Les Cahiers du GERAD ISSN: 0711–2440

Exploring centrality measures and their extensions through an influence network

G. Caporossi, S. Perron,
L.-Ph. Sirois, J. Turbide

G–2014–28

May 2014

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

Exploring centrality measures and their extensions through an influence network

Gilles Caporossi [a]
Sylvain Perron [a]
Louis-Philippe Sirois [b]
Johanne Turbide [b]

[a] *GERAD & HEC Montréal, Montréal (Québec) Canada H3T 2A7*

[b] *HEC Montréal, Montréal (Québec) Canada H3T 2A7*

gilles.caporossi@gerad.ca
sylvain.perron@hec.ca
lps@hec.ca
johanne.turbide@hec.ca

May 2014

Les Cahiers du GERAD

G–2014–28

Abstract: In this paper, we propose an empirical study of the centrality of actors in network. The data was collected among publicly available information of the boards members of organisations, including charities in Québec. A first contribution of this paper is a set of new measures of centralities and variations on classical measures. From the application point of view, the strategical position of professionals (accountants and lawyers in this case) within the network is shown.

Key Words: Social Network Analysis, Centrality.

Acknowledgments: This work was supported by the *Fondation des comptables agréés du Québec*, the *Fondation HEC Montréal* and NSERC (Canada).

Introduction

Some structures could be used by individuals searching to increase their influence, mainly because they could meet other individuals with whom they may exchange information or gain influence. It turns out that some networks are more suitable than others for this task. It is the case for networks of board of directors of organizations, particularly those of charities and other nonprofit organizations, along with business associations such as boards of trade. Indeed, board membership of such organizations is partly motivated, at least for some individuals, by the potential benefits one gains from inter-personal networking opportunities that may result; that is, opportunities to develop one's human capital. This is especially true for business professionals, such as accountants and lawyers, as their personal contacts network and reputational capital is fundamental to the success of their practice.

As such, we argue that business professionals are more likely than the general population to volunteer as board members of nonprofit organizations and business associations for networking purposes. In turn, this is likely to have an impact on the importance and role these individuals play in a large network of board of directors. Accordingly, we evaluate the performance of network common centrality measures along with some extensions by comparing those measures for the subgroup of business professionals to those of other non-professionals individuals included in a large network of board members. Our network consists of close to 55,000 organizations with over 418,000 board members.

This network is analysed using a wide variety of centrality measures. Indeed, the classical measures were developped for various goals and it is reasonable to think they are not necessarily related to each other. In this article, we expose some of those classical measures and introduce some new ones in order to better evaluate the effectiveness of the networking activity of the professional individuals through the network.

As some measures could be adjusted to better fit some special need, for example by increasing the importance of close influence from an individual to other, we will propose ways to adapt some of these measures. The comparision of the classical, new and adjusted measures will be presented. This research lead to some unexpected results strenghtening the conclusion of the analysis.

1 The Context

There are many reasons why individuals perform volunteer work, and both intrinsic and extrinsic motivations play a role. Proponents of the utilitarian view of volunteering argue and find that extrinsic motivations dominate. This stream of the literature suggests that volunteers invest, through their implications, in their human and social capital by acquiring special skill sets, expanding and deepening their social contacts and overall social network, and signaling their willingness to perform (e.g., [6]). These findings are echoed by repeated survey evidence acknowledging that some volunteers get involve to meet people and develop their personal network of friends and contacts, and some simply to promote their careers (e.g., [5, 9]).

Investing in one's human capital, including personal network of contacts and reputational capital, is especially valuable for business professionals, such as accountants and lawyers, both for their career and the organisations they work for. Indeed, the quality of the services these professionals provide is based on reputational and personal qualities and is overall difficult to assess otherwise. Information about professional firms in general, and individual professionals in particular, flows through formal and informal networks of contacts, such as client referral or college alumni [7]. Among strategies employed to promote and enhance their social network, business professionals frequently engage in volunteer work [1, 8], such as membership to the board of directors of charitable or nonprofit organisation or business associations. In fact, it is not uncommon for lawyers and professional accountants to boast their volunteer work and community involvement on their firm's website[1] or social networking webpage.[2] Moreover, members of accounting and law firms are specifically encouraged to serve on boards of non-profit organisations.[3]

Hence, the board of directors of nonprofit organizations and business associations serves as a vector through which business professional can hope to interact with executives and directors of "for-profit" organisations. Indeed, the latter individuals also frequently serve on such boards for both intrinsic and extrinsic motivations. From this networking activity, business professionals can boost their social capital and possibly recruit new (or simply maintain) "for-profit" clients. We argue the motivation for business professionals to actively engage in networking activities leads them to occupy a more central role than most in a large network of management boards. Accordingly, this offers a references point.

1.1 Data description

We limit our analysis to organisations registered in the Canadian province of Québec given the availability of data. All active organisations in Québec must register to the "Registre des entreprise" (Registrar).[4] These are known as "enterprises" and they must file an annual declaration form to maintain their active status. The Registrar keeps information on individual enterprises (e.g., address of business, operating and legal name, main and secondary economic sector of operation, etc.). More importantly for our study, the list of "owners" (i.e., main shareholders for corporations or all partners for partnerships), along with the name of administrators (i.e., directors for corporations and associations, and partners for partnerships) is also available, with the corresponding address of the individuals. Each enterprise has a unique identification number.

We obtained this information for relevant active registered enterprise as at August 2012 through a combination of Québec's Access to information Act and manual extractions from the Registrar's public web-site. Organisations are grouped into three categories: 1) charities and nonprofit organisations,[5] 2) business associations,[6] and 3) "for-profit" enterprises.

The group of "for-profit" enterprises is included as a key component of the network. Indeed, as argued above, individuals' involvement on a nonprofit's or business association's management board may in part be motivated to establish links with key representatives of the "for-profit" sector, especially for business professionals. These organisations are identified from various sources.[7] The objective is to capture a significant portion of the Québec business sector.

---

[1] See for example: `http://www.fasken.com/en/lawyers/detail.aspx?professional=3987`.

[2] See for example: `ca.linkedin.com/pub/luc-villeneuve/b/831/601`.

