are some aspects, studied by economists using network analysis.

The main methodological problem tackled in this paper is to find a way of improving the measurement of cohesion in the European Union. In order to address that problem, innovative measures are computed using network analysis (graph theory) based on data from the European Statistical System (ESS). Diverse empirical illustrations are provided. This research, whose results are partially presented here, constructs new indicators based on EU countries network characterization, creation and change. Those indicators can be produced on a regular basis and complement the current official statistics.

The paper is organized as follows: first, main network concepts are presented; then network indicators for European flows among countries and for different years are computed and discussed; and finally, conclusions are presented.

2. Concepts, Network Measures

2.1. Networks, graphs and related concepts

Before computing the network measures and indexes, some fundamental concepts must be introduced.

A directed graph G is a pair (V,E), where V is a finite set and E is a binary relation on V. The set V is called the vertex set of G, and its elements are called vertices also named nodes, agents or egos. The latter two names are more frequent in Social Sciences and the former in Mathematics, Computational Science and Artificial Intelligence. The set E is called the edge set of G, and its elements are called edges (Cormen *et al.* 2005:1080).

Figures 1a and *1b* represent two networks. Each vertex represents one country (circles), and edges are represented by arrows. *Figure 1a* is a hypothetical representation of EU population migrations and is a pictorial representation of a directed graph on the vertex set {1,2,3,4,5,6}. Self-loop edges from a vertex to itself, like vertex 2, are inexistent in this case by definition. Figure 1.b represents an undirected graph based on countries which are neighbours.

A directed graph is strongly connected if every two vertices are reachable from each other. The strongly connected components of a directed graph are the equivalence classes of vertices under the mutually reachable relation. A directed graph is strongly

[1] Input-output (I-O) analysis, which computes the I-O 'technical coefficients', can also be considered a relational analysis. They allow inter-industry analysis. Developed over several decades and generally associated with relations among sectors in National Accounts or regions. At the present input-output systems are also used to evaluate environmental issues (Ward, 2004). I-OA shares with the network analysis, among other aspects, the original data organization as a adjacency matrix, also called the association or network matrix.

connected if it has only strongly connected components. The graph in the Figure 1a has three strongly connected components: {1, 2, 4, 5}, {3}, and {6}. All pairs of vertices in {1, 2, 4, 5} are mutually reachable.

Figure 1a
Directed Graph – Group of six countries
(example: population immigrations)

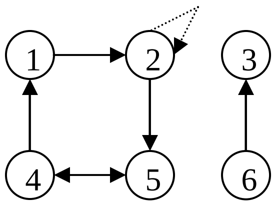
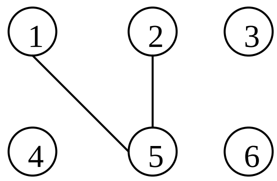


Figure 1b
Undirected graph- – Group of six countries
(example: countries which are neighbours



2.2. Networks measures and indicators

Multiple relational or network measures or indicators can be computed using relational data: measures which describe the network structure (e.g. size, density); measures related with one node (ego-measures), usually associated with the role it performs in the network (e.g. in-degree, out-degree, power, isolated).[2]

Density -The density of a graph (d) is computed as:

$$d = \frac{l}{n(n-1)/2}$$

where *l* is the number of lines and *n* the number of vertices.

Cohesiveness- Cohesiveness is a network structure characteristic that is usually applied to a subset of nodes. The cohesiveness of a network is important because the more links each node (e.g. country) has with the others, the more it is influenced by group performance (e.g. financial system behaviour). Mutuality and Frequency of ties among network members also affect cohesion positively.

Centrality- There are different measures of centrality. They may be associated with the whole graph or with the node position in the network that influences the role and the performance of all the net.

Reciprocity- Reciprocity, one of the indices of mutuality, is only defined for dyads. A dyad is a pair of nodes and all ties between them. A formal definition of dyad is an unordered pair of nodes (labeled *i* and *j*) and the arcs that exist between the two nodes in the pair, represented by *Dij=(Xij, Xji), i* and *j* being different.
The measures previously presented and others associated will be applied and discussed in next point.

[2] In network analysis, some characteristics are only computed for nodes, networks or a subset of nodes. Other characteristics, like degree, can be computed both for nodes and networks. For a brief presentation see Monge and Contractor (2003). For reasons of space limitations only a few results and associated indicators will be presented.

