Available online at www.sciencedirect.com
SCIENCE DIRECT®
Social Networks 28 (2006) 442–448
SOCIAL NETWORKS
www.elsevier.com/locate/socnet

Core/periphery structure models: An alternative methodological proposal

Ana Salomé García Muñiz *, Carmen Ramos Carvajal [1]

University of Oviedo, Faculty of Economics, Department of Applied Economics, Avd. Cristo s/n, 33071 Oviedo, Asturias, Spain

Abstract

The idea of a structure made up of a core and periphery is a common, classic paradigm in many fields of science. Following this line, in 1999, researchers Stephen Borgatti and Martin Everett developed a model of structural analysis based on the delimitation of a core formed by a group of densely connected actors, in contrast to a class of actors, more loosely connected and forming the periphery of the system.

The original approach of these authors is modified, employing measures that, in our opinion, show a larger degree of coherence and accuracy in the proposed objectives.

Keywords: Periphery structure; Coherence; Cohesive group

1. Introduction

Trying to form the concept referent to a cluster of actors that form a core around which observed activity revolves is challenging. The idea of a structure made up of a core and periphery is a classic paradigm occurring in many fields of science.

In 1999, the researchers Stephen Borgatti and Martin Everett developed a model of structural analysis based on the delimitation of a core formed by a group of densely connected actors. This cohesive group was characterized by a high density of interrelations in contrast to a more loosely connected class of actors forming the periphery of the system.

We modify the original approach of these authors employing measures that, in our opinion, show a larger degree of coherence and accuracy in the proposed objectives. In this work, we

* Corresponding author. Tel.: +34 985105058; fax: +34 985105050.
E-mail addresses: asgarcia@uniovi.es (A.S. García Muñiz), cramos@uniovi.es (C. Ramos Carvajal).
[1] Tel.: +34 985105054; fax: +34 985105050.

0378-8733/$ – see front matter
doi:10.1016/j.socnet.2005.11.001

concentrate on estimation and evaluation of core/periphery models based on valued graphs without loss of generality.[2] A synthesis of obtained results concludes this paper.

2. Core periphery model: general approach

Following the line of development started by Borgatti and Everett (1999), it is possible to build a model to determine the core of activity as opposed to the nodes of the periphery. The delimitation of the interested zones is based on the comparison with an ideal structure from available network data. It is assumed that the actors with intense relations will be the ones to form the main core.

The measure of proximity between real and theoretical structure proposed by Borgatti and Everett is:

$$\rho = \sum_{i=1}^{n}\sum_{i=1}^{n} x_{ij}\delta_{ij}, \quad \delta_{ij} = c_i c_j \tag{1}$$

where x_{ij} shows the observed relations between nodes i and j, δ_{ij} the ties in the ideal image and c_i represents the degree of centrality in such a way that $c_i \geq 0$.

The theoretical structure will have relatively large values for pairs of nodes that are both high in coreness, while it will be formed by intermediate values in the cases where only one node is high. Finally, the periphery actors will entail small values.

The estimation of core values c_i is made through a maximization procedure of the correlation between the observed or real structure and the theoretical one. However, in general, such a method cannot always be correct. The existence of high correlation between structures does not imply necessarily that both will be identical, only that their behaviour will be similar. The correlation coefficient measures the strength and direction of relationship between two variables, but not the concordance[3] of the observation value. In fact, a change in the units of one magnitude does not affect correlation, but it does affect concordance.

Furthermore, neither is the first method employed in the resolution of maximization suitable. The iterative applied algorithm – proposed by Fletcher and Powell – introduces variability in the results according to the considered point of starting. It argues the uniqueness and representativeness of the solution due to the possibility of offering a local maximum, though not a global one, as a solution of the function (Everitt, 1987).[4]

The use of a measure that is not suitable for all situations and the limitations of an iterative process show the necessity for a methodological improvement, both in estimating coefficients and analysing reliability. These questions are tackled in the next epigraphs.

3. Estimation process

We propose to employ information measures in order to estimate the theoretical structure presenting minimum divergence with real observed structure.

One of the pioneering theories of information measures is the so-called entropy of Shannon employed like other similar measures as an indicator of diversity being present in a population.[5]

[2] The techniques of reliability employed can be extended to dichotomy graphs.

