The Evolving Social Network of Marketing Scholars

Jacob Goldenberg, Barak Libai, Eitan Muller and Stefan Stremersch

Database Submission to *Marketing Science*

September 2009

Jacob Goldenberg is Professor of Marketing at the School of Business Administration, Hebrew University of Jerusalem, Jerusalem, Israel. E-mail address: msgolden@mscc.huji.ac.il.

Barak Libai is Associate Professor of Marketing at the Leon Recanati Graduate School of Business, Tel Aviv University, Tel Aviv, Israel.

Eitan Muller is Chaired Professor of Marketing at the Leon Recanati Graduate School of Business, Tel Aviv University, Tel Aviv, Israel, and Professor of Marketing at the Stern School of Business, New York University.

Stefan Stremersch is Chaired Professor of Marketing and Desiderius Erasmus Distinguished Chair of Economics at the Erasmus School of Economics, Erasmus University Rotterdam, and Visiting Professor of Marketing, IESE, Universidad de Navarra, Barcelona, Spain.

The authors acknowledge the research assistance of Karni Marcus, Boaz Chen, Albert Kimyagarov, Arik Veeder and Laura Wijnhold. They especially thank Arik Veeder of Netwise for invaluable conceptual and computational assistance. They thank the Erasmus Research Institute for Management and the Israeli National Science foundation for their financial support.

Abstract

The interest in social networks among marketing scholars and practitioners has sharply increased in the last decade. One social network, of which network scholars increasingly recognize the unique value, is the academic collaboration (co-author) network.

We offer a comprehensive database of the collaboration network among marketing scholars over the last 40 years. Based on the ProQuest database, it documents the social collaboration among researchers in dozens of the leading marketing journals, enabling us to create the networks of active marketing researchers. Unlike most of the published academic collaboration research, our database is dynamic and follows the evolution of the field over many years.

In this paper, we describe the database and point to some basic network descriptives that lead to interesting research questions. We believe this database can be of much interest to researchers interested in the evolution of social networks over time, as well as the specific evolution of the marketing discipline.

Introduction

Social networks are increasingly attracting the attention of marketing scholars and practitioners. In the past four years, it has been identified by the members of the *Marketing Science Institute* (MSI) as the highest priority research topic in marketing. Research in this area often examines the way various levels of personal networks aggregate to construct a social system. In order to analyze how networks form a certain social system, one must be able to document social contacts within the system, preferably from its inception. Such data are often hard to construct in large social systems, and so much of social network analysis has been conducted on relatively small networks, and most of them cover relatively short time frames, usually after the network is already formed and became stable.

To examine larger networks, researchers have begun to take advantage of academic publication databases, where the social network of collaboration among scientists is documented in detail. Indeed, recent advances in social network analysis have used the networks of academic collaborations in areas such as: mathematics (Grossman and Ion 1995); biology, physics, computer science (Newman 2001); and neuroscience (Barabasi et al. 2002). An additional advantage of these inquiries is that they allow a better understanding of the mechanisms that drive the development and structure of the specific academic disciplines (Moody 2004). These attempts suffer however from a shortcoming: the documented networks are quick "snapshots" of a network rather than a fully documented evolution. One reason is that the networks used are built using data in established ("old") fields (e.g. Biology, Mathematics): the size and the fact that electronic documentation began only recently made it difficult to construct a dynamic picture of the network from its inception.

To address this issue, we hereby publicize the database composed in the context of the marketing connectivity project (www.mconnectivity.com). This database documents scientific collaboration in marketing, from the time the modern marketing social network has begun to today. Based on the ProQuest database, the database documents the social collaboration among researchers in dozens of the leading marketing journals, enabling us to create the networks of active marketing researchers. Even though, as we will review below, collaboration network analysis has been undertaken in other disciplines, our database is novel for several reasons. First, unlike most of the published academic collaboration research, our analysis is dynamic, and follows the evolution of the field over many years. Therefore, one can make inferences regarding the evolution of a social system from a relatively small number of players to the more than 22,000 connected players at the end of 2008. Second, our data provide more richness in the descriptors of both links between authors and author “demographics”. Third, our database offers the first collaboration network data in the marketing discipline and thus allows novel descriptions of the evolution of marketing science, also as compared to other sciences, which allows rich scientometric analyses of the marketing discipline. Such data were unavailable and are not trivial to replicate (as conceded in the limitations of Stremersch et al. (2007), who used a publicly available snapshot of our data).