[3] For example, the US affiliate of Deloitte, a large international public accounting firm, states on the website: "Our people are encouraged to serve on boards of nonprofit organizations in their community. Nearly half of our partners, principals and directors currently serve on at least one board." (see: `www.deloitte.com/view/en_US/us/About/Community-Involvement/nonprofit-board-service/index.htm`, accessed April 3rd, 2014). Stikeman-Elliott, a large Canadian law firm writes on its website: "(...) The firm will match donations of up to $5,000/year/person for firm members who sit on charitable boards and who also make a financial contribution (...)" (see: `www.stikeman.com/cps/rde/xchg/se-en/hs.xsl/12257.htm?`, accessed April 3rd, 2014).

[4] See: `http://www.registreentreprises.gouv.qc.ca/en/`.

[5] Not-for-profit organisations are identified as all active enterprises in the Registrar incorporated under *Part III of the Companies Act* (provincial or federal regime), which applies to non-profit organisations. This represents the majority of nonprofit organisations in Québec. For other nonprofit organisations not incorporated under this law, we complement this list by adding all active charities and foundations residing in Québec and registered with the Canada Revenue Agency that had filled their T3010 2011 return by September 2012. Registered charities are nonprofit organisations registered with the Agency and that can issue tax receipts for donations. This is the case, for example, of faith-based organisations. Finally, we exclude student associations incorporated under *Part III of the Companies Act*.

[6] "Business associations" also operate as nonprofit organisations but are incorporated under a different law and are not charities. They include for the most part chambers of commerce or professional association. We identify these organisations by selecting all active enterprises reporting "Commercial Associations" as their main or secondary economic activity, regardless of the constituting law or judicial form.

[7] These are comprised of 270 publicly listed companies based in Québec to which we add companies from the *Les Affaires* (Québec based business weekly) Top 500 of Québec companies and Top 300 Small-Medium-Enterprise listings for 2011 (`http://www.lesaffaires.com/classements/`), and all individually managed member branches and entities associated with the cooperative financial group Desjardins, a key economic player in Québec (`http://www.desjardins.com/ca/about-us/index.jsp?navigMW=pp&`).

In total, the base data for our network consists of 54,485 organisations and 418,580 board members, with a vast majority of 52,666 as nonprofit organisations (402,417 members), 1,129 business associations (11,168 members) and 1,282 "for-profit" organisations (10,410 members). The size distribution of management boards for all organisations in the sample is presented on Table 1 (the size distribution is similar across all three organisation types).

Table 1: Size distribution of management boards for all organisations in the sample

Total number of organisations	54485
Total number of board members	418580
Distribution of board members per organisations	
Mean	7.68
Standard deviation	4.94
Min	2
1st decile	3
Q1	4
Median	7
Q3	10
10th decile	14
Max	66

Multiple board memberships, key for linking organisations and individuals together, are identified by matching individual entries per organisations to entries of all other organisations in the database based on standardised full names and postal codes of personal addresses as reported in the Registrar. Perfect matches are assumed to accurately identify a unique person. Several manual random checks confirm this assertion.

Business professionals are identified as partners (owners) in a partnership or board directors for corporations of all active enterprises operating in Québec as accounting or law firms.[8] The names and addresses of these individuals are kept and matched to the full sample of 54,485 organisations and 418,580 board members. Note that this definition of business professionals does not include all practicing registered professional accountant or lawyer in Québec. Indeed, we retain only a relatively small portion of these professionals that are also business owners; usually the most senior and influential representatives of their respective firms who are also more likely to engage in networking activities.

As a result of the matching process, we note that the total 418,580 board members in the network correspond to a total of 350,427 unique individuals, with 1,703 identified as business professionals. Table 2 presents the distribution of individual board memberships by individual type. We note that multiple board memberships are overall rare, although more frequent for professionals, as expected.

1.2 Description of the network

There are different ways to represent the networking information from the boards of "non-profit" organizations. In one of them, each vertex corresponds to an actor, an individual, and a link between two actors occurs if they belong to at least one common board. A variations of such a network could involve weights representing the strength of that link. Another model involves both the individuals and the boards, a link representing the belonging of an individual to a board. In such a model, there cannot be a link between two individuals or between two boards. In such a model, links can only lie between nodes of different kind (individuals vs boards). Such a graph is called "bipartite" as it is possible to split its nodes in two groups and only observe links from one group to the other, but not among the same group.

[8] Specifically, we retain all active enterprise reporting to the Registrar "Office of accountants and professional accountants" or "Law and public notary offices" as their main or secondary economic activity, regardless of the constituting law or judicial form.

Table 2: Distribution of individual board membership per individual type

Number	Non-prof Frequency	Non-prof %	Non-prof Cum. %	Prof Frequency	Prof %	Prof Cum. %
1	302282	0.86	0.87	1131	0.66	0.66
2	34073	0.098	0.96	333	0.20	0.86
3	7867	0.02	0.99	135	0.08	0.94
4	2619	0.01	0.99	65	0.04	0.98
5	938	0.002	1	16	0.009	0.99
6	461	0.001	1	7	0.004	0.99
7	238	0.0006	1	7	0.004	0.99
8	116	0.0003	1	3	0.002	1
9	61	0.0002	1	3	0.002	1
10	36	0.0001	1	1	0.0006	1
11 +	33	9.46307e-05	1	2	0.001	1
Total	348724			1703		
Mean	1.192524747			1.596007046		
Max	16			13		

However, whatever the model, the graph is not connected, and is composed of 18,700 connected components, one of them being very large while the others are very small. For example, the second largest component has 164 vertices and 166 edges with a maximum degree 18 and 124 pending vertices, which reveals the nature of these parts of the graph that consists in rather isolated organizations that could certainly not be considered for "networking" purpose.

This main component involves 230,765 individuals related to 32,597 organizations. The bipartite graph therefore has 263,362 vertices. The number of edges in the bipartite graph is 294,280, the graph is extremely sparse. In the case of the non bipartite model, the number vertices is 230 765 (the number of individuals) and there are 1,583,312 edges, which is also rather sparse.

The large number of vertices in the main component suggests that the corresponding organizations are good places for "networking", the most important part of the individuals being directly or indirectly connected.

