A Preliminary Social Network Analysis of MPACT

Yong-Mi Kim
School of Information, University of Michigan, 1075 Beal Ave., Ann Arbor, MI 48109-2112 kimym@umich.edu

The dissertation author and committee relationships in the MPACT dataset were modeled as social networks. Exploratory network analysis was performed primarily on the network composed of dissertation advisors and advisees to investigate the utility of structural prestige for measuring mentoring impact. Three prestige measures were investigated: 1) outdegree, 2) output domain, and 3) proximity prestige. All three measures were highly correlated when computed across the entire dataset, but correlation was lower for those scoring in the top 25 of any of the three. Correlation with citation ranking was low for this subset.

Introduction

University faculty productivity has typically been examined in terms of research productivity, usually measured through number of publications or citation counts. But another aspect of faculty productivity is that of the production of new Ph.D.'s (Crosta & Packman, 2005; Marchionini, Solomon, Davis & Russell, 2006). Various mentoring activities are undertaken towards this end, the most direct form being participation in a student's dissertation committee, either as chair or as a committee member. Such involvement can require a significant investment of time and effort from a faculty member. This mentoring ensures the production of scholars capable of making continuing contributions to their respective fields. One question that arises is how to measure mentoring impact.

Marchionini et al. (2006) explored this question by assembling a dataset of dissertation authors, advisors and committee members for dissertations in information and library schools, and investigating a number of exploratory metrics for mentoring impact. The measures were all based on raw counts of direct dissertation author-committee chair or advisor relationships. In this paper a number of measures used in social network analysis categorized are applied to networks constructed from the MPACT dataset. Three kinds of prestige measures are explored: 1) outdegree, 2) output domain, and 3) proximity prestige. Analysis focused on advisor-advisee networks, excluding non-advisor committee

relationships. This was because mentoring impact measures that included indirect relationships were being explored, and dissertation committee members were not expected to have mentoring impact beyond the dissertation author to the dissertation author's advisees.

MPACT Dataset

The MPACT dataset was obtained from the MPACT team in February of 2007. The analysis presented in this paper is based on the state of the MPACT dataset in February of 2007. Marchionini et al. (2006) report complete datasets for six schools, Drexel University, Florida State University, Indiana University, the University of California, Los Angeles, the University of Illinois, and the University of North Carolina at Chapel Hill. The dataset for the University of Michigan was completed by the author as part of this study. The dataset includes 2469 dissertations, 1713 of which have complete advisor and committee information. Advisors have been identified for 2039 dissertations (83%) and committee members have been identified for 1714 (69%) of dissertations. In addition, there is complete advisor data for 14 other schools. Thirty-two schools in the U.S. and Canada are represented for the 1964-2004 time period.