We believe that this database offers unique research opportunities. It allows researchers of social networks to examine various questions that so far because of a lack of data, remained unanswered. Befitting the increased attention within marketing for scientometric investigations (Stremersch et al. 2007; Stremersch and Verhoef 2005), present and aspiring marketing scholars can use it to study the growth of our discipline, by either studying the growth of the network itself or relating it to other concepts, such as productivity, diversity, innovativeness or impact.

The Data

This section first presents briefly the network definitions and characteristics, then the data collection procedures of the marketing connectivity project, after which it turns to the database structure and the choices made when constructing the database. Networks consist of nodes (typically people in case of social networks) and links (typically social ties). In scientific collaboration networks, nodes are defined by authors and links are defined by their joint papers. We count each year how many papers each pair of authors published together and the journal they published in. This count defines a "temperature" of the collaboration link and tie strength of each pair is measured by the cumulative number of published papers. By defining a link as co-authorship, it is assumed that this kind of activity consists of interactions, communication and knowledge sharing. The source for the database is the intersection of the ProQuest ABI/INFORM database, a large-scale business literature computerized database, and the Theoharakis and Hirst (2002) list of top 40 marketing-related journals. We considered the 40 top journals out of the total of 55 journals that were the basis of the survey. Out of these 40 journals, five did not appear in ProQuest, or appeared in a way that could not be used, and therefore were deleted from our list. Our final list therefore contains the following 35 journals listed in Table 1 (in parenthesis: date of journal inception; date of first inclusion in ProQuest).

The dynamic nature of this database is evident when one considers the fact that while the current list applies to 35 journals, at the end of the 1960s, only two marketing journals—*Journal of Marketing* and *Journal of Marketing Research*—are recorded. The rapid increase in the number of journals is consistent with the sharply increasing number of scholars with the years. This phenomenon may be one of the appealing aspects of our database, as it allows the study of large dynamic networks. For each published study in the database the following details are retrieved: the authors' particulars, the journal, and the year of publication. Based on this database, and consistent with common practice in network analysis, we constructed a large matrix wherein each researcher is a node. The matrix records which nodes are connected through collaboration on a paper. Thus, the user has access to separate databases for the years 1973-2008.

Table 1: Included journals (year of inception; year of inclusion in ProQuest)

1. California Management Review (1959; 1969)	19. Journal of Marketing (1936; 1939)
2. Decision Science (1970; 1972)	20. Journal of Marketing Research (1964; 1964)
3. European Journal of Marketing (1967; 1975)	21. Journal of Personality and Social Psychology (1965; 1986)
4. Harvard Business Review (1922; 1971)	22. Journal of Personal Selling and Sales Management (1980; 1983)
5. Industrial Marketing Management (1971; 1972)	23. Journal of Product Innovation Management (1984; 1984)
6. International Journal of Advertising (1982; 1982)	24. Journal of Public Policy and Marketing (1982; 1982)
7. International Journal of Research in Marketing (1984; 1985)	25. Journal of Retailing (1925; 1971)
8. International Marketing Review (1984; 1986)	26. Journal of Services Marketing (1986; 1987)
9. Journal of Advertising (1972; 1972)	27. Journal of Service Research (1998; 2000)
10. Journal of Advertising Research (1961; 1971)	28. Journal of the Academy of Marketing Science (1973; 1973)
11. Journal of Business (1928; 1973)	29. Management Science (1955; 1971)
12. Journal of Business Research (1973; 1974)	30. Marketing Letters (1989; 1989)
13. Journal of Consumer Marketing (1983; 1983)	31. Marketing Science (1982; 1984)
14. Journal of Consumer Research (1974; 1974)	32. Organizational Behavior & Human Decision Processes (1966; 1971)
15. Journal of Economic Psychology (1981; 1984)	33. Psychology and Marketing (1984; 1992)
16. Journal of International Business Studies (1970; 1970)	34. Sloan Management Review (1959; 1970)
17. Journal of International Marketing (1993; 2000)	35. Strategic Management Journal (1980; 1980)
18. Journal of Market-Focused Management (1996; 1999)	

One should also note that analyses of and results stemming from any research on this database heavily depend on the journals included in the dataset. For example, authors in less known journals might also be publishing less frequently and with less central co-authors. This will have immediate implications on the network formation and metrics. In order to provide more information about the authors, we constructed two (linked) datasets: Authors Database and Links Database.

Authors Database includes author (with his/her ID); Average separation; Cumulative number of papers; Cumulative number of journals s/he published in; First year of publication and last year of publication.