From a technical point of view, this main component being huge, some measures cannot be computed. For example, measures that explicitly requires the use of a $(n \times n)$ matrix cannot be considered as the required memory is too large. It is the case of the random walk based betweenness centrality proposed by Newman [10].

For the purpose of this work, a parallel program was developed in order to handle the graph and compute the most relevant measures.

2 Some classical centrality measures

Prior to the definition of the centrality measures used here, let us give some notation. Let $G = (V, E)$ be an undirected simple graph without loops and multiple edges. Note $n = |V|$ the number of its vertices (or actors) and $m = |E|$ the number of its edges (or links). Note $A = \{a_{ij}\}$ be its adjacently matrix, with $a_{ij} = 1$ iff i and j are joined by an edge. Note d_{ij} the geodesic distance between vertices i and j.

- **Degree centrality.** The degree d_i of the vertex i is the number of edges adjacent to the vertex i.

$$d_i = \sum_{j=1}^{n} a_{ij}. \tag{1}$$

- **Extended degree centrality.** The extended degree centrality ed_i^α is the number of vertices that are at distance at most α from i. Here, we will only consider $\alpha = 2$.
- **Eccentricity centrality.** The eccentricity centrality is computed from the maximum distance from the vertex i as follows :
$$ecc_i = \frac{1}{\max_j d_{ij}}. \tag{2}$$
- **Closeness centrality.** The closeness centrality c_i is based upon the sum of the distances from i to all the other vertices as follows :
$$c_i = \frac{1}{\sum_{j=1}^n d_{ij}}. \tag{3}$$
- **Harmonic centrality.** The harmonic centrality h_i of the vertex i is the sum of the reciprocical of the distances from i to all the other vertices as follows :
$$h_i = \sum_{j=1}^n \frac{1}{d_{ij}}. \tag{4}$$
- **Betweenness centrality.** The betweenness centrality [3, 4] is a centrality measure based upon the number of shortest paths between pairs of vertices that uses the considered vertex.
$$b_i = \sum_{j=1}^n \sum_{k=1}^n \sum_{l=k+1}^n \frac{s_{ij}^{kl}}{s^{kl}}. \tag{5}$$
where s_{ij}^{kl} is the number of shortest paths between vertices k and l that use the edge (i, j) and s^{kl} is the total number of shortest paths between k and l. We note that in the context of social networks, Newman [10] suggests to consider a random walk based measure, instead of using the shortest path, for the computation of the betweenness, but the computation of this centrality requires the use of a $n \times n$ matrix, which is not possible for the current application for memory reasons.
- **Eigenvector centrality.** The eigenvector v associated to the largest eigenvalue λ_1 of the adjacency matrix A is another centrality measure. It is a value such that the equation
$$e_i = \frac{1}{\lambda_1} \sum_{j=1}^n a_{ij} E_j \tag{6}$$
is respected for all i.
- **Random walk centrality.** Suppose that we randomly move in the network, starting at a random vertex. The Random walk centrality is the defined in a similar way as the eigenvector centrality by the following equation :
$$r_i = \sum_{j=1}^n \frac{r_j}{\sum_{k=1}^n a_{jk}}. \tag{7}$$

3 Distance weighted centrality measures

When studying centrality measures in a social network, it is reasonable to assume that the influence of a vertex i on a vertex j will likely decrease when the distance d_{uv} between i and j increases. We the influence of a pair i, j of vertices on centrality measures should decrease when this distance increases.

However, for example in the equation (5), a pair of vertices (i, j) will globally have more impact if d_{ij} is large, because it will add one unit to the betweenness of $d_{ij} - 1$ vertices. Such a measure is very well fitted for telecommunication networks, where a connection between vertices that are far from each other will have impact on a larger number of transmitters, but it is not the case in a social network.

To correct this behavior, one must weight the contribution of each pair of vertices to reduce it when the distance increases. The betweenness definition b_i of the vertex i from the equation (5) could then be replaced by :

$$b_i^\alpha = \sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=k+1}^{n} \frac{s_{ij}^{kl}}{s^{kl}} \frac{1}{d_{kl}^\alpha}, \tag{8}$$

where α is a non negative constant. Increasing the value α will reduce the impact of the pairs of vertices when their distance increases.

If $\alpha = 0$, $b_i^\alpha = b_i$. If $\alpha = 1$, we have $\sum_{i=1}^{n} b_i^\alpha = \frac{n(n-1)}{2}$, which means that the overall contribution of each pair of vertices is the same. When α tends to infinity, b_i^α tends to the degree d_i.

In the context of social networks, it is expected that information does not circulate between vertices that are at larger distance, or it does not as much as for closer pairs. Should the centrality of a vertex in the context of a social network represent the real flow of information that is forwarded by the corresponding, it is likely that larger values of α would better fit the reality.

Beyond the betweenness centrality, it is possible to define another eigenvector centrality in a similar way. Instead of using the adjacency matrix, one could also consider vertices that are not adjacent to the vertex i, but with an influence that decreases with the distance. Hence, instead of using the adjacency, one could use the weighted inverse distance matrix $IWD = \{w_{ij}\}$, with $w_{ij} = 1/d_{ij}^\alpha$. The first eigenvector of the matrix IWD could be computed in a similar way as the first eigenvector of the adjacency matrix, using the power iteration algorithm by the mean of the following equation :

$$de_i^\alpha = \frac{1}{\mu_1} \sum_{j=1}^{n} \frac{1}{d_{ij}^\alpha} de_j^\alpha. \tag{9}$$

This formulation is an extension of the eigenvector centrality, since both measures tend to converge if α tends to infinity.

4 Variations on the betweenness centrality

Betweenness centrality, specially in its distance weighted version seems appropriate to identify key persons, i.e., actors which have a key position for the transmission of information through the network. However, when the whole network extends beyond *the world* of individuals, some different features seems more important. Indeed, more that being in the center of the whole network, which is actually a tool that goes beyond the influence of any individual, being at the intersection of different groups seens more important from a practical point of view. In this sens, the clique centrality which counts the number of cliques to which an actor belongs seems more appropriate. This measure is used to identify weak points for the spreading of deceases, it could also make sense in the context of information transmission. In the current graph, each clique finally corresponds to a board, and the clique centrality would correspond to the degree centrality in the bipartite representation of the network. The clique is also a very local measure, from that point of view, very comparable to the degree centrality. Instead, we propose a variation of the betweenness centrality.