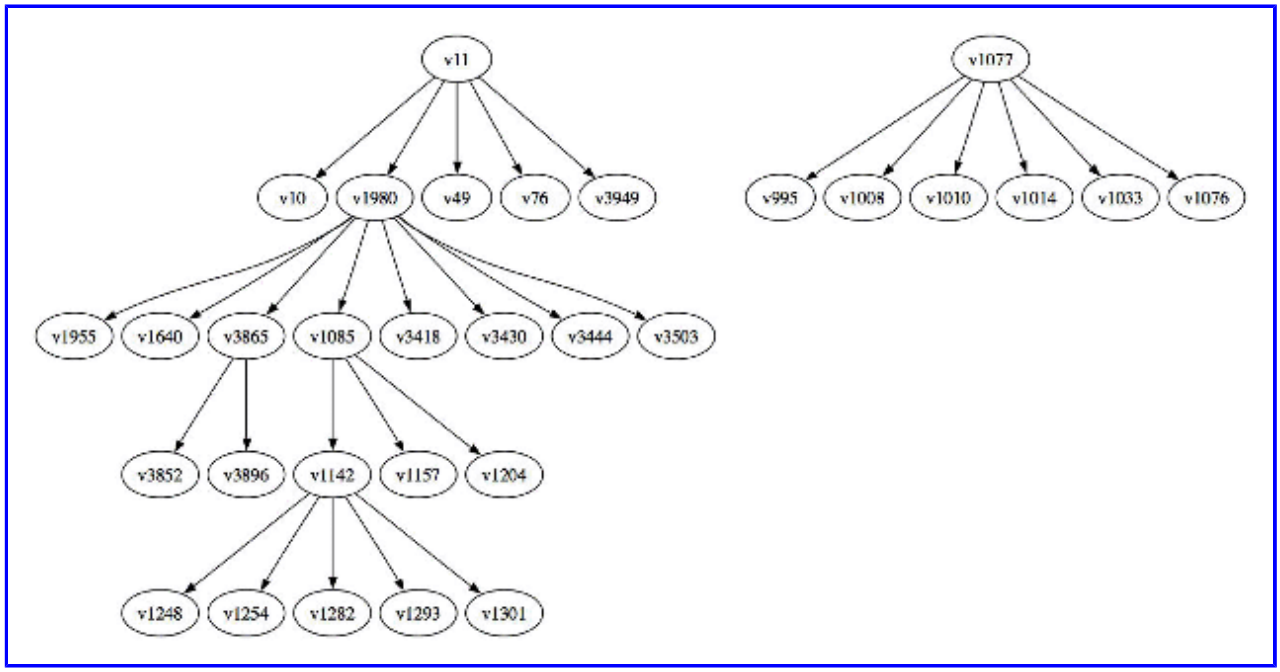
This dataset was used to construct networks to examine different aspects of dissertation committee work. A network was constructed with only dissertation advising relationships, omitting committee relationships. This was motivated by the higher coverage of dissertation advising in the dataset and in the perceived difference in the amount of faculty involvement in advising versus committee work. In addition, it was felt that contribution to mentoring impact from students other than one's immediate advisees was more likely to come from their advisees than from students linked through committee participation. A complete dissertation committee network with both advisors and committee members was also constructed for comparison purposes. Several Python scripts were used to generate a MySQL database containing the MPACT dataset in a form more amenable to analysis, as well as generating input files for use with Pajek (de Nooy, Mrvar & Batagelj, 2005). GUESS (Adar, 2006) was used for visualization.

Prestige in Social Network Analysis

In social network analysis, prestige is a property derived from the patterns of social ties of a particular social network. Sometimes referred to as *structural prestige* (de Nooy et al., 2005), it may reflect conventional assignments of prestige depending on the particular relations and attributes modeled in the network. Three prestige measures, outdegree, output domain, and proximity prestige, are illustrated below through a common example.

Figure 1 shows a network with two advisors, v11 and v1077, and their respective advisees. v11 advised five dissertations, and one of these dissertation authors, v1980, later advised eight dissertations, from which emerged further generations of dissertations. v1077 advised six dissertations. This network has a total of 31 nodes and 29 directed edges (arcs).

Figure 1. Example advisor-advisee network



The *outdegree* of a node is the number of arcs originating from that node. In this example outdegree corresponds to the number of advisees supervised by a faculty member. Node v11 has outdegree of 5, while node v1077 has outdegree of 6. A node with no outgoing arcs, such as v3852, has outdegree of 0. Outdegree corresponds to the measure A (number of advisees) used by Marchionini et al. (2006).

The *output domain* of a node is the number of nodes for which there is a path to that node. For example, v3865 has output domain 2, as two nodes, v3852 and v3896, can be reached from it, while v3503 has output domain 0. v11 has output domain 23, which is the total number of its descendants. The v1077 has output domain 6, the same as its outdegree, as there are no additional nodes to be reached from its neighbors. So this measure takes into account v11's impact across multiple generations of dissertation authors, but all contributions are counted equally. That is, v1282 contributes the same as v49 to v11's mentoring impact,

Proximity prestige refines output domain by having closer neighbors count more than far-away neighbors. Following de Nooy et al.'s definition, the proximity prestige of a node is the proportion of all nodes in the network (excluding itself) that are in its output domain divided by the mean distance from all nodes in its output domain. By definition all nodes