Links Database includes two authors (with their IDs); Cumulative number of joint papers; List of journals at which they published their joint papers.

As will be explained shortly, these databases are provided on an annual basis that allows for exploring the dynamics of and changes in the network. The representation of the links database is suited for large networks and is compatible with common network software packages such as the widely used UCinet and *Pajek* - a free software package that can be downloaded at http://vlado.fmf.uni-lj.si/pub/networks/pajek/.

We made several choices when constructing the database of which its prospective users should be aware. First, we chose a certain list of journals, which are considered a good representation of the marketing domain (Theoharakis and Hirst 2002), but other scholars may feel journals were omitted. Second, we applied an automatic algorithm to detect cases in which authors' names may be misspelled (a well-known issue in scientometrics), combining it with a manual inspection. In the database each author is assigned a unique id number in addition to the name, to help users with easier identification. However, despite our procedures, we cannot guarantee that some authors may appear twice under small variations.

Third, the articles covered start from 1964. Being a relatively young discipline, only a few marketing journals have a long history. A notable exception is the *Journal of Marketing* that is covered in ProQuest and has been published since the late 1930s. One should note that most journals' coverage starts with the early 1970's, for some due to ProQuest coverage issues. Therefore, the yearly databases we provide start from the early 1970's. At the time, the modern marketing network was still very small, and can still be considered in its infancy stage.

Description of Network over Time and across Disciplines

To trigger future research ideas that other scholars can test using this public database, we provide some basic descriptives of the network of marketing scholars over time and compared with other disciplines. We first derive core network measures by which one can describe the network.

Network Measures

Visually, a network is a graph composed of nodes with two types of links between them: direct and indirect (Monge and Contractor 2003). The total number of links for a certain node is labeled the *degree* of that node. The *shortest distance* (also known as *degree of separation)* between two nodes reflects the number of nodes needed to get from one node to the other via the shortest path (usually termed the *geodesic path*). Therefore, a node's network represents the distribution of the degree of separation of this node from the rest of the network. As a preliminary examination of the data by potential users and as a general coherence demonstration we describe the marketing collaboration network through three basic concepts: the largest component, separation, and the clustering coefficient. Our description relates to the time period up to the end of 2008, and on the main component of the network.

In network databases, few different *components* are frequently identified. A component is a cluster (sub-network) of nodes that are linked between themselves but are not linked outside to other nodes. Scientific social network researchers have traditionally focused on the largest component (usually termed the *main component*), which is typically larger in at least one order of magnitude, as compared to the second largest component. A common measure of interest is the *ratio of the largest component to the entire network,* or, the number of scholars in the largest group of nodes that are connected to each other, to the total number of scholars in the network.

We discern average separation and expected average separation. The relevant measure for average separation l represents the mean shortest separation between the n node pairs in the network ("geodesic separation", see Newman, 2003):

(1) $$l = \frac{1}{\frac{1}{2}n(n-1)} \sum_{i \geq j} d_{ij}$$, where d_{ij} is the shortest separation between nodes i and j.

It is generally proposed that in random networks, *l* scales logarithmically compared to the network size *n*, and the value may be comparable also in social networks which are not random. It is suggested that the *expected average separation* should be about $l \approx \frac{\ln(n)}{\ln(k)}$ where *k* is the average degree (number of co-authors in our case) of the network (Watts 1999).

The third basic concept relates to the phenomenon of *clustering* or *transitivity*, a known element of social networks that distinguishes them from random graphs (Newman 2003). Clustering means that if Anne is directly connected to Bart, and Bart is directly connected to Chak, then there is a heightened probability that Anne is also directly connected to Chak. In a random graph, the clustering coefficient C should be in the range $C \sim \frac{k}{n}$ (Watts 1999). One of the ways to represent clustering on the network level is to measure the network level *cluster coefficient C*, which can be defined as:

$$C = \frac{3 \;\; x \;\; \text{number of triangles in the network}}{\text{number of connected triple nodes}} \qquad (2)$$

Where a connected triple node indicates any node connected to a pair of other nodes. C is the mean probability that two nodes that are in a direct link to a third node are in a direct link themselves. Note that other measures of the cluster coefficient are also possible ((Newman 2003).

