Complementing the Numbers:

A Text Mining Analysis of College Course Withdrawals

Greg V. Michalski

Florida State College at Jacksonville

Author Notes

Paper presented at the Association for Institutional Research Annual Forum, Toronto, May 24, 2011. Correspondence concerning this paper should be addressed to Greg V. Michalski, Office of Student Analytics and Research, Florida State College at Jacksonville, 501 W. State Street, Jacksonville, FL, e-mail: gmichals@fscj.edu

Abstract

Excessive college course withdrawals are costly to the student and the institution in terms of time to degree completion, available classroom space, and other resources. Although generally well quantified, detailed analysis of the reasons given by students for course withdrawal is less common. To address this, a text mining analysis was performed on open-ended, verbatim, student comments in which students explained their reason(s) for course withdrawal. The text for all comments was extracted from the course withdrawals database of Florida State College at Jacksonville, a large, diverse, multi-campus institution located in northeast Florida. An initial set of 616 comments from the fall 2010 term was used to develop a preliminary text mining model which categorized 96.1% of all records. The model was retained and further tested using a second set of 679 comments from the spring 2011 term and found to categorize 98.7% of the term records. Combined data from both terms (n = 1,295) was used to produce a final text mining model containing eleven node categories. Model node categories were labeled referencing a framework of prior empirical work in the area of student course withdrawal. Leading academic rationales include course characteristics (especially those involving student preparedness, satisfaction, and delivery mode), faculty satisfaction, and schedule adjustments. Leading non-academic rationales include personal issues especially involving job/work, family, financial, and health. Record classification data from the model were also exported and explored to further group and summarize results. Principal Components Analysis of all data from both terms revealed four components which accounted for 45% of the total variance with the first two components involving instructional delivery and student personal issues accounting for 24% of the variance. Hierarchical Cluster Analysis and Multiple Correspondence Analysis were also used to confirm results suggesting major academic withdrawal reasons to include negative course perceptions and to a lesser degree negative faculty perceptions. Non-academic rationales were found to center on job-work, personal, and time-schedule issues. Limitations and implications for institutional research and practice are presented and discussed.

Keywords: text mining, text analysis, college course withdrawals, educational data mining

Complementing the Numbers: A Text Mining Analysis of College Course Withdrawals

This paper discusses the use of text mining to complement more traditional methods typically used to track and analyze student course withdrawals. Many institutions routinely track course withdrawals numerically expressing these numbers in reports as frequencies, ratios, rates, trends, and so on. However, text mining/analysis studies of student comments describing precise reasons for course withdrawal are less common. While traditional quantitative descriptive analyses effectively answer "who, what, when, where, and how" type questions regarding course withdrawals, text mining focuses on the question of "why" students withdraw. By combining and integrating both approaches, institutions will be better positioned to take action to improve service to their students.

Text mining is generally considered to be part of the broader field of data mining (Nisbet, Elder, Elder, & Miner, 2009). The field of data mining is relatively new and still evolving. Data mining has been defined in various ways as "extracting useful information from large data sets", or "the process of exploration and analysis, by automatic or semi-automatic means, of large quantities of data in order to discover meaningful patterns and rules", or "the process of discovering meaningful correlations, patterns and trends by sifting through large amounts of data stored in repositories" (Shmueli, Patel, & Bruce, 2010). And while part of this evolution includes specialization in specific disciplines, including education (Romero, Ventura, Pechenizkiy, & Baker, 2011), less attention has been devoted specifically to text mining. Nevertheless, the growing accessibility of textual knowledge applications and online textual sources has also contributed to an increase in text analytics and text mining research.

Text mining is a form of qualitative analysis involving the discovery of new, previously unknown, information extracted and organized from different written sources. In brief, text mining involves the discovery of useful and previously unknown "gems" of information from textual document repositories based upon patterns extracted from natural language (Zhang & Segall, 2010). Currently a topic of considerable importance in academia and industry, text mining theory and practice has also benefited from increased multidisciplinary interest spanning the public and private sectors involving, for example, multinational government, business and industry, and university research (Berry & Kogan, 2010). While many post-secondary institutions track or otherwise monitor student course withdrawals via quantitative analyses of transactional data, text mining studies based upon verbatim student comments made at the time of withdrawal remain scarce.

Given this a secondary purpose of this paper involves stimulating discourse in this underrepresented area by exploring and extending the use of text mining to more completely understand college course withdrawals and complement quantitative measures of such. Due to the nature of text analysis model building and refinement the findings presented and discussed here are probably best viewed more as emerging or developing rather than conclusive or definitive. From an internal institutional perspective further research is required to examine the reliability and stability results over time. This involves the need for further work and fine tuning of the extraction and categorization process which in turn involves the continued development and refinement of linguistic resources, coding, and categorization strategies. Beyond this further progress and additional results, models, and analytic strategies need to be developed and shared between institutions.

Literature Review

This applied study uses tools, terms, and techniques of text analysis and text mining to better understand student rationale for college course withdrawal. This presents the possibility to examine and review briefly (or at least acknowledge) prior work in at least three fundamental and distinct areas involving (1) text analysis independent of text mining technology, (2) text mining and analytics, and (3) prior empirical work in the area of college course withdrawals. Due to current limitations, none of these is treated exhaustively here.

The first two areas are related but distinct. The area of text analysis, independent of mining technology, acknowledges both the history and well-developed body of knowledge that encompasses the analysis of text prior to the availability of powerful analytical software and mining technologies. In the broadest sense, this can be examined across a wide spectrum of ontological and epistemological frames and lenses with varying emphasis on methods. For example, in studying text analysis as a guide for research in art education, Ettinger & Maitland-Gholson (1990) acknowledged the work of Geertz (1983) in terms of its influence on social science research. The authors commented on efforts in many academic disciplines to redefine the object, methods, and aim of research in social disciplines and further cautioned against the use of superficial descriptions of methods as being either qualitative or quantitative. According to Ettinger and Maitland-Gholson, "At its core, this redefinition involves an implicit questioning of the nature of reality and truth. The important questions at issue appear to be: (1) What is reality? and (2) How do we know it?" Such commentary effectively captures the essence of text analysis as an endeavor transcendent of—or at least secondary to—methods (quantitative, qualitative, mining, or otherwise). With this in mind the current review is limited to a brief historical consideration of text analysis from a positivist perspective that leads up to and then includes text mining and analytics.

The third area of fundamental review involves the study of college course withdrawal and particularly student rationales, inclinations, and motivations that explain or are associated with such. Again, this area is related to, but distinct from, student withdrawal from higher education overall which has been addressed by a robust body of knowledge spanning at least four decades. (see, e.g., Tinto 1987a, b, 2006; Charlton, Barrow, & Hornby-Atkinson, 2006). Given this, the present review focuses on a select group of studies concerning student withdrawal from college courses (but not necessarily college itself).