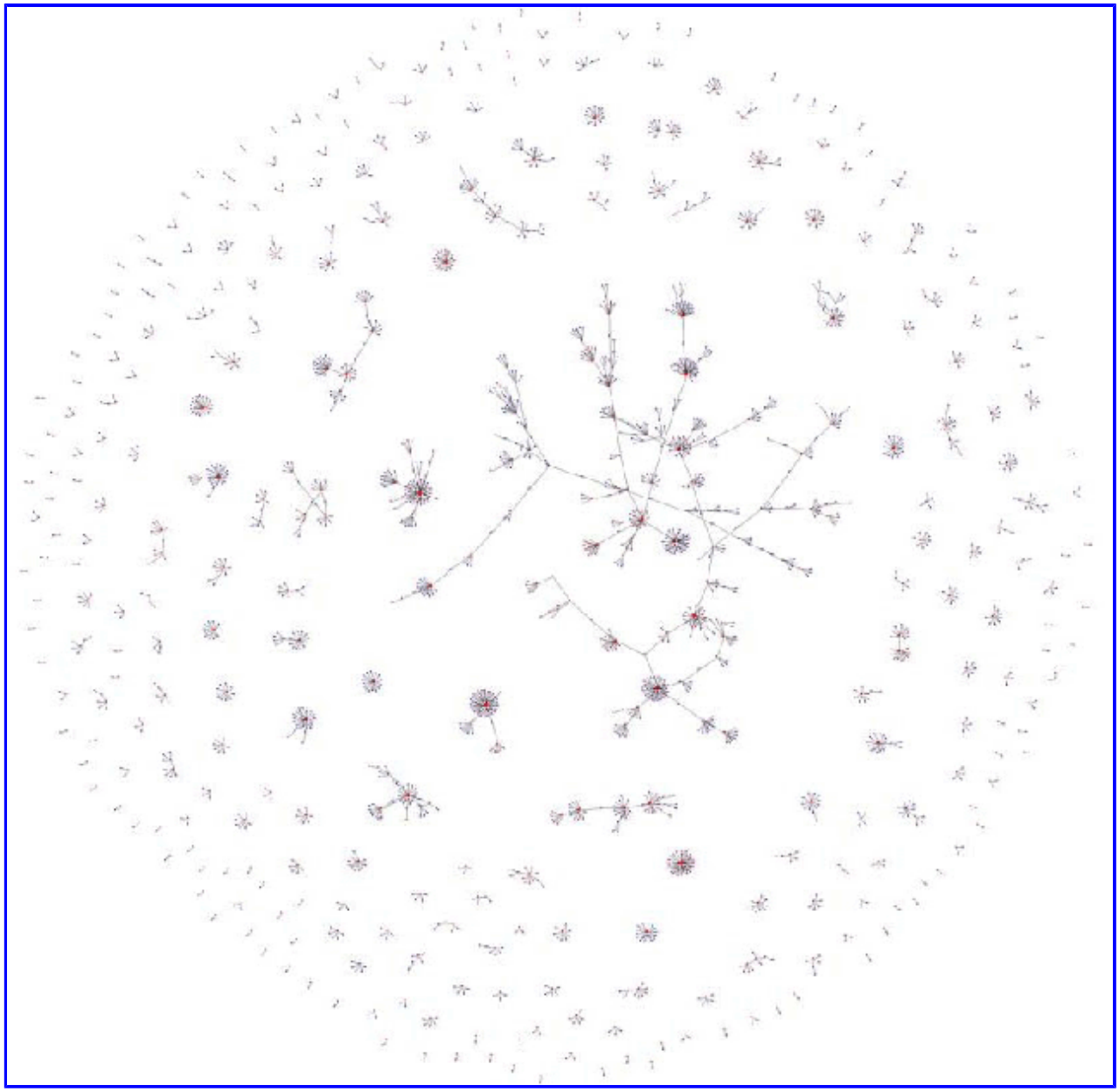
with zero output domain get zero proximity prestige. The output domain of v3685 contains two nodes (v3852 and v3896), which is 2 out of 30 (0.0667). Average distance within the output domain of v3865 is 1, so proximity prestige for v3865 is 0.0667 divided by 1. For v11, it has input domain of 23 divided by 30 (0.767) and average distance of 3.304, for proximity prestige of 0.232. v1077 has proximity prestige 0.2. Advisor v11, having more descendants than advisor 1077, has higher proximity prestige.