Comparing Collaboration Networks across Disciplines

Table 2 compares the collaboration network of marketing scholars at the end of 2008 to the collaboration network of scholars in other disciplines: biomedicine, physics, and mathematics (adapted from Newman 2004). Two points should be considered here. First, for two of the three other disciplines, the period covered is very short (five years), and the sample was constructed at very mature stages of the network. Second, the reference list in Table 1 includes some journals in which inter-disciplinary research is published (e.g., *Journal of Business Research*). Some of the researchers in these outlets are non-marketing academics, and therefore not connected to the marketing main component. In our comparison, we, therefore, concentrate on the main component of each network. Table 2 shows that the marketing discipline is much smaller than the other three, with a smaller, average, number of co-authors. Out of the 32,381 researchers in the 2008 social network, the largest component includes 22,278 researchers. The group that was left out, comprising of 10,103 researchers is highly fragmented. The largest component in this group includes just 32 individuals. The ratio of the largest component to the entire network in marketing (69%) is lower compared to the other disciplines, possibly reflecting the existence of more "out-of-the consensus" researchers.

Table 2: Core network concepts: Marketing vs. three other disciplines

	Biomedicine	Physics	Mathematics	Marketing
Years covered	1995-1999	1995-1999	1940-2000	1964-2008
Average number of coauthors	18.1	9.7	3.9	4.3
No. of authors in largest component	1,520,251	52,909	253,339	22,278
Ratio of largest component to complete network	92%	85%	82%	69%
Average separation	4.6	5.9	7.6	7.5
Expected average separation	4.9	6.1	9.1	6.9
Clustering coefficient	6.6%	43%	15%	50%

Similar to Mathematics, average separation in the marketing discipline (7.5) is considerably higher than biomedicine and physics. Yet for all networks, the average separation is very low compared to network size. How much can the average separation say about the level of connectivity in the network? Following the above, one should take into account the network size and the average degree in the network as well. For example, the biomedicine network size is very large, over 1.5 million, yet the average number of co-authors is considerably larger than that of the other disciplines. Comparing the expected average separation with the average separation shows that in biomedicine and physics the two are not that far apart. Considering the short time span (which implies closer relations), these disciplines' shorter separation compared to marketing may not say much regarding connectivity. However, also in mathematics, where the time span is large, the average separation is lower compared to the expected one. Collaboration among mathematicians, it seems, creates a more coherent network compared to that among marketers.

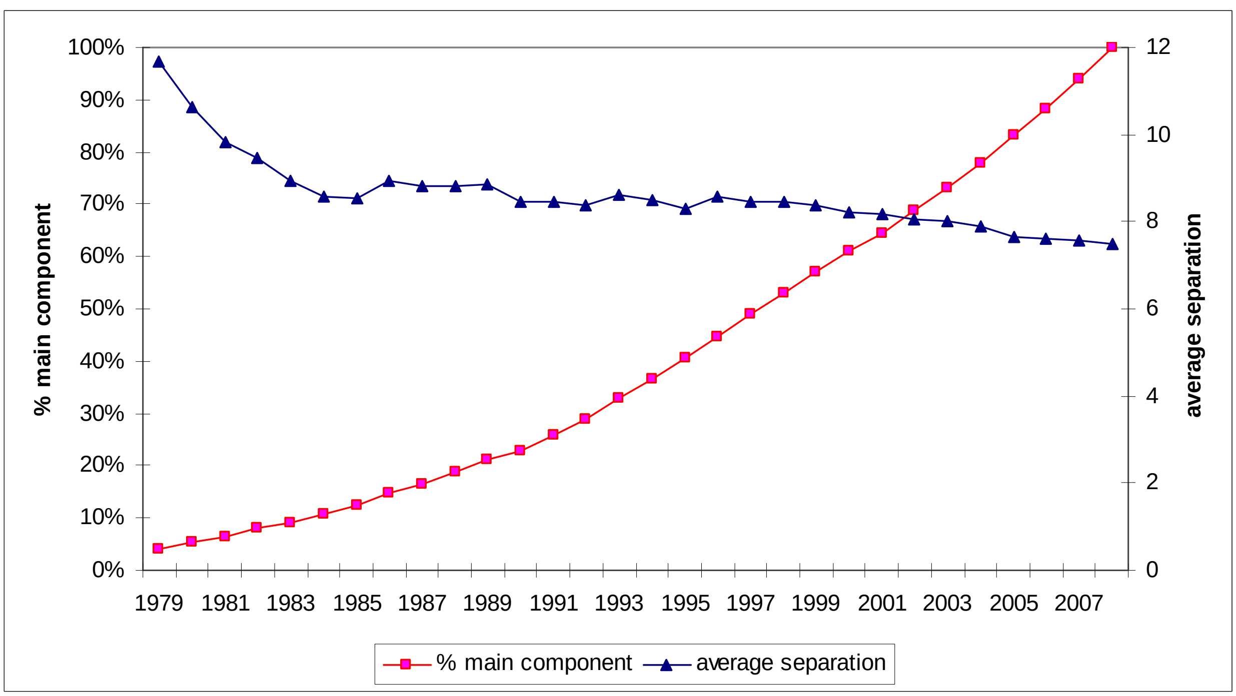
Finally, we can also see from Table 2 that the clustering coefficient is higher for marketing than the other disciplines (50%), meaning that a larger proportion of the co-authors a marketing researcher collaborate among themselves as well. Thus, if a researcher has collaborated with two other researchers, in about half of the cases the two researchers have also collaborated independently. This also helps to partially explain the stark difference in the average number of co-authors between biomedicine and marketing: On average not only do researchers in biomedicine collaborate with more co-authors and write more papers, but with this small clustering coefficient, they write these papers with different individuals.