Text Analysis and Text Mining

Text analysis encompasses a broad class of qualitative and quantitative methodologies and techniques for the social scientific study of communication. Although the technical ability to mine and analyze textual information has undergone vigorous growth largely concurrent with advances in information technology, the idea of analyzing symbolic information in the form of written or printed text is far from new. Depending on how questions regarding the origins of text analysis are framed methodologically evidence can be found to substantiate efforts to analyze printed material using quantitative means at least as far back as the 18th century (Popping, 2000). Beyond that era newspaper content analysis began in the early 20th century and has been characterized as developing in five methodological stages described as (I) frequency analysis, (II) valence-analysis, (III) intensity-analysis, (IV) contingency analysis, and (V) computer analysis (Van Cuilenburg, 1991; cited in Popping, 2000).

Viewed from a North American perspective of qualitative research as a field, the developmental period of these five stages (which largely precede current text mining technology) is situated in what Denzin and Lincoln (2005, p.3) refer to as the second historical moment (modernist or golden age) encompassing a period from 1950 to 1970. Viewed from this perspective text analysis done with the tools and assistance of powerful state-of-the-art information technology is but one among many possible approaches available to the researcher as “bricoleur” seeking to extract meaning from communication as the written comments of others. As a methodological alternative (or complement) to quantitative means, text analysis has also been widely performed through the use of what may be viewed as more traditional qualitative methods. Miles and Huberman (1994), for example, discuss the role of the conceptual framework and development of various manual coding schemes in relation to the text analysis process. Although following a fundamentally different approach compared to automated linguistic based text mining (e.g., based upon natural language processing or other means) these more qualitative approaches also enjoy support from various applications designed to facilitate the process (see, e.g., NVivo9). So decisions about methods and technology choices remain wide and varied. Although the current study happens to have used a particular application and approach, the spectrum of alternative applications and approaches remains wide and open to new inquiry and research in both epistemological and methodological terms.

Text mining is also referred to as text data mining. It involves the discovery of novel information such as associations, hypotheses, or trends that are not explicitly present the text sources being analyzed (Nisbet et al., 2009). The field of text mining and the many applications now available to engage in such has evolved rapidly over the past two decades and is tied closely to the concurrent growth of foundational technologies in areas that include computer science, artificial intelligence, and machine learning among others. Although distinctions are made between purely statistical approaches and those based upon artificial intelligence, one direction of development in the area of text mining is based upon Natural Language Processing (NLP) which is rooted in realm of machine learning traceable back to the work of Turing (1950).

With the onset and widespread adoption and use of database technology and specifically textual databases, steady advancements were made toward the goal of automating human analysis of text. Particularly relevant to the current study is the work of Nasukawa and Nagano (2001) who help to lay the groundwork for current text analysis and knowledge mining. According to these authors

> Large text databases potentially contain a great wealth of knowledge. However, text represents factual information (and information about the author's communicative intentions) in a complex, rich, and opaque manner. Consequently, unlike numerical and fixed field data, it cannot be analyzed by standard statistical data mining methods. Relying on human analysis results in either huge workloads or the analysis of only a tiny fraction of the database. We are working on text mining technology to extract knowledge from very large amounts of textual data. Unlike information retrieval technology that allows a user to select documents that meet the user's requirements and interests, or document clustering technology that organizes documents, we focus on finding valuable patterns and rules in text that indicate trends and significant features about specific topics. (p. 967)

They go on to compare several document handling technologies in terms of function, purpose, technology, data representation, natural language processing, and output. Document handling functions include searching, organizing, and knowledge discovery. The purpose of knowledge discovery is characterized in terms of extracting interesting information from content using natural language processing, mining, and visualization through semantic and intention analysis with the output being "digested information (trend patterns, association rules, etc.)" (p. 968). Given the application employed for the current study this work is of particular interest for its obvious technology overlap (including the application output visuals contained in the article).

The application employed in the current study uses the technology described by Nasukawa and Nagano (2001). After acknowledging the possibility of manual coding (i.e., having people read survey responses, note their contents, determine key concepts and assign codes), and its strengths and merits, chiefly in terms of accuracy, the application's user's guide notes the limitations of manual coding in terms of inter-rater reliability as well as labor intensity and time requirements associated with manual coding. The guide goes on to state:

> There are many different automated solutions to choose from, including statistical and linguistic solutions. SPSS Text Analytics for Surveys offers a combination of automated linguistic and statistical techniques to yield the most reliable results for each stage of the process. In this product, linguistic-based techniques are used to extract the key concepts from the responses automatically, and both linguistic and statistical techniques can be used to create the category definitions (codes) that are assigned to responses. (SPSS, Inc., 2009, p. 5)

Key steps in the overall text mining process involve the extraction of key concepts and the categorization of these into a number of labeled model "nodes". Extraction is done using linguistics based analysis which employs machine-based understanding to increase reliability over purely statistical approaches. Linguistic resources for the process include one or more libraries as well as type and synonym definitions. The main steps in the extraction process include (1) inputting data conversion into a standard format, (2) identifying candidate terms, (3) identifying equivalence classes and integration of synonyms, (4) assigning a type, (5) indexing and, (6) matching patterns and events extraction.

Categorization involves the organization of extracted concepts and can be accomplished in different ways within the application. Two broad text categorization approaches include (1) knowledge engineering approach in which expert knowledge about categories is encoded directly and (2) machine learning in which a category is constructed from a set of existing examples according to a general inductive process (Feldman & Sanger, 2007). Referencing the application used, category building refers to the generation of category definitions and classification through the use of one or more built-in techniques and categorization refers to the scoring, or labeling, process in which unique identifiers are assigned to the category definitions for each record. Both categorization and category building happen simultaneously. During category building, the concepts and types that were extracted are used as the building blocks for categories. Records are automatically assigned to pre-built categories if they contain text that matches an element of a category's definition. Automated category building techniques include peer/sibling grouping and parent/child grouping techniques.

Peer/sibling grouping involves the horizontal association of concepts and patterns and includes (1) shared root concept derivation, (2) semantic network among siblings, (3) co-occurrence or paired usage of concepts. As implied by the name parent/child grouping refers to the grouping of concepts and patterns in a vertical manner based upon subsets. It includes techniques of (1) concept inclusion or word subsets, and (2) parent-child semantic network based upon hyponyms meaning that one concept is a sort of a second concept in hierarchical relationship. Finally, it is important to note that most automated settings within the application can be fine-tuned or otherwise adjusted and tailored to suit any given text mining context. In fact, one of the recommendations for further work in this paper is the suggestion to publish and share library resources among similar institutions or usage groups involving, for example, institutional research in higher learning based upon common research tasks and interests.

Student Withdrawals

The literature on college course withdrawal is related to but distinct from that concerning complete withdrawal of the student from college. While constituting a perhaps more optimistic outcome compared to complete withdrawal from college, individual course withdrawal is problematic in its own right. According to the Florida Department of Education (March, 2011):

> When students enroll in, but fail to complete a course, it costs the student and the state money, reduces available classroom space, and increases the amount of time for the student to complete their degree. Clearly, many withdrawals are necessary for personal and academic reasons, but when withdrawals become excessive they pose a significant burden on the student, the college, and the state. (p. 1)

The report also states system total of 668,854 withdrawals, representing 11.3% of the total course enrollments between 2007-08 and 2009-10. Clearly there is room to improve and a more complete understanding of the reasons students withdraw from courses can assist.