3. Data and Methodology

3.1. Data

From the European Statistical System (ESS) several Eurostat indicators were selected: immigration by country of previous residence (*migr_immiprv*), immigration by citizenship (*migr_immictz*), Erasmus students and teacher's mobility, foreign direct investment net flows by partner country (*bop_fdi_flows*).[3] The selection aims to include several domains for EU: mobility of people, capital and knowledge. The availability of data by countries of origin and destination of the flows (this information is scarcer than similar information by country groups) restricted the selection.

During the process of construction of the adjacency matrices for computing the network indicators, several problems arose. One main problem was the absence of information for some years and/or countries. Because the adjacency or association matrix must be square ($n \times n$), some information became lost during the process.[4] Another methodological problem arose when several years are compared. Because comparison among networks is only possible for networks of the same size, some information for one year must be ignored if similar information does not exist for the other year.

This paper is merely illustrative of the possibilities for constructing network indicators. Nodes and flows can be analysed at diverse levels of aggregation, points in time and kinds of links.[5]

3.2. Methodology

The main steps for computing the network indicators were:

(i) Selection of the relational information from the ESS databases;

(ii) Study of the comparability among years and countries and selection of subgroups of countries or years. Imputation of values to the missing data by country based on the values for $t+1$ or $t-1$;

(iii) Construction of the association or adjacency matrix, adopting for the column the representation of destination countries and for the line the origin countries. When the original data have a different format, the matrix is changed (e.g. transposed);

(iv) Weighting the original values according to two weights based on origin stock or based on destination stock with which the flows are related. For example, the immigration flows are weighted by the population of the origin of the flows (weight 01) and also by population of destination (weight 02);

(v) Construction of the adjacency matrices with the weighted values;

(vi) From the adjacency matrices, the associated graphs are represented. For each network several graphs were designed, considering all the directed connections or only the strongest connections. The importance of the connection was evaluated according to different criteria and levels. The different simulations for graphs not

---

[3] The authors thank Sonia Santos helpful information about foreign direct investment variables.All errors or omissions are our own.

[4] For example, because there are no data for France about immigration by country, all information about the French population in all other European countries must be ignored in the analysis, even if data for this are available.

[5] Examples of other flows studied by the authors but not presented here are: trade among countries and regions, flows of knowledge-based services, flows of tourists' movements (trips and nights). Network indicators can be extended and used for computing interactions between individuals (inside firms, households, cities). Fontainha (2008), matching the surveys of individuals of the same household (the household as a network) applied them to study family members economic role.

shown in this paper are available from authors upon request. Reciprocity, isolated and other node characteristics are flagged in the graphs;

(vii) Based on the adjacency matrices, several indicators are computed for network and nodes[6]. Some measures consider the dichotomized matrices, ignoring the intensity of the links, and only considering their existence or absence.

(viii) Finally the results are compared and discussed for each area in a dynamic perspective (e.g. did the density increase in this network between 1998 and 2006?) and in a cross- domain perspective (e.g. are the indicators similar between students and teachers participating in the Erasmus mobility programme?)

3. Results

3.1. An example: intra-EU immigrant flows

Network and graphs

Figures 2a to *Figure 2d* map, total and partially, for 2002 and 2006, the migrations among several European countries, all of them members of EU at the present. The Eurostat data used was the indicator *migr_immiprv*. For both years the association matrices are 17x17 and they are obviously asymmetric. Each node represents a country. The graph representations was carried out considering two kinds of links: all the links independent of the intensity of the migrations (*Figures 2a* and *Figure 2c*); only the links which represent flows equal to or greater than 30 migrants per 100000 habitants of the country of origin.[7] The red arrows represent mutuality between two countries (dyad), the blue inexistence of mutuality. From 2002 to 2006 the reciprocity increases as shown by the decrease of dark lines from *Figure 2a* to *Figure 2c*.

One country, Germany, retains an important position in the group in both years. Spain, gains importance in 2006. Because *Figure 2a and b, Figure 2c and 2d* represents the flows respectively before and after EU enlargement from EU15 to EU25, not surprisingly the flows from the new EU25 countries increase both to neighbour countries and to others.