The Collaboration Network of Marketing Scholars over Time

As already noted, the strength of this database is the fact that it covers a time dependent evolution of the network from its very first stages. Here we provide few simple descriptive figures to help potential users get familiar with the database, and to allow for some quick yet interesting observations. Figure 1 shows the evolution of the collaboration network over time, for two focal concepts: (1) the growth of the main component over time as percent of the main component size in 2008; and (2) average separation. The numbers that are presented are from the years 1979-2008, as the size of the main component before that was very small, and may not be a good indication of the network parameters that we examine.

The line labeled “main component” presents the size of the main component through time, as a percentage of the size in 2008. We see that the size of the main component grows in a nearly linear trend from the mid 1980’s. Our data also show a consistent growth in the percentage of the main component out of all researchers in our database, from about 23% in 1979 to about 69% in 2008. Thus, the collaboration network in marketing seems to show increasing marketing research coherence with time. An interesting result regards the declining average separation, also shown in Figure 1. This finding may be surprising given the increasing size of the system, as network models generally predict that average separation should increase with network size (Bollobas 1985). In this case, in 2008, network size is more than 25 times that of 1979; yet average separation declined more than 35%. This intriguing phenomenon which may be related to the changing collaborating patterns over time is yet to be explored.

Figure 1: Main component and average separation over time



Research Opportunities

The database we have configured provides researchers with opportunities of three kinds: (1) social network research; (2) scientometric analysis of the marketing discipline; and (3) study of marketing journals. In social network research, this database allows a unique observation of how a network evolves over time, from its early inception to maturity. Therefore, we believe this database to be of significant value to scholars from marketing, sociology, economics and physics, to understand network growth. Examples include: (1) social hub formation and their influence on network structure or outcomes; (2) mechanisms of network growth, such as preferential attachment; (3) the role of tie strength and "temperature of the tie".

In scientometric research, any social network can be associated with outcomes for individuals as well as the entire network. Outcomes that scholars may consider to study are: (1) citations, e.g. to test Merton's (1968) argument that scientific impact of scholars depends on their network position; and (2) the magnitude of scientific breakthroughs. An equally interesting endeavor would be to study antecedents of scientific collaboration. Why do some scholars become popular co-authors to many others, while other scholars remain isolated? Finally, from a descriptive perspective, scholars may also examine whether research practices differ across research areas within marketing.

A third area of research is the study of marketing journals and their respective positions, possibly extending earlier work by Pieters et al (1999), Pieters and Baumgartner (2002) and Tellis, Chandy and Ackerman (1999). As we provide information on which journals connect scholars, one can infer which journals serve as bridges between otherwise disconnected scholars. One can connect network descriptors of journals to journal outcomes, such as a journal's standing (credibility as measured in surveys, e.g. Theoharakis and Hirst, 2002), journal submission statistics, or a journal's ISI impact factor.

Limitations of Data

There are some limitations to the data we provide. First, we cannot claim that our data contain the inception of the marketing scholar network. While our data start very early, marketing scholars did already publish in *Journal of Marketing* and much of marketing science was published in journals that we now consider outside our field, such as *Management Science*, as well as in operations research, psychology, and economics outlets. Second, even today marketing scholars collaborate outside the marketing journals. Thus, the results one finds based on our data are bound to the journals we included.