Compared to overall student retention in higher education which has been widely studied and now comprises an extensive body of emerging theory and research spanning at least four decades (Tinto, 2006), the set of empirical studies focusing on selective or discretionary student withdrawal from individual courses is less developed. Nevertheless, support can be found in the literature for at least two general classes of withdrawal reasons. These involve (1) largely academic reasons, related to areas such as grades, instructors, and course, and (2) non-academic reasons related to areas such as family, illness, and military service (Dunwoody & Frank, 1995; Astin, 1997; Wiley, 2009).

On the other hand, student rationales for withdrawing from individual college courses consists of a variety of commonly mentioned reasons ranging from purely logical and/or clearly necessary at one end of the continuum, to ostensibly legitimate and/or tenuously fanciful at the other end. Much of the research done in the area is based upon student surveys and questionnaires which often provide a fixed set of withdrawal reason choices and in some cases open ended comment sections which allow for more precise explanations. Student cited reasons for course withdrawal commonly include those such as *I was not happy with my grade, I didn't understand the material, I didn't like the course/professor, The subject did not interest me* and others according to Dunwoody & Frank, (1995).

From a research perspective, although there is evidence of a general broadening of the methodological spectrum used to describe, understand, and even predict course withdrawals (see e.g., Bambara, Harbour, Davies, & Athey, 2009; Buglear, 2009; Charlton, Barrow, & Hornby-Atkinson, 2006) most studies, especially those involving large institutions, continue to rely heavily (or exclusively) upon traditional quantitative/statistical measures, including the implementation of business intelligence and data mining processes (Wiley, 2009). Such efforts generally involve counting and comparing course withdrawals in relation to various categories or dimensions such as time (e.g., term/semester), course/credit type, student demographics, student major, and others (see e.g., Conklin, 1997; Friedlander, 1981; Hagedorn, Maxwell, Cypers, Moon, & Lester, 2003; Hall, M., Smith, K., Boeckman, Ramachandra, & Jasin, 2003; Lunneborg, Lunneborg, & de Wolf, 1974; Mery, 2001; Reed, 1981; Sumner, 2000; 2001). While such enumerative comparisons are certainly useful, the availability of increasingly powerful and specialized tools to mine and analyze large volumes of purely textual data represents an opportunity to develop a more complete understanding and improved institutional response to student course withdrawals. A key to this strategy involves the ability to efficiently collect and analyze large volumes of textual data.

Institutional Context

The context for the study is Florida State College at Jacksonville, a large multi-campus institution with an annual (2009-2010) unduplicated student enrollment of over 84,000. According to the Florida Department of Education (March, 2011) the college had a course withdrawal rate (withdrawals as a percentage of total course enrollments 2007-08 through 2009-10) of 8.1% based on 364,179 enrollments and 29,655 withdrawals. In reference to student course withdrawal the current catalog states the following:

> *A student may withdraw without academic penalty from any course up to the published withdrawal date. The assigned grade of "W" is not included in the calculation of any grade point average. Course(s) receiving a grade of "W" are included in attempted courses when determining a standard of academic progress. The student will be permitted to withdraw only in*

the first and second attempt. The student is not permitted to withdraw from the course upon the third attempt. Upon the third attempt a student must receive an "A," "B," "C," "D," "F" or "FN" grade for the course.

Since fall term 2009, the course withdrawal process has required a written statement from the student to explain reason(s) for withdrawal. As part of the course withdrawal process the student encounters the following request: "Please provide your reason for requesting a withdrawal." The student is then provided with a text entry area to provide a response. These open-ended (textual) explanations are collected in a withdrawal database along with associated information (such as a unique withdrawal identification number, course reference number, student identification number, and withdrawal submission date). This database now contains well over 10,000 records. As such, reading through the complete set of written reasons given by students to explain their course withdrawals—some several paragraphs in length—is not practical. In an effort to identify and implement an automated text mining approach to extract useful information for decision making, a range of internal and external (commercially available) options were evaluated and considered. IBM/PASW Text Analytics for Surveys (v. 3.0.1) was adopted and purchased to analyze a range of student textual data including withdrawals and open ended responses on student surveys.

Methodology

A pilot text mining project was completed in the summer of 2010 to prepare for larger scale projects at the college. The pilot project involved the analysis of student comments on the Florida State College at Jacksonville *Survey of New Student Experience*. In addition to scaled items pertaining to new student experience in the course SLS 1101 (Dynamics of Student Success), the online (web based) survey also contained an open-ended question requesting the respondent to share feedback about experience as a new student. A test set of consisting of 130 student comments during the months of June, July, and August was extracted and analyzed with the text miner application using default library resources. Results revealed eight major categories subsequently labeled (1) College, (2) SLS (Student Life Skills)

Positive, (3) SLS Negative, (4) Student, (5) Faculty, (6) Positive Experience, (7) FSCJ Positive, and (8) Information Resources. All records in the pilot data set were successfully coded. In addition to being used to evaluate new student experience and make improvements in areas such as advising and the student life skills course, lessons learned from the pilot were applied to the current mining project.

The current project was initiated as an analysis of withdrawal comments from the fall 2010 term only and then expanded to include a subset of comments from the spring 2011 term. Comments from the fall term were used to develop and build the model and those from the spring term were used to test and validate the model. The combined comments from both terms were also analyzed. Text comments from the fall term were drawn from open-ended, verbatim, student comments entered for course withdrawals that occurred between September 1 and September 26, 2010. This period corresponds to the first three full weeks of the fall 2010 semester. The period for the spring 2011 term included comments entered for withdrawals that occurred between January 19 and February 6, 2011.

All comments were taken directly from the course withdrawal database using Microsoft SQL Server 2008 Management Studio and then imported to Excel for cleansing and organization, and then imported (as an Excel file) into IBM/PASW Text Analytics for Surveys (v. 3.0.1) for mining and analysis. The cleansing process consisted of performing a descending alphabetical sort to identify and delete non-comment entries. These include entries in which a student simply typed a random character or entered other non-response character combinations such as “n/a”, “no”, “none”, *etc*. Common terms and abbreviations used at the college were standardized. These included commonly used abbreviations such as FSCJ to refer to the college name, as well as others such as “ bb” or “ BB” to refer to “Blackboard” (an academic learning content and management system), and others. Finally, a spell check was performed in Excel. It should be noted that, although the text miner can handle common misspellings, the application manual suggests fixing these prior to importing:

> While the program accommodates some spelling errors, we recommend that you correct such errors before importing your data into the program. Spelling errors can cause problems in text

> analysis for humans as well as for software programs. The more spelling errors you can correct beforehand, the more reliable the resulting categories are. (SPSS, Inc., 2009, p. 38)

The organized and cleansed set of comments in fall term data set consisted of 616 comments and included several associated reference variables (i.e., course, gender, *etc*.) for each record. This data set was then used in the text miner for exploration and initial model development.

