

Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Physics Letters A

www.elsevier.com/locate/pla



# Measuring the flow of information among cities using the diffusion power

B.A. Mello<sup>a,\*</sup>, L.H. Batistuta<sup>b,c</sup>, R. Boueri<sup>b,c</sup>, D.O. Cajueiro<sup>d,e</sup>

<sup>a</sup> Institute of Physics, University of Brasilia, DF, 70910-900, Brazil

<sup>b</sup> Institute of Applied Economic Research (IPEA), Brasilia, DF, Brazil

<sup>c</sup> Department of Economics, Catholic University of Brasilia, DF, Brazil

<sup>d</sup> Department of Economics, University of Brasilia, DF, 70910-900, Brazil

<sup>e</sup> National Institute of Science and Technology for Complex Systems, Brazil

## ARTICLE INFO

### Article history:

Received 17 July 2009

Received in revised form 22 October 2009

Accepted 23 October 2009

Available online 28 October 2009

Communicated by C.R. Doering

### PACS:

89.75.-k

89.65.-s

89.70.Hj

### Keywords:

Centrality measures

Complex networks

Diffusion power

Domination power

Page rank

## ABSTRACT

In this Letter, we define here the so-called *diffusion power* – an extension of the dominance power, which considers the interaction between neighbors of higher orders. Using this measure, we analyze the centrality of cities in two networks of the flow of information among these cities, namely a network of calls among the cities and a network of radio stations. Finally, we explain the centralities of the cities evaluated using the diffusion power in terms of the specific characteristics of the cities that belong to the network.

© 2009 Elsevier B.V. All rights reserved.

## 1. Introduction

Main stream research in the statistical physics field has been to characterize the dynamics of systems such as airports [1], financial institutions [2–4], web pages [5], flow of information networks [6,7], information city networks [8–11], social networks [12–14] and chaotic interacting functions [15] that may be described by complex weblike structures. Several comprehensive reviews on this subject are now available [16–18].

In the literature of complex networks, one of the most important concepts called *centrality* is used to define the relative importance of a node in a network. This concept can be defined in several different ways. For a typical undirected unweighed network, the simplest definition is the so-called degree of a node.<sup>1</sup> Other popular definitions such as *closeness centrality* [19], which is the inverse of the average distance from one node to all other nodes, *graph centrality* [20], which is the inverse of the maximal

distance from one node to all other nodes and *efficiency* [21], which is the average of the inverse of the distance from one node to all other nodes, are based on the characteristic path length of a node. Another important concept in this context is the so-called *betweenness centrality* [22,23] that counts<sup>2</sup> the number of times that a node lies in the path between the others. Some of these measures built for undirected and unweighted networks may be generalized for the case of directed and weighted networks [25,4]. For instance, if one can associate weights to the edges of a network, the degree of a node can be generalized to be the *in-strength* and *out-strength* of a node. While the former is the sum of the weights of the edges that arrive to a given node, the later is the sum of the weights of the edges that leave a given node. Furthermore, if one is dealing with a geographic (spatial) network where the distances between the nodes are previously defined, closeness centrality, graph centrality and efficiency may be trivially extended for the case of directed and weighted networks.

\* Corresponding author.

E-mail addresses: bernardo@fis.unb.br (B.A. Mello), rogerio.boueri@ipea.gov.br (R. Boueri), danielcajueiro@unb.br (D.O. Cajueiro).

<sup>1</sup> This concept is a kind of first order approximation to the centrality of a node, since it considers only the neighbors of first order.

<sup>2</sup> The way that the number of times that a path lies in the path between the others is counted may vary depending on the definition. For instance, one may consider any path or only the shortest paths between two nodes. A discussion about this topic may be found in [24].

An interesting revision of measures of centrality may be found in [26].

A constraint that usually arises is when the distances between the nodes are not previously defined, but only the weights of the edges of a given network. In this situation, the issue is how to define the distance from the previously defined weights. A path to circumvent this problem is to use, for instance, in-strength or out-strength measures defined above or the so-called *dominance power* [27] of a node, which is a measure of the influence of a node in all other nodes of the network relative to the influence of all other nodes (this measure will be precisely defined in Section 2). However, as in the case of the strength measure, this measure also only considers the neighbors of the first order.