Note that ProQuest and similar citations databases receive data from multiple sources and formats. Small changes in names or other information can lead to misclassification. We have aimed to correct such problems where noticeable, yet given the magnitude of the database, it is certainly not perfect. While the picture of the formation of the marketing network is robust to infrequent misclassifications, this may be an issue if one tries to use our database for an individual level analysis, where each article may be of significant importance. Given the fact that we do not cover all information sources as described above, the fact that focus on collaborations and do not report single-author papers, and the limitations of databases such as ProQuest were some misclassifications can occur, readers are cautioned not to use this database as a source of individual level analysis such as for academic promotions.

In addition, some of the future research we suggest may require additional data. Two particular research issues come to mind. A first issue is the relation between authors' network position and their impact, as measured by citations. There are two main sources for citation data, namely ISI and Google Scholar. ISI's citation data cannot be easily manipulated by scholars, as it includes only cites in other peer-reviewed journals. Access to ISI's citation data is easy through Web of Science and one can also buy such data from bibliometric institutes with subscription to ISI (e.g. CTWS at Leiden University, The Netherlands, cfr. Stremersch et al. 2007). Its main disadvantages are that it is biased towards English-language outlets, only contains full name and first initials of the author, thus leading to potential errors, and is proprietary to ISI (which is the main reason we did not include citations in our database). Google Scholar data are publicly available, contain full author names - leading to less measurement error across authors with the same last name and initials, as compared to ISI - but can easily be manipulated by scholars as it contains citations from any source, published in a peer-reviewed journal or in students' theses. For Google Scholar, we have included a query tool in our database submission that already contains all names in our network database, which one can activate by pressing the command button, after which the query tool retrieves the total citation count across each author's 100 most impactful papers. Google Scholar typically allows retrievals of the total number of citations of around 500 scholars before it blocks access (data checked: November 27, 2008), as a protection against hackers. Therefore, researchers that want to use this tool should make multiple queries spread out over time or bound their inquiry to a certain subset of the network.

Additional data will also be needed to examine the reason why a scholar is a popular co-author. To answer this research question, one needs to inventory information beyond a scholar's impact and network position, such as on their school affiliation, education history, editorial board positions, probably through their bios. Despite these limitations of the data we offer, we hope it provides a starting point for fruitful inquiries in this exciting field.

References

Barabasi, A-L, H. Jeong, E. Ravasz, Z. Neda, A. Schuberts, T. Vicsek. 2002. Evolution of the social network of scientific collaborations. *Physica* A **311** 590-614.

Bollobas, Bela. 1985. *Random Graphs.* London, U.K.: Academic Press.

Grossman, J.W., P.D.F. Ion. 1995. On a portion of the well-known collaboration graph. *Congressus Numerantium* **108** 129-131.

Merton, Robert K. 1968. The Matthew effect in science. *Science* **159** (3810) 56-63.

Moody, James. 2004. The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review* **69**(2) 213- 238.

Monge, Peter R., Noshir S. Contractor. 2003. *Theories of Communication Networks.* Oxford, U.K.: Oxford University Press.

Newman, Mark E.J. 2001. The structure of scientific collaboration networks. *Proceedings of the National Academy of Science USA* **98** 404-409.

Newman, Mark E.J. 2003. The structure and function of complex networks. *SIAM Review* **45**(2) 167-256.

Newman, Mark E.J. 2004. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Science USA* **101** 5200-5205.

Pieters, Rik, Hans Baumgartner, Jeroen Vermunt and Tammo Bijmolt (1999), "Importance and Similarity in the Evolving Citation Network of the International Journal of Research in Marketing," *International Journal of Research in Marketing*, 16 (2), 113-127.

Pieters, Rik and Hans Baumgartner (2002), "Who Talks to Whom? Intra- and Interdisciplinary Communication of Economics Journals," *Journal of Economic Literature*, 40 (2), 483-509.

Stremersch, Stefan, Isabel Verniers, Peter C. Verhoef. 2007. The quest for citations: Drivers of article impact. *Journal of Marketing* **71**(July) 171-193.

Stremersch Stefan, Peter C. Verhoef. 2005. Globalization of authorship in the marketing discipline: Does it help or hinder the field? *Marketing Science* **24**(4) 585-594.

Tellis, Gerard J., Rajesh K. Chandy, David S. Ackerman. 1999. In search of diversity: The record of major marketing journals. *Journal of Marketing Research* **36**(February) 120-131.

Theoharakis, Vasilis, Andrew Hirst. 2002. Perceptual differences of marketing journals: A worldwide perspective. *Marketing Letters* **13**(4) 389-402.

Watts, Duncan J. 1999. Networks, dynamics and the small-world phenomenon. *American Journal of Sociology* **105**(2) 493-527.