The graphic representation of the network is visually attractive, but woefully inadequate at measuring the network and node characteristics.[8]

Network and network measures

Several measures are computed from this network. *Table 1* shows some network indicators[9]

The network density increases between 1998 to 2006 from 8.4 to 13.2, considering the flows weighted by the citizenship. Using an alternative weight (population of previous residence of the immigrant) and for different period (2002-2006) the density of the network also increased. In general, the computations adopting the citizenship or the last residence are similar in results (Table 1, columns 1 and 2).

---

[6] Many software packages are available for doing network analysis as UNICET and Pajek (Borgatti et al 2002, Nooy et al 2005).

[7] This means that if the ratio (migrants/population of origin country) is less than 30/100000, the link is ignored in the computation and graph. Because the represented links in each year are only the links with values equal to or greater than 30/100000, the two graphs represent the migrations partially.

[8] Note that in the graph the representation is random and the size of the lines has no significance.

[9] For the discussion of the meaning and computation of network measures see, for example, Jackson (2008) or Monge and Contractor (2003).

Figure 2a- Immigration flows
(origin: country of previous residence)
2002
All network (251 ties, 17 nodes)

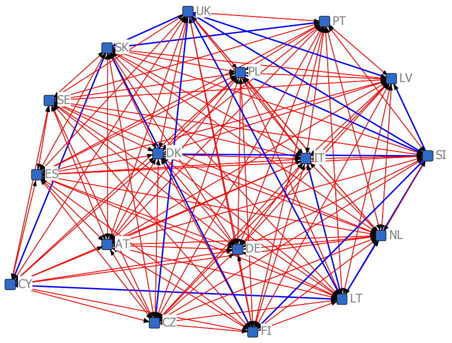


Figure 2b- Immigration flows
(origin: country of previous residence)
2002
Partial network (27 ties, 17 nodes)

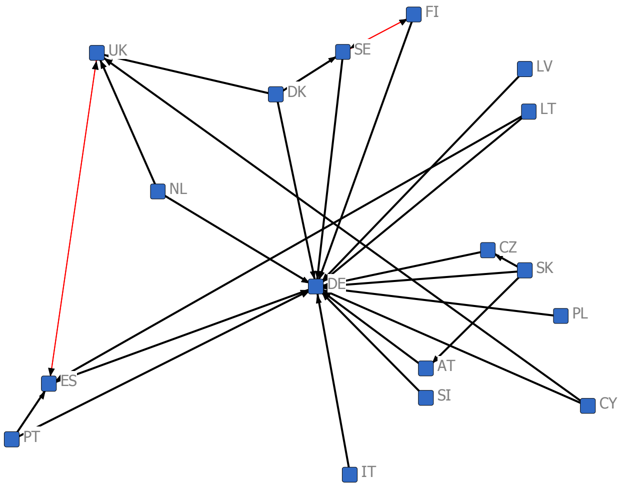


Figure 2c- Immigration flows
(origin: country of previous residence)
2006
All network (267 ties, 17 nodes)

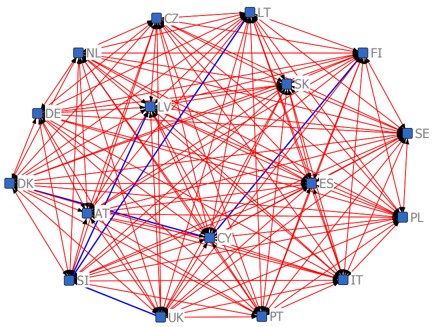
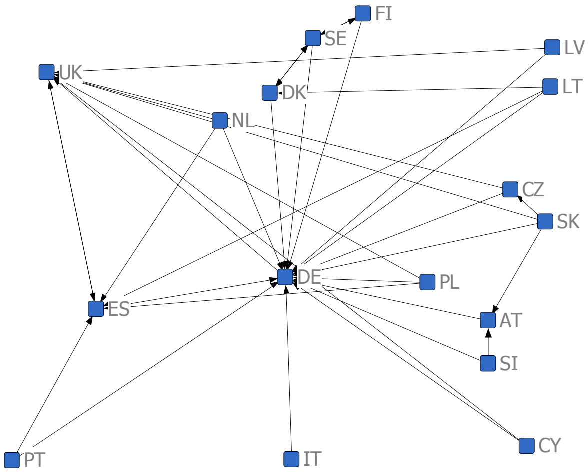