The proposed measure is based upon a remark : il the network is a tree, then the computation of the betweenness centrality could be measured by the following equation :

$$b_i = \sum_{j/(i,j)\in E} n_i \times n_j, \tag{10}$$

where n_i is the number of vertices that are closer to i than j and n_j the reverse. Of course, applying this formula is much faster than counting the shortest paths if the distance information is available.

It turns out that this definition could be extended to the case of graphs in general, even if the result will differ from the classical betweenness.

Each edge being associated to a value $b_{ij} = n_i \times n_j$ that will increase if the edge (i, j) makes a link between communities, and if these communities are large. This centrality based upon geodesic distances could be called geodesic centrality, defined as follows :

$$g_i = \sum_{j/(i,j)\in E} n_i \times n_j. \tag{11}$$

The geodesic centrality seems to be a good measure to identify actors that make links between communities.

It turns out that the geodesic centrality is a good measure to identify vertices that are at the intersection of various communities, but its performance is reduced when the neighbors of the vertex i are connected together, in which case they are at the same distance from i and j, and are therefore omitted.

To correct this problem, another measure, very similar could consist in using pairs of neighbors of i, instead of edges adjacent to i. This measure would better measure the splitting capability of the vertex (how removing i would separate communities). The so called split centrality would then be defined as follows :

$$s_i = \sum_{j/(i,j)\in E} \sum_{k/(i,k)\in E} n_j \times n_k. \tag{12}$$

From a computational point of view, the geodesic and split centralities are much easier to compute than the betweenness centrality if the distance matrix is available. If not, the betweenness centrality is easier to compute than the geodesic, which in turn is faster to compute than the split centrality (specially for vertices with high degrees).

In order to better represent the reality of social networks, these two measures could also be weighted by a distance function, as we proposed for the betweenness, as follows.

The distance weighted geodesic centrality is thus :

$$g_i^\alpha = \sum_{j/(i,j)\in E} f_i^\alpha \times f_j^\alpha, \tag{13}$$

where

$$f_i^\alpha = \frac{1}{d_{ik}^\alpha} \quad \forall k/d_{ik} < d_{jk}. \tag{14}$$

Similarly, the distance weighted split centrality is defined as follows :

$$s_i^\alpha = \sum_{j/(i,j)\in E} \sum_{k/(i,k)\in E} f_j^\alpha \times f_k^\alpha. \tag{15}$$

5 Relations between centrality measures in the network under study

The first question that arises is the choice of the proper model to use. The bipartite model, even if it may not appear as natural as the other seems more appropriate. Indeed, in the case of the non bipartite model, as soon as two individuals are associated to the same organization, they will share as common neighbors all the members of that organisation. The non bipartite network representation will then be composed of cliques corresponding to organizations. Should one of those organization have a huge number of members, the graph will have a very large clique, which will drastically affect the nature of the network, even if the interactions among the members of that organization are limited (due to the large number of individuals involved).

From a computational point of view, the bipartite graph is easier to handle for most centrality measures, and it turns out that in the present case, the results are comparable. Table 3 indicates the correlation between the values obtained for the bipartite and the non bipartite models. Except for d, rw and ed^2, the measures for both models are very correlated (more than 0.99). It is then reasonable to think that the results will not depend on the model for the present study. However, the split centrality measure, which is sensitive to the degree of vertices is much longer to compute and was not achieved on the non bipartite model for that reason.

Table 3: Correlation between bipartite and non bipartite models for each centrality measure

c	1	b^3	0.999	de^5	0.97
ecc	1	b^4	0.999	g^0	0.873
h	1	b^5	0.998	g^1	0.898
ed^2	0.774	rw	0.642	g^2	0.925
d	0.641	de^1	1	g^3	0.95
b^0	0.999	de^2	1	g^4	0.943
b^1	0.999	de^3	1	g^5	0.88
b^2	0.999	de^4	0.992	e	0.84

In general, centrality measures are not necessarily correlated, therefore the choice of the proper centrality measure is important. In the present case, the correlation of two distinct measures is never below 0.2, ant the worst values involve the eccentricity ecc, which is not surprising as this measure is clearly not robust.

5.1 Correlation among classical measures

Table 4 is the correlation matrix of the main centrality measures that were tested against the model. Except between eccentricity and betweenness which is 0.22 (in italics), the correlations are always at least of 0.29. In bold face are represented the strong correlations (0.9 or more). We notice that there are 4 groups of measures. (i) the distance based measures, ecc, c and h, (ii) the extended degree-2 and the eigenvector centrality, (iii) the degree and random walk centralities and (iv) the betweenness centrality.

Among those groups, the eigenvector/extended degree and random walk/degree are rather well correlated (0.68 and 0.74). There are then 3 groups. The distance based measures, the degree based measures (which are related to the eigenvector and the random walk centralities) and the betweenness centrality. Should the betweenness centrality be associated to one of the otehr groups, it would be associated to the degree based measures with correlations higher than 0.50 while these values are at most 0.30 with the distance based measures.

Thus, there are two main dimensions in centrality measures for this networks: distance based and degree based. In some way, the degree based measures are local measures while, at the other opposite, the betweenness centrality is a measure that is related to the whole network, regardless the distances (the centrality of

Table 4: Correlation matrix for the main centrality measures

	ecc	c	h	e	ed^2	d	rw	b
ecc	**1.00**	**0.91**	**0.90**	0.41	0.41	0.31	0.31	*0.22*
c	**0.91**	**1.00**	**1.00**	0.49	0.49	0.36	0.36	0.29
h	**0.90**	**1.00**	**1.00**	0.50	0.51	0.36	0.36	0.30
e	0.41	0.49	0.50	**1.00**	**0.98**	0.74	0.74	0.58
ed^2	0.41	0.49	0.51	**0.98**	**1.00**	0.68	0.68	0.58
d	0.31	0.36	0.36	0.74	0.68	**1.00**	**1.00**	0.54
rw	0.31	0.36	0.36	0.74	0.68	**1.00**	**1.00**	0.54
b	*0.22*	0.29	0.30	0.58	0.58	0.54	0.54	**1.00**

a vertex is increased when it lies on the shortest path between other vertices, regardless the distance that separates them).