[3] Consider concordance in the sense of equality of values between structures.

[4] Nowadays, the network programs like UCINET have improved the algorithm employing deterministic methods that overcome this limitation.

[5] See Rao (1982) and Good and Smith (1985).

The entropy of a distribution can be understood as there being disorder in the distribution, i.e. the uncertainty associated with a determined phenomenon. This uncertainty can be quantified considering a probability system. So if a phenomenon is associated with a uniform probability distribution, it will show a greater load of uncertainty than any other.

Starting with the entropy measure proposed by Shannon (1948),[6] Kullback and Leibler (1951) we introduce a measure of divergence between two distributions. Consider two discrete random variables X and Y whose probabilities are $\mathbf{p} = \{p_1 \ldots p_n\}$ and $\mathbf{q} = \{q_1 \ldots q_n\}$, respectively. The Kullback–Leibler distance is defined as:

$$D(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \tag{2}$$

Following the line developed by Golan et al. (1994), the next expression arises from the last divergence of Kullback and Leibler giving two joint probability distributions $\mathbf{P} = \{p_{ij}\}$ and $\mathbf{Q} = \{q_{ij}\} \forall i, j = 1, \ldots, n$.

$$\sum_{i=1}^{n} \sum_{j=1}^{n} p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{3}$$

This measure allows us to detect the divergence between two-dimensional discrete distributions. In addition, it offers an alternative to the estimation of theoretical matrix (δ_{ij}) that supposes a minimum divergence with real observed structure (x_{ij}).

The specification of a core–periphery model is inspired by the minimum divergence principle subject to a series of restrictions:

$$\begin{aligned}
&\mathrm{Min} \sum_{i=1}^{n} \sum_{j=1}^{n} p_{ij} \log \frac{p_{ij}}{q_{ij}} \\
&\text{s.a.} \sum_{i=1}^{n} \sum_{j=1}^{n} p_{ij} = 1 \\
&0 \leq p_{ij} \leq 1 \\
&p_{ij} = \frac{\delta_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{n} \delta_{ij}} \\
&\delta_{ij} = c_i c_j \\
&c_i \geq 0; \quad c_j \geq 0
\end{aligned} \tag{4}$$

where p_{ij} and q_{ij} indicate the proportions of theoretical (δ_{ij}) and observed (x_{ij}) ties in the structures collected in Table 1, where

$$p_{ij} = \frac{\delta_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{n} \delta_{ij}}, \quad q_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij}} \tag{5}$$

[6] Shannon gave his well-known measure of entropy for the probability distribution $\mathbf{p} = \{p_1 \ldots p_n\}$ as: $S(p) = -\sum_{i=1}^{n} p_i \ln p_i$.

Table 1
Structural organization chart

Theoretical network				Real network			
	1	...	n		1	...	n
1	δ_{11}	...	δ_{1n}	1	x_{11}	...	x_{1n}
...	...	...	...	...	...	...	...
n	δ_{n1}	...	δ_{nn}	n	x_{n1}	...	x_{nn}

The solution to minimize the process can be obtained by employing Lagrange[7] multipliers (Golan et al., 1994).

It is interesting to emphasize some properties of this process,[8] as a unique solution, that shows this method is more suitable potentially than the algorithm employed by Borgatti and Everett which does not allow us to guarantee the representativeness of the solution.[9]

The method also provides similar structures where as many zeros as the intensity of variables are maintained. The important or relatively high coefficients q_{ij}, represented by p_{ij} are relatively considerable too in the theoretical matrix and, the coefficients, null or positive in the observed matrix, will show the same pattern in the image structure.[10]

Moreover, the possibility of including non- linear equations together with inequalities supposes an operational advantage which other less fast and efficient methods lack. This feature allows the inclusion of additional information derived from other studies or sources, as it is possible to define from proper equations.

4. Analysis of reliability

Subsequent to estimating the model, an interesting point of analysis is evaluating the degree of adaptability of the theoretical matrix to the observed one.

Following the leading theme through information theory, one measure of the degree of accuracy between structures that has been applied thoroughly for its suitable properties is the Kullback and Leibler divergence, which has been employed in the estimation procedure.