Within the miner application used, linguistic resources are composed of libraries, templates, and compiled resources used for term extraction and development. Libraries include lists of words, relationships, and other information to specify and/or tune an extraction through iterative refinement. For the present study the budget, core, opinions, customer/product satisfaction, and variations libraries were used for initial extraction and and subsequent development and analysis. These were subsequently developed, tuned, and improved through iterative testing to create a custom Text Analysis Package (TAP). A TAP is a bundle of linguistic resources that can be applied to the analysis of text in a mining project. The TAP contains category sets and mining project resources used to extract terms, types, concepts, and patterns. A TAP can be constructed from the contents of any mining project that contains at least one category and some linguistic resources including concepts, types, rules, and patterns.

For the present study a custom TAP was developed to mine withdrawal comments in the fall 2010 data set. The final TAP used to create the model described was labeled and saved as an application file (StudentWithdrawF2010.tap) within the miner project. To enable additional and future comparisons among the final model categories, a preliminary set of seven reference variables was also defined. The reference variables defined were (1) campus/location, (2) class time block, (3) course credit type, (4), course identifier, (5) student gender, (6) student race, and (7) instructor name. Of these, only the course identifier is discussed in the present study. Several preliminary comparisons were made using the remaining reference variables; however, a discussion of those results is beyond the scope of the present study. The course identifier was used to check and compare the proportionality of courses in the withdrawal comment data set against that of all courses withdrawn from in the fall 2010 term (i.e., academic history W grades for the term).

Coded results from both terms were exported as a PASW file for quantitative analysis. Using the course identifier reference variable, proportionality comparisons of courses withdrawn from by term, as well as between terms, were made. Correlation analysis was also performed in addition to several additional exploratory procedures including cluster analysis, principal components analysis, and multiple correspondence analysis. These procedures were used to further examine natural groupings of comments coded together within model nodes.

Results

The final set of categories extracted from the fall 2010 data set accounted for 96.1% of all responses. Referencing the course withdrawal literature framework summarized previously, eleven major final model node categories were identified and labeled. The categories were named (1) time-schedule, (2) job-work, (3) family, (4) health, (5) financial, (6) personal-other, (7) information technology, (8) faculty negative, (9) course negative, (10) online course, and (11) federal service. Table 1 shows an individual count of records both within, and shared between, categories. The categories, also referred to as nodes in the model web diagram, are shown in Figure 1 which depicts the categorization of 592 of the 616 total responses (96.1%) into 11 nodes based on the fall 2010 data. Additional figures showing web diagrams for shared responses between each category are contained in the Appendix. Each web diagram uses relative circle diameter to represent the number of cases (responses) categorized into each node. The graphical model also uses line thickness to represent the number of responses shared between nodes. Additional detail can been seen in Figure 2 which is a category bar chart showing total number of responses coded into each category of the model based on fall term data. As shown the time-schedule node contains the most comments with 331 and the federal service node contains the fewest comments with 11. Next, several brief examples of comments used to develop the model are provided to illustrate both the scope of the comments how they are coded into one or more categories.

Time-Schedule

Several examples of comments coded into the time-schedule follow. Note that because a comment can be coded into more than one category the list of categories in which the comment was coded is shown in [brackets] following the comment. Time-schedule withdrawal comments examples include,

- I do not have the time to perform my best in this class. Also, my schedule doesn't work well with this class. [time-schedule, personal-other]
- Don't have the time needed to complete this class with my current job. [time-schedule, job-work]
- No time. Work. [time-schedule, job-work]

Job-Work

Many students expressed reasons for withdrawal related to job and work and there is substantial overlap between this category and the time-schedule category. Examples include,

- With work and other classes, I don't have the time available to commit to this class. [time-schedule, job-work]
- Did not realize when I registered for the class that it was from 8:00 AM - 12:05 PM. I work and this class doesn't work with my work schedule. I will take the class next semester and be more conscious of the class times. [time-schedule, job-work, personal-other]
- I just can't find the time due to the fact that so busy at work. I cannot apply myself as needed. I will try to work it out before next semester. [time-schedule, job-work, personal-other]

Family

Several categories relate to and have substantial overlap with others. For example, because the health category includes the health of the student as well as others (e.g., family members) a withdrawal comment involving health of a family member is coded into both the health and family (as well as other possible) categories. Examples of comments coded into the family category include,

- My son has started Kindergarten; I have a full course load, three girls, a husband, and a home to care for. I am extremely busy. I need to allow time to properly teach my son and be sure there is time for everything else I have going on. [family, time-schedule, personal-other]
- The schedule times interfere with my daughters' daycare. The class ends at 6 and her daycare closes at 6. [family, time-schedule]
- We've recently been forced to deal with an estate issue on her father's behalf and have realized that a full schedule is too much while taking care of our special needs three year old. [family, time-schedule, personal-other]

Health

As mentioned, course withdrawal explanations related to health may include those related directly to the health of the student as well as others close to the student such as core and extended family members. In many cases comments categorized as health are also included in the time-schedule and/or personal-other categories. Examples include,

- Due to newly received medical treatment on this day I was advised by my doctor to reschedule this course. I was advised that if I don't make changes to my schedule it may affect my treatment. I did not know at the beginning of class that I would have to receive treatment on this day. [health, time-schedule, personal-other]
- I haven't had time for my studies due to my grandmother being admitted to hospice. I have been trying to spend time with her before her passing. [health, time-schedule, family]

- Right now I need to focus on myself and getting myself healthy, physically, mentally, and spiritually. It is too much stress and anxiety to worry about classes too. [health, time-schedule, personal-other]

Financial

Financial withdrawal comment examples include the following:

- I'm requesting a withdrawal because although I'm enrolled in this course now, I don't have the money to pay for my books right now. So I was hoping to drop this course and re-register in a later dated course so that my financial aid will pay the expenses. [financial, time-schedule]
- It was explained to me on 9/7/10, that there was no funds to pay for class therefore the system would automatically drop the class and I did not have to drop it myself and will not be responsible for anything. [financial, time-schedule, info technology]
- I am withdrawing because my class wasn't paid for at the time I registered. [financial]

Information Technology

Information technology withdrawal comments include a wide range of explanations describing personal computer issues, internet connectivity issues, online learning systems issues, learning styles and preferences, and others. Several examples include,

- (I) lost internet connection for approximately 2 weeks during the beginning of the semester (and) missed an important assignment that would not allow me to continue. [info technology, personal-other]
- If I had known that I would have assignments due every other day and that the class requires a computer with Microsoft 2007, I wouldn't have signed up for this class in the first place. I received my book two weeks late for an 8 week course, even though I ordered the book 2 weeks before the class started, and at that point, even if I made a hundred percent on everything else in the class, I would make an 86 final grade for the

class at the absolute maximum, which for me, is not acceptable. I could not do anything without the book and my instructor has been slow to respond and tell me what if anything I can do to save my grade in this course. [info technology, time-schedule, personal-other, faculty negative]

- I would rather take the class in a classroom. I don't like taking the class by computer. Too much information too fast and too many distractions at home. I will sign up for it again after I have taken some math classes. [info technology, personal-other, course negative, online course]

Faculty Negative

A range of student perceptions of and reactions to faculty are represented in the faculty negative category. Many were seen to include comments related to instructional style or method. Examples faculty negative comments include,

- The reason I am dropping this class is because I am very lost in the class and the instructor teaching method is very poor very hard to follow and before I fail this class I want to drop the class and pick it up with another instructor. [faculty negative]
- The teacher sucks at teaching. All he does is read off a power point. That is NOT teaching! I can't learn that way and I'm surprised anyone else can!! [faculty negative, course negative]
- This teacher does not seem to understand that the reason for taking online classes is because people have busy schedules but still want to be able to go to school. I think that this teacher was very rude in his e-mail. [faculty negative, time-schedule]
- Better instructor [faculty negative]

Although faculty negative comments were found to be positively correlated with course negative comments, there are also examples of purely course negative comments in which a student may even specifically express his or her satisfaction with the instructor but not the course.