5.2 Correlations among extensions and new measures

Concerning the newly proposed indices, we first notice that the de_i^α centrality is correlated to the harmonic centrality (and other distance based measures) when α is small, but tends to correlate to the eigenvector centrality as α increases, which is not surprising given the nature of the matrice on which it is based. This property makes this measure suitable to represent centralities in context's where a combination of distance and eigenvector centralities may be suitable.

The geodesic centrality seems interesting as well because it seems to correlate rather well with the betweenness, which is not surprising as the graph is very sparse, but its correlation with eigenvector tend to increase with α.

Overall, increasing the value α modifies a measure which becomes more local, α is thus a parameter which could be adjusted depending on the context.

6 Comparing centrality for professionals and other individuals

A second step in the analysis of the network was, of course to see the importance of professionals in the network. This analysis aims at testing for which of the centrality measures the hypothesis that the centrality is larger for professionals than otehr individuals. The underlying goal is to understand which centrality measures could predict the influence of some individuals in the network. In Table 5, various ranking measures are computed according to all the centrality measures. Table 6 indicates the proportion of each category (professionals and non professionals) that appear in the top decile.

p-values of Wilcoxon two-sample rank-sum tests. For all centrality measures, all tests suggest that the distributions of centrality scores in the two groups differ significantly, at the 0.001 significance level, or better, with higher centrality scores observed for the "professional" group.

p-values of χ^2 test for differences in proportions. For all centrality measures, all tests suggest that the proportion of "professional" that fall into the top-decile for a given measure, based on the full population, is greater than the proportion of "non-professionals". Tied values in centrality measures are assigned the smallest of the corresponding decile-rank (conservative, results are robust to alternative ranking of tied values), which is important in the case oh highly degenerate measures such as *ecc*.

Both tests lead to the exact same conclusion for all the centrality measures. It is clear that the influence of the professionals is higher than the other individuals. From a practical point of view, this result is very strong and it was not expected that such a conclusion arises for every measure.

7 Conclusion

In order to better fit potential analysis needs, we proposed new centrality measures, and a parametrization scheme that allows an easy definition of a measure in a context where the needed is neither local nor global, but a mix of both. It turns out that those properties are respected according to the obtained results. The geodesic and split measures are specially developed to identify individuals that are making links between different communities. These measures are important in various cases such as social network analysis, epidemic analysis or viral marketing (whose model is based upon an epidemic scheme).

From this case analysis point of view, the hypothesis that professionals are more central in the network than other individuals seems to be validated as all the centrality measures clearly leads to that conclusion.

Table 5: Centrality measures by individual category (all differences are significant with $p < 0.001$ according to the Wilcoxon rank test)

VARIABLE Measure	NONPROF 1st decile (1)	NONPROF Median (2)	NONPROF 10th decile (3)	PROF 1st decile (4)	PROF Median (5)	PROF 10th decile (6)
c	1.99588e-07	2.42743e-07	2.85006e-07	2.15155e-07	2.64039e-07	2.97764e-07
ecc	0.0277778	0.03125	0.0333333	0.0294118	0.03125	0.0333333
h	14162.6	17414.45	20716.5	15327	19071.3	21702.4
d	1	1	2	1	1	3
ed^2	7	13	27	8	16	42
e	0.000610326	0.00122065	0.00274647	0.000712047	0.00162754	0.00457745
rw	2.97751e-06	2.97752e-06	5.95503e-06	2.97751e-06	2.97752e-06	8.93255e-06
b^0	263361	263361	3950320	263361	263361	10248300
b^1	14264.2	17737.4	247077	15980.2	20097.6	637873
b^2	794.693	1243.85	15620.4	1005.44	1606.22	42519.7
b^3	47.4752	93.0809	1047.42	68.17	135.348	3016.25
b^4	4.32017	8.78182	77.6986	6.10781	13.9973	234.365
b^5	1.44044	1.98309	8.82349	1.60561	2.82901	24.5162
g^0	263361	263361	9036070000	263361	263361	26532800000
g^1	15127.4	19152.3	50657100	17058.1	22068.7	173094000
g^2	903.638	1468.5	308877	1162.43	1983.54	1367380
g^3	65.2697	129.949	3528.15	91.3647	205.862	17210.1
g^4	12.0569	23.2282	205.927	15.1006	38.5672	666.626
g^5	6.96247	13.9438	68.9221	7.93134	20.9544	184.21
s^0	0	0	357719000	0	0	4616250000
s^1	0	0	2976740	0	0	34720900
s^2	0	0	44591	0	0	367474
s^3	0	0	1762.08	0	0	8523.78
s^4	0	0	227.695	0	0	833.803
s^5	0	0	119.422	0	0	450.662
de^1	0.00154633	0.00190924	0.00228223	0.00167644	0.00209702	0.00239425
de^2	0.00113109	0.001762615	0.00258978	0.00134039	0.00216294	0.0028875
de^3	0.000723773	0.0014695	0.00278781	0.000949776	0.00207881	0.00340228
de^4	0.000296133	0.000849259	0.00243735	0.000449687	0.0015218	0.00348631
de^5	4.43504e-05	0.000193205	0.00122915	7.34299e-05	0.000513742	0.00253666

Table 6: Proportion of individuals of specific type in full population top decile (all differences are significant with $p < 0.001$ according to the χ^2 test, except ecc for which $p = 0.001$)

Measure	non prof.	professionals
c	0.1	0.212
ecc	0.045	0.077
h	0.1	0.214
d	0.053	0.141
ed^2	0.098	0.199
e	0.099	0.214
rw	0.07	0.17
b^0	0.1	0.203
b^1	0.1	0.216
b^2	0.1	0.216

Measure	non prof.	professionals
b^3	0.1	0.22
b^4	0.1	0.23
b^5	0.1	0.24
g^0	0.1	0.24
g^1	0.1	0.25
g^2	0.1	0.243
g^3	0.1	0.252
g^4	0.1	0.243
g^5	0.1	0.243
s^0	0.1	0.232

Measure	non prof.	professionals
s^1	0.1	0.234
s^2	0.1	0.232
s^3	0.1	0.245
s^4	0.1	0.234
s^5	0.1	0.221
de^1	0.1	0.214
de^2	0.1	0.212
de^3	0.1	0.212
de^4	0.1	0.227
de^5	0.1	0.225

However, a much deeper analysis, including qualitative analysis, of the results obtained by each measure will be needed in order to verify whether the geodesic and split measures really indicates the inter-community properties as they are supposed to predict.