In this Letter, we analyze the centrality of cities in two networks of the flow of information among these cities, namely a network of phone calls among the cities and a network of radio stations. Since we do not have a previously defined distance among the nodes of this network, we define here the so-called *diffusion power* – an extension of the dominance power, which considers the interaction between neighbors of higher orders. Furthermore, we explain the centralities of the cities evaluated using the diffusion power in terms of the specific characteristics of the cities that belong to the network.

The subject has some resemblance with the topic of Opinion Formation, since, in that field, there is information flow among agents [28] and it may depend on an influence network [29].

The remainder of this Letter is structured in the following way. Section 2 revises the measure known as the dominance power and defines the concept of diffusion power. Section 3 describes the process to construct the networks. Section 4 presents the results of the application of the diffusion power in the evaluation of the centrality of the cities that belong to the networks. Finally, Section 5 presents the main conclusions of this Letter.

## 2. Diffusion power

In order to define the diffusion power, we first revise the measure of centrality known as dominance power introduced by [27]. For a directed weighted network, the dominance power measures the influence of a node  $i$  on its neighbors normalized by the influence of all other nodes of the network on these neighbors as given by equation

$$\beta(i) = \sum_{j \neq i} \frac{w_{ij}}{s_{in}(j)} \quad (1)$$

where  $w_{ij}$  is the weight of the edge that comes from  $i$  and goes to  $j$  and  $s_{in}(j)$  is the in-strength of node  $j$  given by

$$s_{in}(j) = \sum_k w_{kj}. \quad (2)$$

As already stressed, since this measure defines centrality considering only the neighbors of first order, it does not measure the effect of the propagation of the influence of a node by its neighbors.

Therefore, we extend this concept accounting for the propagation of the higher order neighbors as given by equation

$$D(i) = \sum_{j \neq i} \frac{[1 + fD(j)]w_{ij}}{s_{in}(j)} \quad (3)$$

where  $f$  is a free parameter defined in the interval  $[0, 1)$  that measures the effects of higher order interaction among the nodes of a network. Furthermore,  $f$  defined in this interval ensures that the solution of the linear system defined in Eq. (1) may be found iteratively and it will converge to a fixed point [30]. Finally, if  $f = 0$ , the diffusion power is equal to the dominance power.

The idea proposed here to evaluate the diffusion power is quite similar to the idea of the Google founders [31] used to calculate the page rank:

$$P(i) = (1 - d) + d \sum_{j \neq i} w_{ji} \frac{P(j)}{s_{out}(j)}. \quad (4)$$

In the page rank expression above,  $w_{ji}$  is 1 if there is a link pointing from page  $j$  to page  $i$ , and zero otherwise.

The damping factor ( $d$ ) used in page rank represents the probability that a surfer who reaches a given page clicks in a link on that page, moving to the next page. In our model, the damping factor is the probability that an information received by a node is forwarded to another node. There is not a consensus about the value that should be used for the damping factor, but the original paper from Brin and Page [31] suggested using 0.85.

A detailed discussion of the page rank may also be found in [26]. It is also worth remarking that the page rank has been extended in several directions. A revision of these attempts may be found for instance in [32].

## 3. Description of the networks data

In this Letter, by measuring the centrality among the cities, we are trying to find the most influential cities. In particular, we consider here two different networks of Brazilian cities, namely an undirected weighted network where the weight is the number of phone calls from one city to the other and a directed network where the weight is the number of radio stations of a city that reaches the other city.

### 3.1. The network of calls among the cities

Although there are few examples of networks of calls considered in the literature such as [33,34], we believe that the idea of using a network of call in order to identify the most influential cities is new.