Course Negative

Course negative comments are generally focused specifically on the course but may also be coded into multiple categories. Comments in this category were found to range from very simple and straightforward expressions of how the student found the course to be "boring" to more detailed reasons. Some examples include,

- Don't like the class. [course negative]
- The class is boring and not engaging. [course negative]
- Not enough time, boring class. [course negative, time-schedule]
- Do not feel comfortable in the class. It has been years since I took my last Algebra class. I am not catching on to the concepts fast enough. [course negative, time-schedule]
- I'm requesting a withdrawal, because I'm not learning anything from this class, I'm not a student who can teach (my)self, I didn't sign up for an online class. [course negative, faculty negative]

The last comment expresses the student's frustration suggesting a mismatch in expectations between how much active "teaching" was expected versus perceived. This comment is also notable because, although it contains the words "online course" it was not coded into that labeled node category because the term was only used in a descriptive sense and not as are reason for withdrawal. The next section illustrates the online course withdrawal category.

Online Course

Many comments in this category involve students withdrawing from a course because they would prefer to take the same course in a traditional (classroom) setting rather than online. Examples include,

- I think I need to do this one in a classroom atmosphere. I am worried about it being online. [online course]

- I would rather take the class in a classroom. I don't like taking the class by computer. Too much information too fast and too many distractions at home. I will sign up for it again after I have taken some math classes. [online course, personal-other, info technology]
- Not able to take a hybrid course due to my conflicting schedule. Would rather take a normal class and actually be taught. Personally, I'm not a self-learner. [online course, time-schedule]
- Spanish is a hard class for me to do online. I will retake in person. Nothing wrong with school or class, just wrong format for me to do well. [online course, personal-other, time-schedule]
- The class is hard for me to follow online and I can't afford to fail. I am going to take it in a class setting next semester. It is nothing against the instructor I just need to be in a class setting for English. [online course, personal-other, time-schedule]
- Online classes not as challenging. [online course]

In most cases, students indicate a preference to withdraw from the online course and take the same course in a traditional classroom setting because they find the online format too challenging, however, as indicated by the last example (above), the opposite can also be true. Taken together, the course negative and online course categories lend support to the priority and importance of effective instruction as well as the alignment of student instructional expectations.

Federal Service

With the formalization and growth of the newest college division (Military Public Safety & Security), the federal service category includes withdrawal comments related to military deployments, but also other federal service commitments as well. Because Jacksonville has a large naval presence, and the college also serves other naval locations (e.g., Pensacola, Great Lakes, San Diego), this category includes many comments specifying service in the navy , but also other federal service commitments (e.g., other military branches, homeland security, etc.), as well. Examples include,

- Deploying to Iraq soon. [federal service]
- Going Active Army. Cannot Move forward in Class. [federal service]
- I have to withdraw from this class. I am a contractor for the Department of Homeland Security. I have to travel to Guantanamo Bay, Cuba every month for work and do not have time at this point to have an on campus dedicated class. My other two classes are online. If that is an option for this class I would like to do it online as well. [federal service, time-schedule, online course, job-work]

With the eleven categories established and the coding rules set, the model was further tested using withdrawal comments taken from an equivalent period during the first three weeks of the spring 2011 term. A a check of reliability, the relative proportion of courses withdrawn from in the fall 2010 terms was compared to those of the spring 2011 term. A similar check was also made by comparing the proportion of courses withdrawn from in the fall 2010 to complete term data (retrieved as "academic history" grades) at the conclusion of the full fall term.

Course Withdrawal Frequency and Proportionality Comparisons

To ensure that the data from fall term used to develop the model accurately reflected overall course withdrawal proportions for the entire fall term, as well as those obtained from the spring term, the course identifier reference variable was used to examine and compare course frequencies and proportions. The proportionalities were found to match closely based upon a comparison of the top six most frequently withdrawn from courses. Figure 3 summarizes a withdrawal comparison of the top six courses in fall term text analysis data ($n = 616$) with all withdrawals from the full fall 2010 term ($n = 84{,}083$). The same subset of six courses was present in both the text analysis sample and full fall term grade set. The Pearson correlation between the two was positive and significant (Pearson's $r = 0.84$, $p < 0.01$). Table 2 contains detailed frequency count, rank, and cumulative percentage comparisons between the text analysis sample and full fall term.

Course frequency proportion comparisons were also done by term. Because the model produced by the text miner was developed using withdrawal comments taken from the fall term only it was considered important to test the model using results from an additional term. The comparison set of 679 withdrawal comments taken from the spring 2011 term was used for this purpose. The idea was to compare the proportions of comments categorized into each of the eleven node categories specified in the original model using the same text miner library resources and extraction settings on the spring 2011 text data. Table 3 shows an individual count of records within and shared between categories for the spring 2011 term. Comparing these results to those from fall 2010 (Table 1), the top three categories for both terms include time-schedule, personal-other, and job-work with the remaining category counts being proportionately similar for both terms as shown in Figure 4 which depicts record coding frequencies as counts for both terms.

Correlation Analysis

Beyond comparing counts of comments coded into (i.e., shared between) two nodes using the category web tables, a cross-tabulation matrix can be used to efficiently view record counts shared between all model nodes. Additionally, correlation analysis and other related procedures can also be performed to further explore how records are coded into two or more nodes produced by the text miner. The results can be used to further understand the complexity of course withdrawal rationales reflected in the text mining model.

The number of comments coded into each category and shared between categories for fall 2010 is shown in Table 4. To more completely understand relationships between comments that were coded into more than one category, correlation analysis was used to examine results from each term as well as for all records from both terms. As a nonparametric measure of the rank-order association between two variables regardless of their distributions, Spearman's rho (ρ) was calculated for each term as well as for both terms combined. Internode rank correlations were also calculated and tested for statistical significance. The internode rank correlation results for the fall 2010 term are contained in Table 5. Positive and significant correlations were observed between several node categories including course-negative and faculty-negative ($\rho = 0.301$, $p < 0.001$) as well as information technology and course negative ($\rho = 0.215$, $p < 0.001$). As discussed earlier, in reading the verbatim comments it makes intuitive sense to see a positive, significant correlations between several nodal categories (e.g., faculty negative and course negative, health and family, etc.). For other categories the result is less intuitive. An example is the relationship between information technology (info_tech) and course negative. In reading the comments that were coded into these categories, however, an often cited reason for withdrawal involves students who originally register for an online section of a course and who subsequently withdraw in favor of a face-to-face (classroom) version of the same course. Many such comments contain information technology terms as well as negative comments associated with the online course. A similar correlation analysis was also performed on the spring 2011 term data and the correlation results for all records from both terms combined is shown in Table 6. As for the fall 2010 term alone, multiple significant correlations were observed. The results of several multivariate analyses used to more completely explore and understand the correlation patterns and natural groupings in the exported data are described next.