Given two probability distributions $\mathbf{p} = \{p_1 \ldots p_n\}$ and $\mathbf{q} = \{q_1 \ldots q_n\}$, a divergent indicator between two probability distributions can be understood as a measure of an error committed to consider that distribution q is the correct one. Its expression is:

$$D(\mathbf{p}||\mathbf{q})\sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \tag{6}$$

It shows the inefficiency or the error committed to assume that correct distribution is q when in fact it is the unknown distribution p. In this sense, it is an indicator of the degree of adaptability of each node to the core periphery model.

[7] The problem has to be resolved numerically given its characteristics.
[8] See the properties in Blien and Tassinopoulos (2001).
[9] See Everitt (1987) and Cover and Thomas (1991) for a revision of both methods, respectively.
[10] See Blien and Graef (1997).

It is a measured limited by zero. It takes this value if and only if $\mathbf{p} = \mathbf{q}$. As the divergence between distributions becomes higher, such an indicator will increase.[11]

Now let us consider two probability distributions $\mathbf{P}_{xy} = \{p_{ij}\}$ and $\mathbf{Q}_{xy} = \{q_{ij}\}$ defined before and whose elements p_{ij} and q_{ij} correspond to proportions of theoretical (δ_{ij}) and observed (x_{ij}) ties, respectively.

The Kullback and Leibler (1951) distance associated with these two-dimensional variables offers a global concordance indicator between theoretical and real structure with analogue behaviour to the measure expressed for a one-dimensional variable:

$$D(\mathbf{P}_{xy}||\mathbf{Q}_{xy}) = \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij} \log \frac{p_{ij}}{q_{ij}} \tag{7}$$

So, the information measures not only offer a global measure of concordance, but allow a differentiation of this level according to observed actors and maintain the theoretical base of information theory employed in the estimation process.

5. A case study: European economy

The input–output analysis is an important tool in economic studies because it offers complete information as to relations between sectors and generated demand. Input–output models themselves are a general class of network flow models used widely in other scientific disciplines (Olsen, 1992).

In this work, we apply the core/periphery methodology to an input–output table. The concept is not new in economics, but the combination of this idea and the theory of networks is novel. Specifically, we take the European input–output table of 1995[12] (EUIO95) in order to analyze the global economy of this area.

The results of applying the core/periphery model are set out in Table 2. The degrees of centrality (c_i) estimated by proposed information methods are in the first column denoted by (G&R). We have used GAMS to programme it. On other hand, the results of the Borgatti and Everett model obtained by UCINET[13] are in column two under the symbols (B&E).

The ideal core–periphery structures have a general pattern in common with some differences with regard to the observed image: The information method (G&R) concedes a central position to metal products except machinery, paper and printing products, inland transport services. All of them are sectors with a more peripheral status in the Borgatti and Everett model (B&E). On other hand, agricultural and industrial machinery and credit and insurance services and institutions are central in the latter, but not in the information method.

The Kullback–Leibler divergence offers a valuation of these models that allows a choice between both techniques. In this case, the divergence for the proposed model is 0.364 and for the other 0.388. So, the estimated structure via the information method shows a major degree of adaptation to the observed network. This characteristic will be always true given the applied process of minimum divergence.

[11] It must be taken into consideration that D($\mathbf{p}$| |$\mathbf{q}$) ≠ D($\mathbf{q}$| |$\mathbf{p}$) which is a relevant feature in the decision of the matrix on which it will make the comparison.

[12] At this moment, this is the latest published European input–output table.

[13] The software UCINET VI (Borgatti et al., 2002) is in www.analytictech.com.

