Les Cahiers du GERAD ISSN: 0711–2440

Exploring an influence network through centrality measures

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

G–2015–100

September 2015

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 an influence network through centrality measures

Gilles Caporossi [a]

Sylvain Perron [a]

Louis-Philippe Sirois [b]

Johanne Turbide [c]

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

[b] *Université Laval, Québec (Québec) Canada, G1V 0A6*

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

gilles.caporossi@gerad.ca
sylvain.perron@hec.ca
louis-philippe.sirois@fsa.ulaval.ca
johanne.turbide@hec.ca

September 2015

Les Cahiers du GERAD

G–2015–100

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 organizations, including charities in Québec. The main contribution of this study is to show, with a dataset containing all charity organization, the structure of this influence network. We also show that professionals (accountants and lawyers in this case) have a more strategical position within the network.

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 analyzed using a wide variety of centrality measures. Indeed, the classical measures were developed 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 comparison of the classical, new and adjusted measures will be presented. This research lead to some unexpected results strengthening the conclusion of the analysis.

1 The context, data description

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 reputation capital, is especially valuable for business professionals, such as accountants and lawyers, both for their career and the organizations they work for. Indeed, the quality of the services these professionals provide is based on reputation 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 organization 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 organizations.[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" organizations. 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.

We limit our analysis to organizations registered in the Canadian province of Québec given the availability of data. All active organizations 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. Organizations are grouped into three categories: 1) charities and nonprofit organizations,[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 organizations are identified from various sources.[7] The objective is to capture a significant portion of the Québec business sector.

In total, the base data for our network consists of 54,485 organizations and 418,580 board members, with a vast majority of 52,666 as nonprofit organizations (402,417 members), 1,129 business associations (11,168 members) and 1,282 "for-profit" organizations (10,410 members). The size distribution of management boards

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

[2] See for example: www.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 organizations 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 organizations. This represents the majority of nonprofit organizations in Québec. For other nonprofit organizations 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 organizations registered with the Agency and that can issue tax receipts for donations. This is the case, for example, of faith-based organizations. Finally, we exclude student associations incorporated under *Part III of the Companies Act*.

[6] "Business associations" also operate as nonprofit organizations 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 organizations 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&).

for all organizations in the sample is presented on Table 1 (the size distribution is similar across all three organization types).

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

Total number of organizations	54485
Total number of board members	418580
Distribution of board members per organizations	
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 organizations and individuals together, are identified by matching individual entries per organizations to entries of all other organizations in the database based on standardized 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 organizations 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.

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		

[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.

2 Network modeling

From a technical point of view, even if there are different ways to model the data by a network, it turns out that the network has a large number of vertices. A classical approach when dealing with a huge dataset is to sample the data, and valid the results obtained on the sample afterward. Unfortunately, it is impossible to keep the essence of the network after sampling some of its vertices. It is also impossible to understand the processes involved in the network by a classical exploration (for example, by drawing the network or looking at its links and vertices).

To understand the nature of the network from the influence point of view, the only possible way is thus to evaluate various measures on the network and confront them in order to deduce the main characteristics of that network.

However, some of these measures cannot be computed for algorithmic reasons, and we have to concentrate only on those that are technically possible.

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.

For the current study, we decided to use the bipartite formulation. If such a model yields more vertices and may suffer from the interactions between the two kind of vertices, the modeling of the interactions is more precise as multiple pairs of actors will be linked according to the different organizations to which they are associated. Another approach consists in weighting the link between two individuals to express the strength of the relation, however, it misses some information. For example, two individuals sharing more than one board will appear with a corresponding weight while their relation to other people will be lost.

3 Network description

The number of connected components in the network is 18 700. One of these components is being very large while the others are all very small. If the main component involves 230 765 individuals, the second only has less than 164 vertices. It is clear that the small components are related to organizations that are rather small and cannot be considered as influence chanels. Furthermore, if the boards of organizations are used for influence reasons, this can only occurs in the main component.