Figure 2d- Immigration flows
(origin: country of previous residence**)**
2006
Partial network (36 ties, 17 nodes)



3.2. EU network indicators: dynamic and cross domain comparison

The domains analysed, present last years, an increase of the network density Immigration between 1998/2002 and 2006, foreign investment between 2002 and 2006 and students and teacher between 2004/2005 and 2006/2007 (*Table 1*)[10]. This means that the number of actual flows (links) among countries increased compared with the total possible links. Each group of EU countries analysed increase the cohesion.

The density of the students' network compared with the teachers' network counterpart is higher. The lowest density is shown in the foreign investment which can be explained by the volatility of this economic variable (*Table 1*).

[10] The original values were weighted by differently. Immigration was weighted by the population of residence or citizenship of the immigrant. The students and teachers by the total population of origin country, and the IDE by the gross capital formation in origin country. For IDE only positive net flows were considered (means disinvestment was ignored).

Networks centralization, in general decreases with two exceptions: the foreign investment (in and out-degree)[11] and student mobility (in-degree). This centralization is computed by Freeman method. The increase in the centrality means that the central node-country used as reference for computing the centrality increase the central role. That central node used as reference is not the same in the two years as shown in *Table 2*. For foreign investment is United Kingdom (max out-degree 2002) and Filand (max in-degree in 2002) and France (max out-degree in 2006) and Belgium (max in-degree in 2006).

Table 1 [12] – Network Measures – Network (EU as a set of networks)

	Immigration(res)		Immigration(ctz)		*FD Investment*		*Students*		*Teachers*	
	[1]		[2]		[3]		[4]		[5]	
	t+n	*t*	*t+n*	*t*	*t+n*	*t*	*t+n*	*t*	*t+n*	*t*
	2006	2002	2006	1998	2006	2002	2006/7	2004/5	2006/7	2004/5
N countries	17	17	17	17	15	15	27	27	27	27
Network size (N of nodes)	17	17	17	17	15	15	27	27	27	27
N of ties	267	251	267	242	80	135	606	573	604	570
Network Centralization (Outdegree) %	6.76	9.14	5.04	5.64	10.27	6.73	11.12	11.34	10.05	13.61
Network Centralization (Indegree) %	20.34	34.34	22.66	40.746	14.45	7.30	13.93	13.77	10.47	13.60
Outdegree Mean (StdDev)	264.73 (173.51)	187.64 (108.58)	211.39 (131.57)	134.31 (65.03)	18.47 (14.75)	15.24 (13.34)	5728.76 (6961.91)	5340.26 (6585.65)	885.77 (724.44)	754.26 (640.40)
Indegree Mean (StdDev)	264.73 (421.70)	187.64 (355.61)	211.39 (348.57)	134.31 (292.92	18.47 (15.85)	15.24 (15.18)	5728.76 (6993.17)	5340.26 (6841.47)	885.77 (729.52)	754.27 (665.03)
Network Density	16.54 (45.64)	11.77 (32.59)	13.21 (35.07)	8.40 (23.15)	1.32 (3.18)	1.23 (3.56)	229.15 (619.03)	213.61 (596.34)	35.4308 (61.00)	30.17 (53.02)

Notes: Data sources are for [1] to [3] Eurostat database and for [4] and [5] EU Erasmus information on line.
Group of countries in each network [1] : AT CY CZ DE DK ES FI IT LT LV NL PL PT SE SI SK UK [2]:AT BE DE DK ES FI HU IT LI LU LV NL PT SE SI SK UK [3]: AT BE CZ DE DK EE ES FI FR HU IT NL PT SE UK [4]: EU27 [5]: EU27.

3.3. EU network nodes: country comparisons

The relevance of nodes (countries in this paper) inside a network is evaluated by different forms. Table 2 illustrates some of those ego-indicators, showing the country top values by domain. The centrality of the countries evaluated by in-degree reveals that for the 3 presented areas (immigration by residence, foreign direct investment and Erasmus student mobility) a small group of countries have a central role in the network (Germany, Spain, France and United Kingdom) for population and students and Netherlands for investment. The same group of countries also present a high out-degree for students and investment (in this case with Sweden). A high value of a country for the in-degree means that that country is a central attractors considering the other countries in the network.