Principal Components Analysis

Principal Components Analysis (PCA) was used to further understand the structure and patterns of correlations in the model for records coded into multiple nodes. A central goal of PCA is extract a small number of uncorrelated variables containing as much of the information as possible in the original data set. Based on the presence of multiple significant correlations including at least one in excess of 0.30 (course/faculty negative) in the fall 2010 correlation matrix, PCA was used to further explore the eleven node variables by individual academic term and for both terms combined. To prepare for PCA the Kaiser-Meyer-Olkin Measure of Sampling Adequacy was calculated and although it was found to be slightly less than 0.60 (0.466 for fall 2010), Bartlett's Test of Sphericity was highly significant (approximate chi-square = 388, $p < 0.0001$) supporting the marginal factorability of the correlation matrix.

For each PCA performed, two different methods were used to determine the number of components to extract. First, scree plots were examined for obvious breaks between components. Next, parallel analysis was used to as a quantitative check of the scree plot examination results. Parallel Analysis (PA) involves the generation of of random correlation matrices to compute eigenvalues to compare with the experimentally obtained data, in this case the coding of comments into one or more of the eleven node categories as reflected in the correlation matrix. Components are retained in the experimental set until their values are found to be less than the corresponding value generated by the PA. According to Watkins (2006):

> PA requires that a set of random correlation matrices be generated based upon the same number of variables and participants as the experimental data. These random correlation matrices are then subjected to principal components analysis and the average of their eigenvalues is computed and compared to the eigenvalues produced by the experimental data. The criterion for factor extraction is where the eigenvalues generated by random data exceed the eigenvalues produced by the experimental data. (Watkins, 2006)

A PCA of the fall 2010 data revealed a clear break in the scree plot after the third component suggesting that three principal components should be retained. This was further supported using Parallel Analysis which showed only three components with eigenvalues exceeding the corresponding criterion values in the generated data matrix of equivalent size (i.e., 11 variables and 616 subjects). The three components explain a cumulative percentage of 36.49% of the total variation. The first component includes the node categories of Info Technology (0.67) and Course Negative (0.66). The second component includes the node categories Personal-Other (0.75), Family (0.46), and Health (0.29). The third component contains the categories of Time-Schedule (0.63) and Job-Work (0.33). Figure 5 contains two-dimensional views of the first three components and Table 7 contains the rotated component matrix showing the component loading values. The three components extracted make sense, especially viewed against the literature which categorizes student course withdrawals as either academically vs. non-academically related. To further explore the results, an independent PCA was carried out using the spring 2011 data.

Using the same process employed to identify a reasonable number of components in the fall 2010 data (i.e. scree plot examination and parallel analysis); five principal components were extracted using the spring 2011 data. The five components explain 55.31% of the total variation. The first component includes the nodes that were labeled Job-Work (0.71) and Time-Schedule (0.64). The second component includes Faculty Negative (0.76) and Course Negative (0.60). The third component includes Family (0.68) and Health (0.64). The fourth component includes Info Technology (0.70) and Online Course (0.68). Finally, the fifth component includes the node categories Federal Service (0.72) and Financial (0.59). Figure 6 contains two-dimensional views of the first three components and Table 8 contains the rotated component matrix showing the component loading values. The component loadings were consistent with those observed in the fall 2010 analysis with the the following component categories making particular sense: (1) Job-Work and Time-Schedule, (2) Faculty Negative and Course Negative, (3) Family and Health, (4) Info Technology and Online Course.

Based upon similarities in the PCA results for fall and spring terms individually, a PCA was performed using data from both terms combined. Similar to the individual term analyses both scree plot examination and PA were used to identify four principal components for extraction. Table 9 contains a view of the total variance explained by the PCA of data from both terms combined (n = 1,295). As shown the four components extracted explain approximately 45% of the total variance in the data set. This table also contains a column for the values obtained using PA. Table 10 contains the rotated component matrix for both terms combined. As shown, the four components extracted agree with those from the individual term analyses and include (1) Course Negative (0.68), Info Technology (0.56), and Online Course (0.51); (2) Job-Work (0.69) and Time-Schedule (0.60); (3) Faculty Negative (0.46), Federal Service (0.44), and Financial (0.23); and (4) Family (0.69) and Health (0.58). Figure 7 contains a labeled two-dimensional component view of both terms combined. Based on the component loadings of the node categories Course Negative, Info Technology, and Online Course, the first component was labeled "Instructional Delivery" to represent the close relationships among the comments contained in these categories. Similarly component 2 was labeled "Student Personal" reflecting the presence of its contents (Job-Work and Time-Schedule). Together these first two components represent 24% of the total variation. The third category which is composed of the node categories Faculty Negative, Federal Service, and Financial, is more difficult to interpret and is, therefore, and not labeled. The fourth component, which is composed of the Family and Health categories, was similarly not labeled (although it makes intuitive sense and corresponds well with the results obtained from the individual term analyses). To further better understand and classify the results a cluster analysis was performed and the results are described next.

Cluster Analysis

Cluster analysis has been used in the area of text mining research. Larsen and Aone (1999) described an unsupervised, near-linear time text clustering system for large-scale topic discovery from text. Their approach involves two main phases which include (1) feature extraction to map each document or record to a point in high-dimensional space and then (2) the use of clustering algorithms to automatically group the points into a hierarchy of clusters.

In the present study Hierarchical Agglomerative Cluster Analysis was used to analyze data from both terms combined. This method, also referred to as Hierarchical Cluster Analysis (HCA) or more simply "cluster analysis" is a multivariate technique commonly used in the social sciences for the purpose of classification (Bartholomew, Steele, & Moustaki, 2008). HCA is primarily used as an exploratory technique to reveal natural groupings (or clusters) within a data set. The objective of HCA is to identify relatively homogeneous groups of variables (or cases) based on selected characteristics. The procedure uses an algorithm that starts with each variable in a separate cluster and then combines clusters until only one is left. In the present study, relationships between the eleven model categories were explored using the median linkage clustering method based on chi-squared counts of category records. Figure 8 depicts the dendrogram produced by the HCA. As shown by the dendrogram the clustering of the node categories corresponds very closely to relative counts of records coded into each category in the overall model. The case processing summary for the HCA is shown in Table 11 and the the agglomeration schedule is shown in Table 12. The HCA results suggest several major cluster groups including job-work, time-schedule, and personal-other as well as course-negative, faculty-negative, and online course. Next the degree of homogeneity of the relationships in the combined data from both terms was further explored using multiple correspondence analysis.