The situation is quite different for the main component which involves 32 597 organizations through which the 230 765 individuals are related. The relative size of the various components is clearly not due to chance. This information seems to indicate that the motivation for belonging to a board is not only the interest in the management of the organization, but likely to be a part of the decision sphere of the society.

In a bipartite model, the graph thus has 263 362 vertices and 294 276 edges, this graph is almost as sparse as a tree. 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. As the number of edges related to a given organization is the number of its members in the case of the bipartite model, and a quadratic function of this number in the other model, the influence of large organizations is clear.

The description of that network could mainly be achieved through the description of its individuals or organizations. For this reason the following centrality measures were computed and analyzed.

- **Degree centrality.** The degree d_i of the vertex i is the number of edges adjacent to the vertex i. In this graph, the degree values are between 1 and 15 for the individuals and between 2 and 66 for the organizations. In the case of individuals, the degree distribution is similar to that of most complex networks (fat tail distribution). Indeed, more than 81% of the individuals belong to only one organization, 13% to 2, 3.3% to 3, 1.1% to 4, while the sum of all the others represents less than 1% altogether. In the case of the organizations, it is slightly different because of the nature of a board. No board has only one member, very few (1.1%) has 2, 10.2% has 3. From 4 to 10, the proportion is somehow stable, varying from 7 to 10%, and this proportion decreases slowly as shows Figure 1.

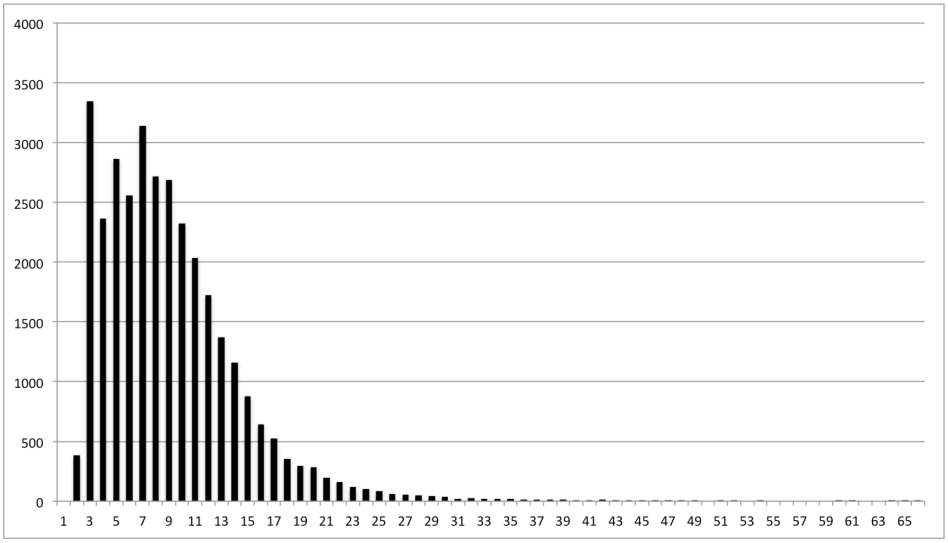

Figure 1: Number of organizations as a function of the number of members (degree distribution of the organizations)

- **Eccentricity centrality.** The eccentricity centrality is computed from the maximum distance from the vertex i as follows :

$$ecc_i = \max_j d_{ij}, \tag{1}$$

where d_{ij} denotes the geodesic distance between the vertices i and j. Since ecc_i is smaller for more central vertices, we will refer to Eccentricity centrality C_i^{ecc} the inverse of the eccentricity.

$$C_i^{ecc} = \frac{1}{ecc_i} = \frac{1}{\max_j d_{ij}}. \tag{2}$$

The minimum eccentricity (also called radius) is 26 while the maximum (the diameter) is 46. These rather large values are not surprising given the sparsity of the graph and the moderate maximum degree, specially because the number of vertices of higher degree is quite limited. The large difference between radius and diameter altogether with the very small density indicates (at least for a large portion of the graph) that the graph has roughly a tree structure. For trees, the radius is about the half of the diameter (depending if this last one is odd or even). On the opposite, a graph for which the radius is close to the diameter is a graph where all vertices have almost the same eccentricity. In such a case, there is no central vertex (in the sense of eccentricity), it is the case of the cycle or the complete graph in which all vertices play the same role.