Appendix

In this appendix are given the correlation tables for all the centrality measures used in this article for the bipartite and the non bipartite models.

Table 7: Complete correlation table for the considered centrality measures on the bipartite network model

	d	ecc	c	h	e	ed^2	rw	b^0	b^1	b^2	b^3	b^4	b^5	de^1	de^2	de^3	de^4	de^5	g^0	g^1	g^2	g^3	g^4	g^5	s^0	s^1	s^2	s^3	s^4	s^5
d	1	0.31	0.36	0.36	0.74	0.68	1	0.54	0.54	0.53	0.52	0.52	0.54	0.36	0.4	0.44	0.48	0.38	0.82	0.81	0.76	0.7	0.69	0.75	0.74	0.72	0.67	0.64	0.62	0.62
ecc	0.31	1	0.91	0.9	0.41	0.41	0.31	0.22	0.23	0.23	0.24	0.24	0.25	0.9	0.88	0.82	0.65	0.4	0.32	0.35	0.34	0.32	0.3	0.29	0.26	0.27	0.27	0.26	0.24	0.23
c	0.36	0.91	1	1	0.49	0.49	0.36	0.29	0.3	0.31	0.31	0.31	0.32	1	0.99	0.94	0.77	0.49	0.37	0.41	0.42	0.4	0.38	0.36	0.3	0.33	0.33	0.32	0.31	0.29
h	0.36	0.9	1	1	0.5	0.51	0.36	0.3	0.31	0.32	0.32	0.32	0.33	1	0.99	0.95	0.79	0.51	0.38	0.42	0.43	0.41	0.39	0.37	0.31	0.33	0.34	0.33	0.32	0.3
e	0.74	0.41	0.49	0.5	1	0.98	0.74	0.58	0.58	0.58	0.58	0.59	0.61	0.5	0.56	0.63	0.73	0.71	0.67	0.72	0.73	0.73	0.75	0.8	0.61	0.64	0.65	0.67	0.69	0.71
ed^2	0.68	0.41	0.49	0.51	0.98	1	0.68	0.58	0.59	0.59	0.59	0.59	0.62	0.51	0.56	0.63	0.74	0.71	0.64	0.69	0.71	0.71	0.73	0.77	0.59	0.62	0.63	0.66	0.68	0.7
rw	1	0.31	0.36	0.36	0.74	0.68	1	0.54	0.54	0.53	0.52	0.52	0.54	0.36	0.4	0.44	0.48	0.38	0.82	0.81	0.76	0.7	0.69	0.75	0.74	0.72	0.67	0.64	0.62	0.62
b^0	0.54	0.22	0.29	0.3	0.58	0.58	0.54	1	1	0.99	0.99	0.98	0.97	0.31	0.37	0.46	0.61	0.58	0.47	0.58	0.69	0.78	0.83	0.78	0.53	0.62	0.72	0.79	0.79	0.75
b^1	0.54	0.23	0.3	0.31	0.58	0.59	0.54	1	1	1	1	0.99	0.98	0.31	0.37	0.47	0.62	0.6	0.47	0.58	0.7	0.8	0.84	0.79	0.53	0.63	0.73	0.8	0.8	0.76
b^2	0.53	0.23	0.31	0.32	0.58	0.59	0.53	0.99	1	1	1	1	0.99	0.32	0.38	0.47	0.63	0.61	0.47	0.59	0.7	0.81	0.85	0.8	0.53	0.64	0.73	0.82	0.81	0.77
b^3	0.52	0.24	0.31	0.32	0.58	0.59	0.52	0.99	1	1	1	1	1	0.32	0.38	0.48	0.64	0.62	0.47	0.59	0.71	0.82	0.86	0.8	0.53	0.64	0.74	0.83	0.82	0.78
b^4	0.52	0.24	0.31	0.32	0.59	0.59	0.52	0.98	0.99	1	1	1	1	0.33	0.39	0.49	0.65	0.63	0.47	0.6	0.72	0.83	0.87	0.82	0.54	0.65	0.75	0.84	0.84	0.79
b^5	0.54	0.25	0.32	0.33	0.61	0.62	0.54	0.97	0.98	0.99	1	1	1	0.34	0.4	0.5	0.67	0.65	0.49	0.62	0.74	0.85	0.89	0.84	0.56	0.67	0.77	0.86	0.86	0.82
de^1	0.36	0.9	1	1	0.5	0.51	0.36	0.31	0.31	0.32	0.32	0.33	0.34	1	0.99	0.95	0.79	0.51	0.38	0.42	0.43	0.41	0.39	0.37	0.31	0.34	0.34	0.33	0.32	0.3