This example network was taken from the actual MPACT dataset. Advisor v11 represents Jesse Shera, from whom emerged four generations of advisees: Thomas J. Galvin (v1980), who advised Marcy Murphy (v1085), who advised Debora Shaw (v1142), who in turn had five advisees. Advisor v1077 represents Mildred Lowell, who was roughly a contemporary of Shera's.

Findings

The dissertation advisor network has 2409 nodes and 2126 directed edges. There are 292 weakly connected components, with component size ranging from 2 nodes to 301 nodes. The dissertation advisor and committee network has 3785 nodes and 7829 edges, and 30 weakly connected components. The network has a giant connected component with 3687 nodes, or 97% of the nodes in the network. The next largest weakly connected component has 20 nodes.

Figure 2. The advisor network (GUESS visualization)



Analysis of prestige measures was carried out on the advisor-only network. A total of 510 individuals had advised at least one dissertation, with 311 advising more than one and 230 advising more than two dissertations. David Kaser and Allen Kent advised the largest number of dissertations, 44 and 43 respectively. Table 1 shows the top 25 faculty members according to the three proximity measures. The last column is taken from Table 3 of Adkins and Budd (2006), which ranks U.S. LIS faculty by most citations to their works. A total of 38 individuals are represented across the three prestige measures.

Table 1. Prestige measures and citations (T denotes tied ranking, numbers differentiate tied rankings)

Rank	Outdegree	Output domain	Proximity prestige	Citations*
1	David Kaser	Herbert Goldhor	Allen Kent	Tefko Saracevic

2	Allen Kent	Ralph Blasingame	David Kaser	Nicholas Belkin
3	E. Blanche Woolls	Alllen Kent	Herbert Goldhor	Paul Resnick
4	John M. Goudeau	Margaret R. Sheviak	Ralph Blasingame	Gary Marchionini
5	Ronald D. Blazek	David Kaser	E. Blanche Woolls	Blaise Cronin
6	F. W. Lancaster	Jay E. Daily	John M. Goudeau	Carol Tenopir
7	Jay E. Daily	E. Blanche Woolls	Ronald D. Blazek	George Furnas
8	Ralph Blasingame	William Goffman	Margaret R. Sheviak	Christine Borgman
9	Ellen Detlefsen	Ernest Deprospo	F. W. Lancaster	Jeff Mackie-Mason
10	Dagobert Soergel (T1)	Rolland Stevens (T1)	Jay E. Daily	Rob Kling (*)
11	Herbert Goldhor (T1)	John M. Goudeau (T1)	William Goffman	Carol Kuhlthau
12	Jane B. Robbins (T2)	F. W. Lancaster (T2)	Ellen Detlefsen	Katherine McCain
13	Linda C. Smith (T2)	Ronald D. Blazek (T2)	Ernest Deprospo	Marcia Bates
14	William Goffman (T3)	Rudoph H. Gjelsness	Dagobert Soergel	Peter Hernon
15	Russell Bidlack (T3)	Michael M. Reynolds	Linda C. Smith	Bruce Schatz
16	Patrick R. Penland (T3)	Evelyn Daniel (T3)	Jane B. Robbins	Andrew Dillon
17	Ronald Rice (T3)	Paul Wasserman (T3)	Rolland Stevens	Les Gasser
18	Edie Rasmussen (T4)	Ellen Detlefsen	Ronald Rice (T1)	Amanda Spink
19	Frank Summers (T4)	Jesse H. Shera (T4)	Russell Bidlack (T1)	Raya Fidel
20	Thomas L. Hart (T5)	Dagobert Soergel (T4)	Patrick R. Penland (T1)	Paul Kantor
21	Ana Cleveland (T5)	Sara Fine	Frank Summers (T2)	John M. Budd
22	Tefko Saracevic (T6)	Linda C. Smith	Edie Rasmussen (T2)	Robert Benjamin
23	Nicholas Belkin (T6)	Jane B. Robbins (T5)	Thomas L. Hart (T3)	Carolyn Haythornthwaite
24	Frank B. Sessa (T6)	Raymond Kilgour (T5)	Ana Cleveland (T3)	Michael Lesk
25	Henry Voos (T7)	Ralph R. Shaw (T5)	Nicholas Belkin (T4)	Dietmar Wolfram
	Jean Tague (T7)		Tefko Saracevic (T4)	

* From Table 3 in Adkins and Budd (2006), p. 379.