The network considered here is based on a data file provided by the Brazilian Agency of Telecommunications with information about the number of calls between two cities and the total number of minutes of the calls between two cities. The data was assembled from all phone calls between wired phones located in one of the 5555 Brazilian municipalities which have access to this service during the month of September of 2008. As one may note, this data allowed to build a directed network. However, we decided to build an undirected network since in a phone call the information flows in both directions. Furthermore, instead of building the weights of the network based on the number of minutes, we decided to consider the number of calls, since most relevant information is given in few minutes.

### 3.2. The network of radio stations

As far as we know, a network based on radio stations is a totally new idea. In Brazil, mainly in the countryside, radio stations still have a fundamental role in the dissemination of information between close cities.

The network of radios was built here based on a data file provided by the Brazilian Agency of Telecommunications and the Ministry of Telecommunications. Two cities are connected in this directed and weighted network if at least one radio station from one city reaches the other city. The network is directed and the weight of the connection is used to count the number of the stations of the source city that reaches the target city.



Fig. 1. Network of Brazilian FM stations.

In this Letter, only FM stations were considered. This is due to the difficulty of evaluating the range of an AM station and the fact that the number of FM stations are much larger than the number of AM stations. The network, shown in Fig. 1, was built using the 7102 FM radio stations working in Brazil. The range of a radio station depends on whether the receptor is or not located in the countryside. In fact, the presence of high buildings in urban areas hinders the reception of low intensity signals. Therefore, since in most Brazilian cities the presence of high buildings is scarce, we decide to build the network based on the range of a radio station evaluated to work in the countryside.

4. Results

In this section we evaluate the diffusion power to the network of calls among the cities and the network of radio stations. It is clear that there is a free parameter  $f$  in the evaluation this measure. Fig. 2 shows the effect of the variation of  $f$  in the ranking of the centrality of the cities of these networks, which  $f = 0$  recovers the dominance power. The effect of the variation of  $f$  is bigger in the network of calls between cities, which is coherent with the higher connectivity of that network compared to the network of radio stations, leading to stronger second order effects.

Unfortunately, we do not have an explicit methodology to choose the value of  $f$ . However, the extremes of the interval  $[0, 1]$  should not be used, since with  $f = 0$  there is no high order flow of information and with  $f = 1$  every piece of information diffuses from one node to the other. We proceed here using an intermediate value, i.e.,  $f = 0.5$ .

Table 1 shows the Spearman correlation coefficients between the diffusion power evaluated using  $f = 0.5$  and other measures of centrality such as dominance power (also present in Fig. 2), strength, geographic closeness centrality and geographic efficiency. The geographic closeness centrality and geographic efficiency are evaluated using the Euclidean distance between the cities.