de^2	0.4	0.88	0.99	0.99	0.56	0.56	0.4	0.37	0.37	0.38	0.38	0.39	0.4	0.99	1	0.98	0.86	0.59	0.41	0.47	0.49	0.48	0.46	0.43	0.34	0.38	0.39	0.39	0.38	0.36
de^3	0.44	0.82	0.94	0.95	0.63	0.63	0.44	0.46	0.47	0.47	0.48	0.49	0.5	0.95	0.98	1	0.93	0.7	0.45	0.52	0.57	0.57	0.55	0.52	0.39	0.44	0.47	0.48	0.47	0.44
de^4	0.48	0.65	0.77	0.79	0.73	0.74	0.48	0.61	0.62	0.63	0.64	0.65	0.67	0.79	0.86	0.93	1	0.89	0.48	0.59	0.67	0.72	0.72	0.67	0.45	0.53	0.59	0.64	0.64	0.61
de^5	0.38	0.4	0.49	0.51	0.71	0.71	0.38	0.58	0.6	0.61	0.62	0.63	0.65	0.51	0.59	0.7	0.89	1	0.38	0.49	0.59	0.67	0.7	0.67	0.37	0.46	0.54	0.62	0.65	0.64
g^0	0.82	0.32	0.37	0.38	0.67	0.64	0.82	0.47	0.47	0.47	0.47	0.47	0.49	0.38	0.41	0.45	0.48	0.38	1	0.97	0.88	0.76	0.67	0.68	0.9	0.85	0.76	0.66	0.59	0.58
g^1	0.81	0.35	0.41	0.42	0.72	0.69	0.81	0.58	0.58	0.59	0.59	0.6	0.62	0.42	0.47	0.52	0.59	0.49	0.97	1	0.97	0.88	0.8	0.79	0.91	0.92	0.87	0.79	0.72	0.7
g^2	0.76	0.34	0.42	0.43	0.73	0.71	0.76	0.69	0.7	0.7	0.71	0.72	0.74	0.43	0.49	0.57	0.67	0.59	0.88	0.97	1	0.97	0.9	0.87	0.88	0.93	0.94	0.89	0.83	0.79
g^3	0.7	0.32	0.4	0.41	0.73	0.71	0.7	0.78	0.8	0.81	0.82	0.83	0.85	0.41	0.48	0.57	0.72	0.67	0.76	0.88	0.97	1	0.97	0.93	0.8	0.89	0.95	0.96	0.92	0.88
g^4	0.69	0.3	0.38	0.39	0.75	0.73	0.69	0.83	0.84	0.85	0.86	0.87	0.89	0.39	0.46	0.55	0.72	0.7	0.67	0.8	0.9	0.97	1	0.98	0.73	0.84	0.92	0.98	0.98	0.95
g^5	0.75	0.29	0.36	0.37	0.8	0.77	0.75	0.78	0.79	0.8	0.8	0.82	0.84	0.37	0.43	0.52	0.67	0.67	0.68	0.79	0.87	0.93	0.98	1	0.74	0.83	0.89	0.95	0.97	0.97
s^0	0.74	0.26	0.3	0.31	0.61	0.59	0.74	0.53	0.53	0.53	0.53	0.54	0.56	0.31	0.34	0.39	0.45	0.37	0.9	0.91	0.88	0.8	0.73	0.74	1	0.97	0.89	0.78	0.69	0.67
s^1	0.72	0.27	0.33	0.33	0.64	0.62	0.72	0.62	0.63	0.64	0.64	0.65	0.67	0.34	0.38	0.44	0.53	0.46	0.85	0.92	0.93	0.89	0.84	0.83	0.97	1	0.97	0.89	0.8	0.77
s^2	0.67	0.27	0.33	0.34	0.65	0.63	0.67	0.72	0.73	0.73	0.74	0.75	0.77	0.34	0.39	0.47	0.59	0.54	0.76	0.87	0.94	0.95	0.92	0.89	0.89	0.97	1	0.97	0.9	0.86
s^3	0.64	0.26	0.32	0.33	0.67	0.66	0.64	0.79	0.8	0.82	0.83	0.84	0.86	0.33	0.39	0.48	0.64	0.62	0.66	0.79	0.89	0.96	0.98	0.95	0.78	0.89	0.97	1	0.97	0.93
s^4	0.62	0.24	0.31	0.32	0.69	0.68	0.62	0.79	0.8	0.81	0.82	0.84	0.86	0.32	0.38	0.47	0.64	0.65	0.59	0.72	0.83	0.92	0.98	0.97	0.69	0.8	0.9	0.97	1	0.99
s^5	0.62	0.23	0.29	0.3	0.71	0.7	0.62	0.75	0.76	0.77	0.78	0.79	0.82	0.3	0.36	0.44	0.61	0.64	0.58	0.7	0.79	0.88	0.95	0.97	0.67	0.77	0.86	0.93	0.99	1