The Spearman's rho rank correlation for the prestige measures for all 510 advisors is shown in Table 2, and for the 38 individuals ranking in the prestige measures is shown in Table 3. Despite the apparent differences in Table 1, correlation is high when all 510 advisors are examined. Spearman rank correlation for the known top 25 cited individuals and for the

prestige measures is low, 0.2153 for all three prestige measures. Rob Kling, ranked 10th in citings, is not even included in the MPACT database.

Table 2. Prestige correlations for all advisors

	Outdegree	Output domain	Proximity prestige
Outdegree	1		
Output domain	0.9407	1	
Proximity prestige	0.9651	0.9949	1

Table 3. Prestige correlations for all advisors

	Outdegree	Output domain	Proximity prestige
Outdegree	1		
Output domain	0.3508	1	
Proximity prestige	0.8606	0.6469	1

For each individual who had advised at least one dissertation, a "dissertation tree" was generated, with this individual as the "root" and generations of advisees, as shown in Figure 2. The number of generations (or tree height) was counted as follows: the root is 0, the advisees are 1, their advisees are 2, and so forth. For example, Jesse Shera in Figure 2 has four generations of advisees. Rudolph H. Gjelsness is the only one with five generations of advisees. Three individuals, Jesse Shera, Rolland Stevens, and Michael Reynolds (who was advised by Gjelsness), had four generations of advisees. Fifteen individuals had three generations of advisees, 83 had two generations, and 409 had one generation of advisees. Eighty percent of advisors in the dataset had one generation of advisees, with 20% having more than one generation.