Multiple Correspondence Analysis

To complement and further explain the results, Multiple Correspondence Analysis (MCORA) was used. As an extension of Correspondence Analysis (CORA) which is commonly used as an exploratory technique to analyze cross-classifications of two or more categorical variables in multi-way frequency tables, an aim of MCORA is to transform a table of numbers into a plot of points in a small number of—usually two—dimensions (Bartholomew et al., 2008). As such, MCORA (also called homogeneity analysis) is a technique that can be used to find optimal categorical quantifications by separating categories from each other as much as possible. This implies that objects in the same category are plotted close to each other and objects in different categories are plotted as far apart as possible. The term homogeneity also refers to the fact that the analysis will be most successful when the variables are homogeneous; that is, when they partition the objects into clusters with the same or similar categories. For each variable, a discrimination measure, which can be regarded as a squared component loading, is computed for each dimension. This measure is also the variance of the quantified variable in that dimension. It has a maximum value of 1, which is achieved if the object scores fall into mutually exclusive groups and all object scores within a category are identical.

In the present study, data from both terms combined were investigated. Figure 9 contains a two-dimension plot of MCORA discrimination measures. As shown the results correspond to prior analyses suggesting close relationships among several node categories including job-work and time-schedule, as well as information technology, online course, and course negative. Finally, of the multivariate techniques applied, MCORA was most effective in discriminating negative course versus negative faculty categorizations.

Discussion

The text mining model developed seems reasonable and finds general support in the prior empirical work discussed in the literature. The validity of the categories that emerged in the text mining model (time-schedule, personal-other, job-work, family, etc.) is generally supported. For example, Friedlander (1981) lists the seven most frequently cited reasons for student course withdrawal in descending order to be (1) job conflict, (2) inadequate preparation for the course, (3) dislike of the class, (4) assignments too heavy, (5) indefinite motivation, (6) illness, and (7) dislike of the instruction. Other reasons often given include transportation problems, personal or family illness, and change in plans. Lunnenborg (1974) includes disappointment with (1) instructor, (2) class, (3) grade/grading system, (4) course load, (5) time-schedule conflict with other activities, and (6) personal/health/family. Based on survey data from the University of North Carolina at Greensboro, Wiley (2009) reported the two most common reasons for student course withdrawal to be (1) medical issues, and (2) work. Based upon a factor analysis of a 15-item questionnaire, Dunwoody and Frank (1995) identified two reasons why students withdraw from classes to involve (1) personal considerations, and (2) course considerations. These findings support several text miner model categories including job-work, course-negative, health, faculty-negative, and personal-other.

Support for the non-academically related categories was also found. In the present study these include withdrawal reasons related to health, family, job-work, time-schedule, financial, and personal-other. For example, in acknowledging the work of Tinto (1993), Charlton, Barrow, and Hornby-Atkinson

(2006) suggest student levels of responsibility associated with age, maturity, marital status, and general family commitments to play a role in student withdrawal:

> Older students are likely to differ from younger students in a number of respects. They are more likely to be married, have children and be based at home, and will therefore typically have more demands on their time resulting in lesser social integration with other students, greater problems in obtaining academic support, and less study time. If commitment to their studies is low, these external pressures can make them particularly prone to withdraw (p. 35).

With a general older student population served by the college these results make sense and are further demonstrated by the quantitative analysis results obtained as described above. Nevertheless, the following section mentions several considerations and limitations.

Limitations

As an exploratory text mining analysis, the results presented here are best viewed as emerging or developing rather than conclusive or definitive. Ideally the results from this institutional case study would be replicated using the same methodological approach at other institutions for comparative purposes. Another consideration is the use of the specific software used. One of the strengths of the miner application used is its flexibility (e.g., enabling modification to its terms, templates, libraries, linguistic resources, text analysis packages and so on). Some may also consider this a weakness in the sense that two researchers could analyze the same text data and arrive at quite different results. The application's manual acknowledges the non-exact and iterative nature of the text mining process. On the other hand, a segment of the withdrawal literature suggests that students tend to withdraw from classes for the same reasons regardless of where, or at which, institution they happen to be. While there may be differences based on certain aspects of the institution, (e.g., public vs. private, large vs. small, urban vs. rural, *etc*.), it seems reasonable to expect that there enough commonality exists between institutions of a certain type (e.g., state colleges in Florida) to enable collaboration leading to a set of common text analysis resources that would allow for data and result sharing across institutional boundaries. The idea of such expanded partnerships and data sharing was in fact a key area of focus at the 50^{th} Annual *Association for Institutional Research* (AIR) Meeting held in Chicago and the work of an esteemed and dedicated group there also resulted in a white paper on the topic that is available on the AIR website (Association for Institutional Research, 2010). More work needs to be done and more results need to be shared.

Conclusions and Recommendations

This study has sought to contribute to a more detailed understanding of student rationale(s) for college course withdrawal and in so doing suggest actions that can be taken by institutions to assist. While there are currently no perfect "automatic" methods to accurately categorize or classify extremely large sets of lengthy and/or detailed written comments provided by students as a reflection of their academic and personal life, the present study represents an organized exploration of (at least the possibility of developing) such. However, is should also be obvious that much still needs to be done in terms of both methodological/analytical refinement and the formulation and implementation of institutional action plans to mitigate excessive course withdrawal. Potential solutions to the latter abound and may involve straightforward interventions such as course redesign (see, e.g., Decreasing Costs and Increasing Student Outcomes: Course Redesign in Maryland, in United States Department of Education, March 2011, p. 21).

In considering an expanded role and application of text analysis, careful attention should be paid to establishing the goal(s) of the analysis and then defining exact criteria used to develop the mining model to reach the goal(s). Especially given the combined and compounded complexity involving both the nuanced interpretation of human language and the technical learning curve associated with the varied and expanding field of text analytics and mining, this is no small task. When the amount of text is relatively small, it is easy to simply read the text and assume (or at least hope for) accurate interpretation and even solid understanding. At very small and perhaps highly specialized institutions in which the number of student withdrawals in any given term is small, perhaps text mining is not needed. At such institutions, those interested in the reasons that students withdraw from courses can read the comments (or simply converse with the actual student). However, such an approach is clearly not practical at very large institutions particularly given the new reality of ever shrinking resources. Nevertheless, it is reasonable to find (and expect to continue to find) many honest, eloquently worded, and thoroughly explained, reasons given by students for course withdrawal. The institution may be in a position to do something about some of the reasons, but it is not in such a position to take action to avert many others. The possibility of finding improved ways to analyze and summarize the comments of withdrawing students, however, offers hope as a means to support and improve the effectiveness of the institution's service to its students especially as much more progress is made.

References

Association for Institutional Research (AIR) Annual Forum. (2010). *Expanding our Partnerships, Consortia and Data Sharing*. Association for Institutional Research (AIR) Annual Forum. Chicago: Association for Institutional Research (AIR). http://www.airweb.org/images/whitepaper2010/WhitePaperDiscussion_2.pdf

Astin, A. (1997). *What matters in college?* San Francisco: Jossey-Bass.

Bambara, C., Harbour, C., Davies, T., & Athey, S. (2009). The Lived Experience of Community College Students Enrolled in High-Risk Online Courses. *Community College Review, 36*(3), 219-238.