Table 8: Complete correlation table for the considered centrality measures on the non bipartite network model

	d	ecc	c	h	e	ed^2	rw	b^0	b^1	b^2	b^3	b^4	b^5	de^1	de^2	de^3	de^4	de^5	g^0	g^1	g^2	g^3	g^4	g^5
d	1	0.41	0.49	0.5	0.88	0.78	1	0.58	0.58	0.58	0.58	0.59	0.61	0.5	0.56	0.63	0.71	0.68	0.7	0.71	0.69	0.66	0.63	0.61
ecc	0.41	1	0.91	0.9	0.4	0.51	0.41	0.23	0.23	0.23	0.24	0.24	0.25	0.9	0.88	0.82	0.7	0.43	0.33	0.33	0.31	0.28	0.25	0.22
c	0.49	0.91	1	1	0.49	0.61	0.49	0.3	0.3	0.31	0.31	0.32	0.32	1	0.99	0.94	0.82	0.53	0.4	0.41	0.39	0.35	0.32	0.29
h	0.5	0.9	1	1	0.5	0.63	0.5	0.3	0.31	0.32	0.32	0.33	0.33	1	0.99	0.95	0.84	0.54	0.41	0.42	0.4	0.37	0.33	0.3
e	0.88	0.4	0.49	0.5	1	0.87	0.88	0.52	0.53	0.54	0.54	0.56	0.58	0.51	0.57	0.66	0.8	0.9	0.55	0.6	0.63	0.63	0.63	0.62
ed^2	0.78	0.51	0.61	0.63	0.87	1	0.78	0.67	0.68	0.7	0.7	0.72	0.73	0.63	0.71	0.81	0.92	0.89	0.65	0.72	0.77	0.78	0.76	0.74
rw	1	0.41	0.49	0.5	0.88	0.78	1	0.58	0.58	0.58	0.58	0.59	0.61	0.5	0.56	0.63	0.71	0.68	0.7	0.71	0.69	0.66	0.63	0.61
b^0	0.58	0.23	0.3	0.3	0.52	0.67	0.58	1	1	0.99	0.99	0.98	0.97	0.31	0.36	0.44	0.55	0.49	0.63	0.72	0.79	0.83	0.84	0.81
b^1	0.58	0.23	0.3	0.31	0.53	0.68	0.58	1	1	1	1	0.99	0.98	0.31	0.37	0.45	0.56	0.51	0.64	0.73	0.8	0.85	0.85	0.83
b^2	0.58	0.23	0.31	0.32	0.54	0.7	0.58	0.99	1	1	1	1	0.99	0.32	0.38	0.46	0.57	0.52	0.64	0.73	0.81	0.86	0.87	0.85
b^3	0.58	0.24	0.31	0.32	0.54	0.7	0.58	0.99	1	1	1	1	0.99	0.32	0.38	0.47	0.58	0.53	0.64	0.74	0.82	0.87	0.88	0.86
b^4	0.59	0.24	0.32	0.33	0.56	0.72	0.59	0.98	0.99	1	1	1	1	0.33	0.39	0.48	0.59	0.55	0.65	0.75	0.83	0.88	0.9	0.88
b^5	0.61	0.25	0.32	0.33	0.58	0.73	0.61	0.97	0.98	0.99	0.99	1	1	0.34	0.4	0.49	0.61	0.57	0.66	0.76	0.84	0.9	0.92	0.9
de^1	0.5	0.9	1	1	0.51	0.63	0.5	0.31	0.31	0.32	0.32	0.33	0.34	1	0.99	0.95	0.84	0.55	0.41	0.42	0.4	0.37	0.33	0.31
de^2	0.56	0.88	0.99	0.99	0.57	0.71	0.56	0.36	0.37	0.38	0.38	0.39	0.4	0.99	1	0.98	0.9	0.62	0.46	0.47	0.47	0.44	0.4	0.37
de^3	0.63	0.82	0.94	0.95	0.66	0.81	0.63	0.44	0.45	0.46	0.47	0.48	0.49	0.95	0.98	1	0.96	0.72	0.51	0.55	0.55	0.53	0.49	0.46
de^4	0.71	0.7	0.82	0.84	0.8	0.92	0.71	0.55	0.56	0.57	0.58	0.59	0.61	0.84	0.9	0.96	1	0.86	0.57	0.63	0.65	0.65	0.63	0.6
de^5	0.68	0.43	0.53	0.54	0.9	0.89	0.68	0.49	0.51	0.52	0.53	0.55	0.57	0.55	0.62	0.72	0.86	1	0.47	0.54	0.59	0.62	0.63	0.62
g^0	0.7	0.33	0.4	0.41	0.55	0.65	0.7	0.63	0.64	0.64	0.64	0.65	0.66	0.41	0.46	0.51	0.57	0.47	1	0.98	0.91	0.82	0.73	0.68
g^1	0.71	0.33	0.41	0.42	0.6	0.72	0.71	0.72	0.73	0.73	0.74	0.75	0.76	0.42	0.47	0.55	0.63	0.54	0.98	1	0.98	0.92	0.85	0.8
g^2	0.69	0.31	0.39	0.4	0.63	0.77	0.69	0.79	0.8	0.81	0.82	0.83	0.84	0.4	0.47	0.55	0.65	0.59	0.91	0.98	1	0.98	0.93	0.89
g^3	0.66	0.28	0.35	0.37	0.63	0.78	0.66	0.83	0.85	0.86	0.87	0.88	0.9	0.37	0.44	0.53	0.65	0.62	0.82	0.92	0.98	1	0.98	0.95
g^4	0.63	0.25	0.32	0.33	0.63	0.76	0.63	0.84	0.85	0.87	0.88	0.9	0.92	0.33	0.4	0.49	0.63	0.63	0.73	0.85	0.93	0.98	1	0.99
g^5	0.61	0.22	0.29	0.3	0.62	0.74	0.61	0.81	0.83	0.85	0.86	0.88	0.9	0.31	0.37	0.46	0.6	0.62	0.68	0.8	0.89	0.95	0.99	1

References

[1] Bay Street Group LLP (2010) New Survey Results: Accounting Firms Dive into Digital Marketing. Accessed October 4th, 2010, available at: http://cpatrendlines.com/2010/07/21/new-survey-results-accounting-firms-dive-into-digital-marketing/.

[2] Caporossi, G., Paiva, M., Vukicević, D. & Segatto, M. (2012) Centrality and Betweenness: Vertex and Edge Decomposition of the Wiener Index. *MATCH Commun. Math. Comput. Chem.*, **68**, 293–302.

[3] Freeman, L.C. (1977) A Set of Measures of Centrality Based on Betweenness. *Sociometry*, **40**, 35–41.

[4] Freeman, L.C. (1978) Centrality in Social Networks Conceptual Clarification. *Social Networks*, **1**, 215–239.

[5] Gauthier, R. (2008) Don, bénévolat et participation : quel est le comportement des Québécois? Centre d'études sur l'action communautaire et bénévole, 70 pages. Available from: http://www.rabq.ca/admin/incoming/20120910100846_554_ECDBP-ComportementQubec-final.pdf.

[6] Hackl, F., Halla, M. & Pruckner, G.J. (2007) Volunteering and Income - The Fallacy of the Good Samaritan? *KYKLOS*, **60**(1), 77-104.

[7] Hitt, M., Bierman, L., Shimizu, K. & Kochhar, R. (2001) Direct and Moderating Effects of Human Capital on Strategy and Performance in Professional Service Firms: A Resource-Based Perspective. *The Academy of Management Journal*, **44**(1), 13–28.

[8] Hooks, K.L., Cheramy, S.J. & Sindch, T.L. (1994) Methods Used by Big 6 Partners in Practice Development. *Auditing: A Journal of Practice & Theory*, **13**(1), 101–114.

[9] Imagine Canada (2006) Board Volunteers in Canada: Their Motivations and Challenges, Knowledge Development Centre and Imagine Canada Research Report, 26 p. Available from: http://www.imaginecanada.ca/files/www/en/library/kdc-cdc/imagine_board_volunteers_report.pdf.

[10] Newman, M.E.J. (2005) A measure of betweenness centrality based on random walks *Social Networks*, **27**(1), 39–54.