Two countries are present the higher out-degrees for immigration in both years: Poland and Slovakia. This higher out-degree of a node can be considered as the most influential importance of that country as origin of immigrants (thought it might matter to which node it is migrating).

The results for node size, as expected, converge with the other results: an increase in the number of ties among EU countries.

[11] The in-degree is related with the oriented links from the node considered and the in-degree is related with the oriented links received by the node under consideration. The size of a network is evaluated by the number of ties (N-1).

[12] Due to space limitations only partial results can be given in Table 1 and Table 2.

Table 2– Network Measures - Nodes (Country as node)

	Immigration(res)		Foreign Investment		Students Mobility	
	2006	2002	2006	2002	2006/7	2004/5
[13]Size (in)	16 (except CY, SI)	16 (except CY, LT, LV, PL, PT, SI, SK)	NL(10) PT (7) SE (7)	BE (11) SE (11)	26 BE ES FR IT DE EE	26 BE DE ES FR IT PT FI SE UK LV LT AT PL FI
Size (out)	16 (except FI, LT, LV, SI)	16 AT, DE, ES, NL, PL, SE, UK	FR (14) NL (9)	AT (13) FR (13) IT (13) NL (13) UK (13)	25 BE CZ DE EE ES FR IT LV LT AT PL FI	25 BE DE ES FR IT NL AT PT FI UK
In-degree	DE (1578) UK(1082) ES(577)	DE (1153) UK (298) ES (280)	BE (60) NL (47)	FI (49) NL (46)	ES (26992) FR (20140) DE (16683)	ES (25217) FR (20200) DE (16688)
Out-degree	PL (701) SK(613)	SK (551) PL (310)	FR (48) ES (42)	UK (47) SE (37)	DE (22717) FR (22421) ES (21918)	DE (2715) FR (21213) ES (20568)

3. Conclusions

The mains results of this research are:

- First, it is possible to compute network indicators from the ESS data. Here some illustrations are presented. These indicators are useful for the understanding of the relations among countries and the network diffusion mechanisms (innovation, crisis, information, etc.)
- Second, there is, in general, an increase in the density of the network across time regardless of the phenomena under analysis (migrations, capital flows, etc.). Cohesion is increasing. EU enlargement is reflected in that increase.
- Third, the role of each country (evaluated by ego or node network indicators) remains stable in some cases but changes in others. For example, with regard to immigration, Spain increased its position as a destination country since 1998.
- Fourth, and as topic for further research, together with economic and cultural factors, the results suggest that geography and language still matter for several EU networks of people, goods and capital and knowledge.

References

Borgatti, S.P., Everett, M.G. and Freeman, L.C. (2002) *Unicet for Windows. Software for social network analysis*, Harvard, MA; Analytic Technologies. [A unicet tutorial by Bob Hanneman is available at http//faculty.ecr.edu/~hanneman.]

Cormen, T.H., Leiserson, C.E., Rivest, R. and Stein, C. (2001) *Introduction to Algorithms*, (2nd edition, 6th reprint 2005). MIT Press.

Fontainha, E. (2008) “With whom are you doing? Applying Network Analysis to Measure Intra-household Resources Allocation, Bargaining and Interactions”, *30th Conference of International Association for Research in Income and Wealth (IARIW)*, Portoroz, Slovenia, August.

Monge, P.R. and Contractor, N. S.(2003) *Theories of Communication Networks*, Oxford University Press.

Nooy, W., Mrvar, A. and Batagelj, V. (2005) *Exploratory Social Network Analysis with Pajek*, Cambridge University Press.

[13] Note that the size is based on dichotomic information. So, for example for investment it does not take the amount of investment but only the existence or absence of it, the existence or absence of links with other countries.

Carrington, P. J., Scott, J. and Wasserman, S., eds. (2005*) Models and Methods in Social Network Analysis*. Cambridge University Press.
Jackson, Matthew O. (2008) *Social and Economic Networks*, Princeton University Press.
Vega-Redondo, F. (2007) *Complex Social Networks*, Cambridge University Press.
Ward, M. (2004) *Quantifying the Word- UN Ideas and Statistics,* Indiana University Press.