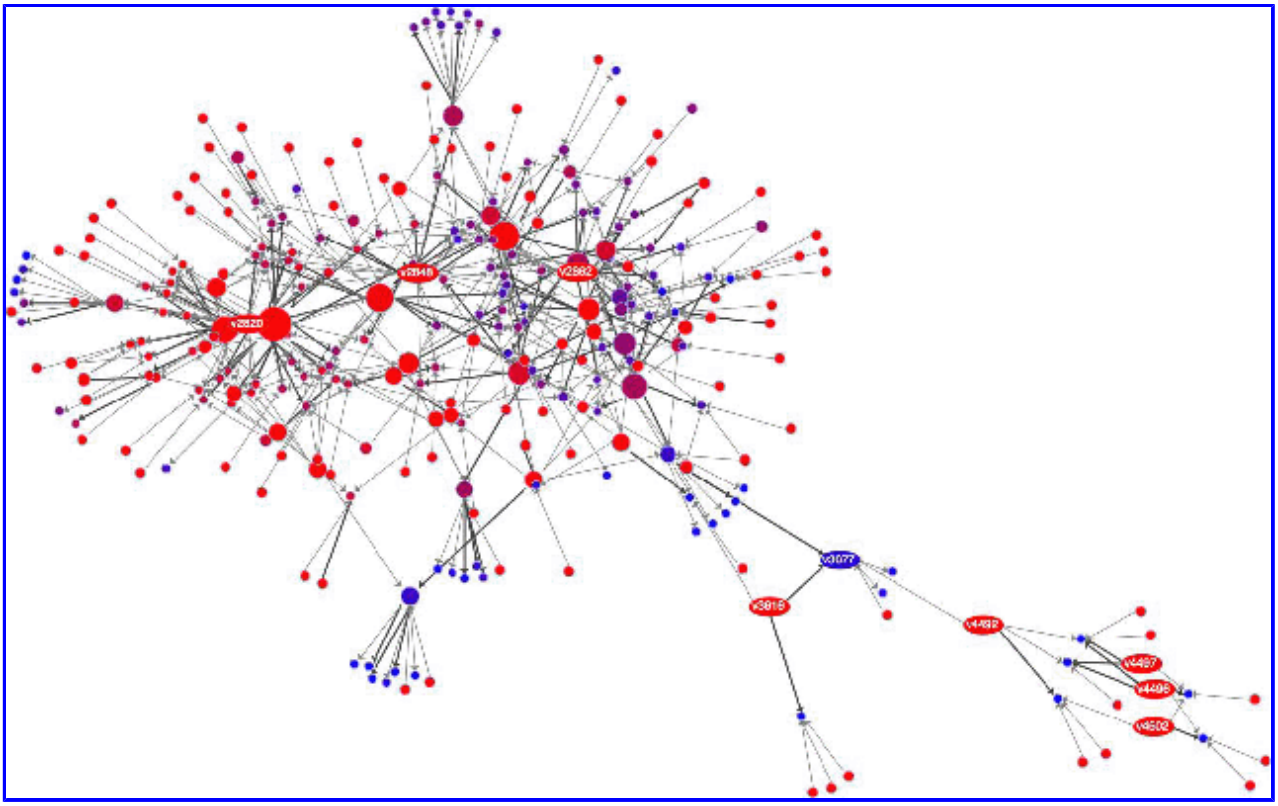
Discussion

While the different measures did yield slightly different names and positions in terms of mentoring impact, correlation was relatively high. The large difference occurs when comparing mentoring impact to research productivity (Persson & Åström,2005; Adkins & Budd, 2006). None of the faculty members listed in Table 1 appears in the top 10 or even top 25 list of citation-based research productivity rankings. Persson and Åström analyzed citation impact in three five-year time periods, and a significant shake-up is observed for the period 2001 to 2004. This corresponds to the time period examined by Adkins and Budd. Future work could involve splitting the data set into different time periods and observing mentoring impact in those different time periods, and comparing this against the research productivity studies.

One area in which LIS programs may differ from other Ph.D.-granting graduate programs is the relatively small number of Ph.D.-granting programs compared to those which only award MLS degrees. Those who earn doctorates in LIS have several alternatives to an academic career, including working as a librarian or entering the private sector. Those who pursue an academic career may do so at an MLS-only institution. Thus a faculty member could supervise many Ph.D. students, none of whom go on to supervise Ph.D. students of their own. Eighty percent of advisors in this dataset had one generation of advisees, which may contribute to the high correlation of outdegree with the other prestige measures for the full set of advisors.

While this study focused on a network containing solely advisor-advisee relationships, another future area of study is analysis of complete dissertation committee networks. Figure 3 shows a GUESS visualization of dissertation committee relationships at Michigan, indicating patterns of collaboration among faculty members. Such a network may provide insight into departmental cohesion, building on previous results by White (1999).

Figure 3. Dissertation authors, advisors, and committee network (GUESS visualization)



ACKNOWLEDGMENTS

The author wishes to thank the MPACT project team for access to the dataset, in particular Gary Marchionini, Terrell Russell, and Cassidy Sugimoto, and Lada Adamic for valuable comments during the development of this study.

References

Adar, E. (2006). GUESS: The graph exploration system. http://graphexploration.cond.org/

Adkins, D. and Budd, J. (2006). Scholarly productivity of U.S. LIS faculty. *Library and Information Science Research,* 28, 374-389.

Crosta, P. M. and Packman, I. G. (2005). Faculty productivity in supervising doctoral students' dissertations at Cornell University. *Economics of Education Review,* 24, 55-65.

de Nooy, W., Mrvar, A., and Batagelj, V. (2005). *Exploratory social network analysis with Pajek.* Cambridge, UK: Cambridge University Press.

Marchionini, G., Solomon, P., Davis, C., and Russell, T. (2006). Information and library science MPACT: A preliminary analysis. *Library & Information Science Research,* 28,

480-500.

Persson, O. and Åström, F. (2005). Most cited universities and authors in Library & Information Science 1990-2004. *Bibliometric Notes,* 7(2). Retrieved April 19, 2007, from http://www.umu.se/inforsk/BibliometricNotes/BN2-2005/BN2-2005.htm

White, W. J. (1999). Academic topographies: A network analysis of disciplinarity among communication faculty. *Human Communications Research,* 25(4), 604-617.