- **Closeness centrality.** The closeness centrality c_i is based upon the sum of the distances from i to all the other vertices, also called transmission t_i as follows:

$$c_i = \frac{1}{t_i} = \frac{1}{\sum_{j=1}^{n} d_{ij}}. \tag{3}$$

- **Harmonic centrality.** In its essence, the harmonic centrality is closely related to the closeness 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.

3.1 Comparing centrality measures to characterize the network

The main centrality measures are compared through their correlations which are given on Table 3.

Table 3: Correlation matrix for the main centrality measures

	ecc	*c*	*h*	*d*	*b*
ecc	**1.00**	**0.91**	**0.90**	0.31	*0.22*
c	**0.91**	**1.00**	**1.00**	0.36	0.29
h	**0.90**	**1.00**	**1.00**	0.36	0.30
d	0.31	0.36	0.36	**1.00**	0.54
b	*0.22*	0.29	0.30	0.54	**1.00**

Except the correlation between eccentricity and betweenness which is 0.22 (in italics), all other 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 degree and (iii) the betweenness centrality.

Should the betweenness centrality be associated to one of the other 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 a vertex is increased when it lies on the shortest path between other vertices, regardless the distance that separates them).

3.2 Comparing centrality among individuals and among organizations

Table 4: Correlation matrix for the main centrality measures among individuals

	ecc	*c*	*h*	*d*	*b*
ecc	**1.00**	**0.91**	**0.91**	0.26	*0.20*
c	**0.91**	**1.00**	**1.00**	0.30	0.25
h	**0.91**	**1.00**	**1.00**	0.31	0.26
d	0.26	0.30	0.31	**1.00**	0.57
b	*0.20*	0.25	0.26	0.57	**1.00**

Table 5: Correlation matrix for the main centrality measures for organizations

	ecc	*c*	*h*	*d*	*b*
ecc	**1.00**	**0.91**	**0.91**	0.24	*0.21*
c	**0.91**	**1.00**	**1.00**	0.28	0.27
h	**0.91**	**1.00**	**1.00**	0.29	0.28
d	0.24	0.38	0.29	**1.00**	0.33
b	*0.21*	0.27	0.28	0.33	**1.00**

4 Comparing centrality for professionals and other individuals

If the board of organizations is used for influence reasons, it is to be expected that professionals (lawyers and accounters) 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 other individuals. The underlying goal is to understand which centrality measures could predict the influence of some individuals in the network. In Table 6, various ranking measures are computed according to all the centrality measures. Table 7 indicates the proportion of each category (professionals and non professionals) that appear in the top decile.

Table 6: 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
b	263361	263361	3950320	263361	263361	10248300

Table 7: 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
b	0.1	0.203

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.

5 Conclusion

The main contribution of this work is the construction of the organization/individual network. The exhaustive aspect of the network with respect to charity organizations makes this network very interesting to study. The first remark is that the size of connected components are very small, except one. This indicates that in a large number of organizations, the board is not related to any other organization, which implies that the individuals in those boards are interested in their tasks mainly for the benefit of the organization. However, the important size of the main component tends to indicate that boards may also be used for networking reasons. Analyzing the various centrality measures, both local and global is also very instructive. The correlation between all measures shows that the individuals that are in the center of the graph (with respect to the geodesic distance) are also those with largest local centrality. In other words, a more central individual will belong to more boards. Interestingly, the same analysis holds for organizations. As the degree of a node representing an organization is the number of its board members, and as that number is related to the importance of the organization, it is clear that important organizations are in the center of the network. If this phenomena is not unknown, it is demonstrated here clearly with a exhaustive dataset. The next and very interesting result is that Lawyers and Accountants tend to be more central than the average. This again is not surprising when we know that some firms clearly express their interests in networking, but it is clearly stated by the results of our study.

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 Joumal 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.