Table 2
Degree of centrality and reliability

Sectors	c_i (G&R)	c_i (B&E)	KL (G&R)	KL (B&E)	Gain (%)
Agriculture, forestry and fishery products	0.145	0.146	0.352	0.383	0.080
Fuel and power products	0.303	0.190	0.179	0.197	0.091
Ferrous and non-ferrous ores and metals	0.267	0.201	0.285	0.348	0.182
Non-metallic mineral products	0.180	0.154	0.189	0.239	0.211
Chemical products	0.238	0.205	0.182	0.213	0.148
Metal products except machinery	0.217	0.180	0.270	0.323	0.163
Agricultural and industrial machinery	0.196	0.183	0.248	0.279	0.108
Office and data processing machines	0.115	0.151	0.270	0.296	0.087
Electrical goods	0.199	0.177	0.209	0.235	0.112
Transport equipment	0.148	0.171	0.241	0.262	0.081
Food, beverages, tobacco	0.142	0.166	0.447	0.472	0.052
Textiles and clothing, leather, footwear	0.121	0.127	0.285	0.301	0.050
Paper and printing products	0.210	0.151	0.184	0.202	0.090
Rubber and plastic products	0.200	0.181	0.276	0.308	0.103
Other manufacturing products	0.134	0.110	0.143	0.181	0.210
Building and construction	0.202	0.168	0.274	0.317	0.134
Recovery, repair serv., wholesale, retail	0.371	0.269	0.240	0.241	0.002
Lodging and catering services	0.141	0.144	0.462	0.462	0.001
Inland transport services	0.247	0.155	0.308	0.338	0.090
Maritime and air transport services	0.089	0.132	0.483	0.464	−0.041
Auxiliary transport services	0.186	0.179	0.336	0.341	0.016
Communication services	0.135	0.092	0.289	0.294	0.014
Services of credit, insurance institutions	0.202	0.198	0.634	0.614	−0.033
Other market services	0.491	0.564	0.271	0.288	0.058
Non-market services	0.160	0.121	0.230	0.210	−0.095
Mean	0.202	0.181	0.292	0.312	0.077

At the same time, if we want a detailed sector analysis of reliability we can apply the Kullback–Leibler divergence by columns instead of by the whole table. The results for both models are in columns 3 and 4 of Table 2. In general, the adaptation is better in the proposed model. Only in three cases, do we have a loss of explicative capacity but even so it is very small.

It can be concluded that the information method is an alternative that allows a better explanation of the network. Besides, it is more in keeping with the objective of the model. Use of the Borgatti and Everett model could give confused results.

6. Conclusion

The core–periphery models developed by Borgatti and Everett (1999) provide a useful analytical tool that represents the classic idea of a core formed by a group of densely connected actors, in contrast to a more loosely connected class of actors making up the periphery of the system.

The original approach of the above-mentioned authors is modified employing measures that we think exhibit a greater degree of coherence and accuracy in the proposed objectives.

References

Blien, U., Graef, F., 1997. Entropy optimising methods for the estimation of tables. In: Proceedings of the 21st Annual Conference of the Gesellschaft für Klassifikation e.V, University of Potsdam, March 12–14.

Blien, U., Tassinopoulos, A., 2001. Forecasting regional employment with the ENTROP method. Regional Studies 5 (2), 113–124.

Borgatti, S.P., Everett, M.G., 1999. Models of core/periphery structures. Social Networks 21, 375–395.

Borgatti, S.P., Everett, M.G., Freeman, L.C., 2002. Ucinet for Windows: software for social network analysis. Analytic Technologies, Harvard, MA.

Cover, T., Thomas, J., 1991. Elements of Information Theory. Wiley Series in Telecommunications, United States of America.

Everitt, B., 1987. Introduction to Optimisation Methods and Their Application in Statistics. Chapman and Hall Ltd., New York.

Golan, A., Judge, G., Robinson, S., 1994. Recovering information from incomplete or partial multisectoral economic data. The Review of Economics and Statistics 76, 541–549.

Good, I.J., Smith, E.P., 1985. The variance and covariance of a generalized index of similarity especially for a generalization of an index of Hellinger and Bhattacharyya. Communication in Statistics – Theory and Methods 14, 3053–3061.

Kullback, S., Leibler, A., 1951. On information and sufficiency. Annals of Mathematical Statistics 22, 79–86.

Olsen, J.A., 1992. Input–output models, directed graphs and flows in networks. Economic Modelling 9 (4), 365–384.

Rao, C.R., 1982. Diversity and dissimilarity coefficients: a unified approach. Theoretical Population Biology 25, 24–43.

Shannon, C.E., 1948. A mathematical theory of communication. Bell Systems Technical Journal 27, 379–423, 623–659.