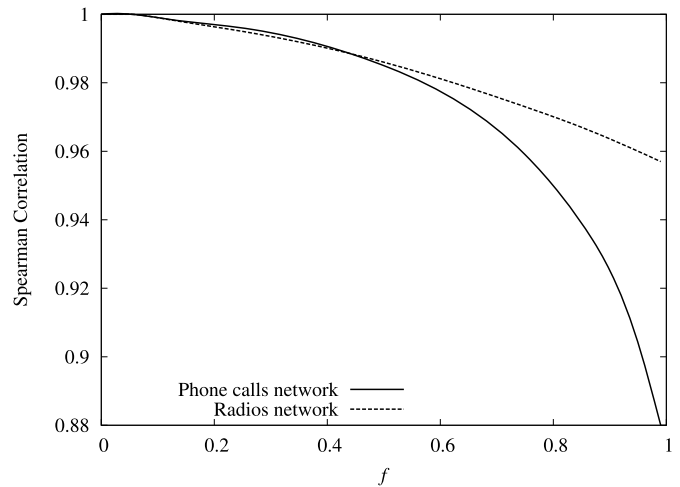


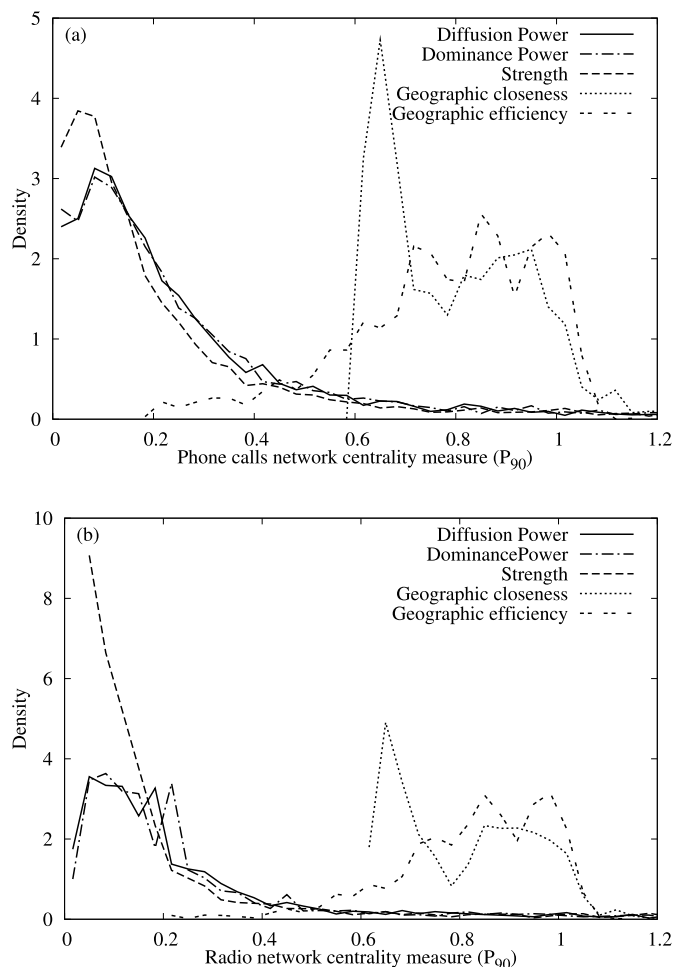
Fig. 2. The Spearman correlation coefficients of the diffusion power evaluated between the rank resulting from  $f = 0$  (dominance power) and the rank resulted from  $f \neq 0$  for the network of calls among the cities and for the network of FM radio stations.

Table 1

Spearman correlation coefficients between the diffusion power, other measures of centrality and economic data.

	Phone calls	Radio
Dominance power	0.985	0.986
Strength	0.99	0.86
Geographic closeness centrality	0.024	0.033
Geographic efficiency	0.037	0.073
Population	0.94	0.55
GIP per capita	0.034	0.048

A surprising result is the very low Spearman correlation coefficients between the geographic measures and the diffusion centrality. However, an intuitive result is that the Spearman correlation



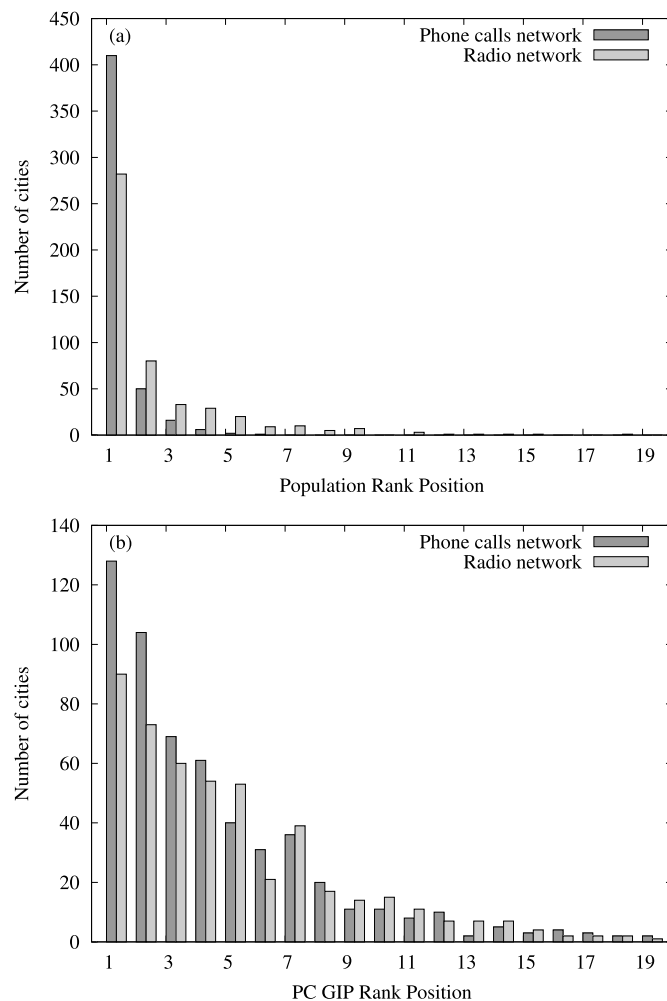
**Fig. 3.** Distribution of several centrality measures for the (a) phone call network and (b) the radio network. The scale of the x-axis is the 90th percentile, i.e., the area of the curves between 0 and 1 is 0.9. The radio network distribution include only the 3469 municipalities with centrality higher than zero.

coefficients for the geographic measures associated to the radio network are larger than the ones associated to the phone calls network. This happens since the FM transmission clearly depends on the distance between the cities.

Fig. 3 shows the distribution of several centrality measures applied to the phone call network and to the radio network. The same similarities between the diffusion power, the dominance power and the strength shown in Table 1 are verified here, placing those centrality measures apart from the geographic measures.

Brazilian cities are clustered in 535 “microregions” which are legally defined administrative areas consisting of spatially connected municipalities. The purpose of our study was to find the municipalities with higher diffusion power within each microregion. We ranked the cities within each microregion according to the diffusion power derived from each network. The comparison of these rankings with the ranking of population and GIP can be seen in Fig. 4.

While Fig. 4(a) shows a strong correlation between having large population and having large diffusion power, Fig. 4(b) shows that this trend is not so strong regarding larger GIP per capita. This is an interesting result since we are explaining the centrality of the cities using economic variables. The same trend can be seen in Table 1. It is interesting to note that, even with the very low Spearman correlation of the GIP per capita, the most connected cities as measured by the diffusion power are usually those with higher GIP per capita, as can be seen in Fig. 4.



**Fig. 4.** Two ranks were build for each of the 535 Brazilian microregions, with cities ordered by their population and GIP per capita. The position in the these two ranks of the cities with higher diffusion power (according to the phone call network and according to the radio network) within each microregion was found. The histograms shows the frequency of each position in the rank ordered by (a) population and (b) GIP per capita.

We have also found that in 57% of the microregions the same cities that are ranked firstly in one network are also ranked firstly in the other network.

### 5. Conclusions

In this Letter, we have defined the so-called diffusion power in order to analyze two networks of the flow of information among these cities, namely a network of calls among the cities and a network of radio stations.

The diffusion power is an extension of the dominance power that considers the neighbors of higher order in the calculation of the centrality. Since, as in the case of the page rank [31], the diffusion power has a free parameter, we have investigated the effect of this parameter in the relative centrality of the cities.

Finally, a very interesting result of this Letter is that we can explain the results of the centralities of the cities by means of their individual economic characteristics.

### Acknowledgements

Bernardo A. Mello and Luiz H. Batistuta thank IPEA for financial support. Daniel O. Cajueiro is grateful to the Brazilian Agency CNPq

for financial support. The authors also thank Roberta da Silva Vieira who is at Instituto de Pesquisa em Economia Aplicada (Brazil) for valuable discussions; Reno Martins who is at the Brazilian Agency of Telecommunications for providing the data; and Thiago Aguiar Soares who is at the same agency, for explanations about radio signal propagation and for providing the ranges of FM stations in the countryside.

## References

- [1] W. Li, X. Cai, *Phys. Rev. E* 69 (2004) 046106.
- [2] G. Iori, G.D. Masi, O.V. Precup, G. Gabbi, G. Caldarelli, *J. Econom. Dynam. Control* 32 (2008) 259.
- [3] G. Iori, R. Reno, G. DeMasi, G. Caldarelli, *Phys. A* 376 (2007) 467.
- [4] D.O. Cajueiro, B.M. Tabak, *Phys. A* 387 (2008) 6825.
- [5] R. Pastor-Satorras, A. Vazquez, A. Vespignani, *Phys. Rev. Lett.* 87 (2001) 258701.
- [6] M. Anghel, Z. Toroczkai, K.E. Bassler, G. Korniss, *Phys. Rev. Lett.* 92 (2004) 058701.
- [7] D.O. Cajueiro, R.S. DeCamargo, *Phys. Lett. A* 355 (2006) 280.
- [8] M. Rosvall, A. Trusina, P. Minnhagen, K. Sneppen, *Phys. Rev. Lett.* 94 (2005) 028701.
- [9] M. Rosvall, A. Gronlund, P. Minnhagen, K. Sneppen, *Phys. Rev. E* 72 (2005) 046117.
- [10] D.O. Cajueiro, *Phys. Rev. E* 79 (2009) 046103.
- [11] D.O. Cajueiro, R.F.S. Andrade, *Europhys. Lett.* 87 (2009) 58004.
- [12] A. Grabowski, R.A. Kosinski, *Phys. Rev. E* 73 (2006) 016135.
- [13] D.O. Cajueiro, *Phys. Rev. E* 72 (2005) 047104.
- [14] M. Boguña, R. Pastor-Satorras, A. Diaz-Guilera, A. Arenas, *Phys. Rev. E* 70 (2004) 056122.
- [15] E.P. Borges, D.O. Cajueiro, R.F.S. Andrade, *Eur. Phys. J. B* 58 (2007) 469.
- [16] R. Albert, A.L. Barabasi, *Rev. Modern Phys.* 74 (2002) 47.
- [17] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.U. Hwang, *Phys. Rep.* 424 (2006) 175.
- [18] L.D. Costa, F.A. Rodrigues, G. Travieso, P.R.V. Boas, *Adv. Phys.* 56 (2007) 167.
- [19] G. Sabidussi, *Psychometrika* 31 (1966) 581.
- [20] P. Hage, F. Harary, *Social Networks* 17 (1995) 57.
- [21] V. Latora, M. Marchiori, *Phys. Rev. Lett.* 87 (2001) 198701.
- [22] L.C. Freeman, *Sociometry* 40 (1977) 35.
- [23] L.C. Freeman, *Social Networks* 1 (1979) 215.
- [24] M.E.J. Newman, *Social Networks* 27 (2005) 39.
- [25] M. Barthélemy, A. Barrat, R. Pastor-Satorras, A. Vespignani, *Physica A* 346 (2005) 34.
- [26] C. Kiss, M. Bichler, *Decis. Sup. Syst.* 46 (2008) 233.
- [27] R. van den Brink, R.P. Gilles, *Social Networks* 22 (2000) 141.
- [28] M. Granovetter, *Amer. J. Sociol.* 83 (1978) 1420.
- [29] N.E. Friedkin, E.C. Johnsen, *Adv. Group Proces.* 16 (1999) 1.
- [30] E. Kreyszig, *Introductory Functional Analysis with Applications*, Wiley, 1978.
- [31] S. Brin, L. Page, The anatomy of a large-scale hypertextual web search engine, in: 7th International World Wide Web Conference, Brisbane, Australia, 1998.
- [32] M. Thelwall, L. Vaughan, *Aslib Proc.* 56 (2004) 24.
- [33] W. Aiello, F. Chung, L. Lu, A random graph model for massive graphs, in: Proceedings of the thirty-second annual ACM symposium on Theory of computing, STOC '00, New York, 2000, pp. 171–180.
- [34] A.A. Nanavati, S. Gurumurthy, G. Das, D. Chakraborty, K. Dasgupta, S. Mukherjee, A. Joshi, On the structural properties of massive telecom call graphs: Findings and implications, in: Proceedings of the 15th ACM international conference on Information, CIKM '06, New York, 2006, pp. 435–444.