Bartholomew, D. J., Steele, F., & Moustaki, I. (2008). *Analysis of multivariate social science data.* Boca Raton, FL: CRC Press.

Berry, M., & Kogan, J. (2010). *Text mining: Applications and theory.* (M. Berry, & J. Kogan, Eds.) Chichester, West Sussex, UK, United Kingdom: John Wiley & Sons, Ltd.

Buglear, J. (2009, Nov.). Logging in and dropping out: exploring student non-completion in higher education using electronic footprint analysis. *Journal of Further & Higher Education, 33*(4), 381-393.

Charlton, J., Barrow, C., & Hornby-Atkinson, P. (2006). Attempting to predict withdrawal from higher education using demographic, psychological and educational measures. *Research in Post-Compulsory Education, 11*(1), 31-47.

Conklin, K. (1997). Course attrition: A 5-year perspective on why students drop classes. *Community College Journal of Research & Practice, 21*(8), 753-759. Retrieved from Education Research Complete database.

Denzin, N. K., & Lincoln, Y. S. (2005). Introduction: The discipline and practice of qualitative research. In N. K. Lincoln & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research, 3rd* (pp. 1-32). Thousand Oaks, CA: Sage Publications, Inc.

Dunwoody, P., & Frank, M. L. (1995, September). Why students withdraw from classes. *The Journal of Psychology, 129*(5), 553-558.

Ettinger, L. F., & Maitland-Gholson, J. (1990). Text analysis as a guide for research in art education. *Studies in Art Education*, 31(2), 86-98.

Feldman, R. & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. New York: Cambridge University Press.

Florida Department of Education. (2011, March). *Commonly withdrawn courses*. Division of Florida Colleges. March 2011-03 edition. Tallahassee, FL.

Friedlander, J. (1981, Jan.). *Why students drop courses.* ERIC Clearinghouse for Junior Colleges. Los Angeles: Junior College Resource Review. (ERIC Document Reproduction Service No. ED196505). from ERIC database.

Geertz, C. (1983). *Local knowledge*. New York: Basic Books.

Hagedorn, L., Maxwell, W., Cypers, S., Moon, H., & Lester, J. (2003). *Course-shopping in the urban community colleges: An analysis of student drop and add activities.* University of Southern California. Los Angeles: Center for Higher Education Policy Analysis (ERIC Document Reproduction Service No. ED477196). from ERIC database.

Hall, M., Smith, K., Boeckman, D., Ramachandra, V., & Jasin, J. (2003). Why do students withdraw from courses. *Paper Presented at the 2003 Southern Association for Institutional Research* (pp. 1-10). San Antonio, TX: SAIR.

Larsen, B. & Aone, C. (1999). Fast and effective text mining using linear-time document clustering. Paper presented at the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 16-22). San Diego, CA.

Lunneborg, C., Lunneborg, P., & de Wolf, V. (1974, Feb.). *Why do UW students withdraw from classes?* University of Washington. Seattle, WA: Educational Assessment Center (ERIC Document Reproduction Service No. ED131777). from ERIC database.

Mery, P. (2001). *Students leaving before census.* City College of San Francisco, Office of Institutional Development, Research, and Planning. San Francisco, CA: Educational Assessment Center (ERIC Document Reproduction Service No. ED455867). from ERIC database.

Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis, (2nd). Thousand Oaks, CA: Sage Publications, Inc.

Nasukawa, T., & Nagano, T. (2001). Text analysis and knowledge mining system. *IBM Systems Journal*, 40(4), 976-984.

Nisbet, R., Elder, J., Elder, J. F., & Miner, G. (2009). *Handbook of statistical analysis and data mining applications.* Amsterdam: Academic Press/Elsevier.

Reed, J. G. (1981, June). Dropping a college course: Factors influencing students' withdrawal. *Journal of Educational Psychology, 73*(3), 376-385.

Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. (2011). *Handbook of educational data mining*. Boca Raton, FL: CRC Press.

Shmueli, G., Patel, N., & Bruce, P. (2010*). Data mining for business intelligence*. Hoboken, NJ: John Wiley & Sons, Inc.

SPSS Inc. (2009). *SPSS Text Analysis for Surveys 3.0.* Chicago: SPSS® Inc.

Stanford University. (2008). *Using SPSS text analysis for surveys.* Social Science Data and Software. Palo Alto, CA: Stanford University.

Sumner, P. (2000). *JCC drop study/Attrition rates, Fall 1999.* Johnson County Community College, Office of Institutional Research. Overland Park, KS: Educational Assessment Center (ERIC Document Reproduction Service No. ED444666). from ERIC database.

Sumner, P. (2001). *Drop study/Attrition rates, Fall 2000.* Johnson County Community College. Overland Park, KS: Educational Assessment Center (ERIC Document Reproduction Service No. ED462993). from ERIC database.

Tinto, V. (1987a). The principles of effective retention. (pp. 1-18). Largo, MD: Paper presented at the Fall Conference of the Maryland College Personnel Association. (ERIC Document Reproduction Service No. ED301267) from ERIC database.

Tinto, V. (1987b). *Leaving college: Rethinking the causes and cures of student attrition*. Chicago: University of Chicago Press.

Tinto, V. (1993) Leaving college: rethinking the causes and cures of student attrition (2nd). Chicago: University of Chicago Press.

Tinto, V. (2006). Research and practice of student retention: What Next? *Journal College Student Retention, 8*(1), 1-19.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind* 49, 433-460.

United States Department of Education. (2011, March). *College completion tool kit*. Accession Number: ED517672. Washington, D.C.

Watkins, M. W. (2006). Determining Parallel Analysis Criteria. Journal of Modern Applied Statistical Methods, 5, 344-346.

Wiley. (2009, April). Withdrawal data helps solve retention issues. *Successful Registrar,9* (2), 6-7.

Zhang, Q., & Segall, R. (2010). Review of data, text and web mining software. *Kybernetes, 39*(4), 625-655.

Table 1

Category Web Table Fall 2010 Term

Category 1		Category 2		Shared Records (Both Categories)
Node Title	Record Count	Node Title	Record Count	
Personal-Other	301	Time-Schedule	331	151
Job-Work	146	Time-Schedule	331	79
Job-Work	146	Personal-Other	301	38
Course Negative	43	Time-Schedule	331	32
Financial	43	Time-Schedule	331	28
Time-Schedule	331	Family	54	21
Online Course	34	Time-Schedule	331	20
Course Negative	43	Personal-Other	301	16
Personal-Other	301	Family	54	16
Time-Schedule	331	Faculty Negative	48	16
Time-Schedule	331	Health	30	14
Financial	43	Personal-Other	301	13
Info Technology	14	Time-Schedule	331	11
Online Course	34	Personal-Other	301	11
Personal-Other	301	Health	30	10
Faculty Negative	48	Course Negative	43	8
Federal Service	11	Time-Schedule	331	7
Job-Work	146	Course Negative	43	7
Course Negative	43	Financial	43	6
Faculty Negative	48	Info Technology	14	6
Faculty Negative	48	Financial	43	6
Info Technology	14	Personal-Other	301	6
Online Course	34	Course Negative	43	6
Info Technology	14	Course Negative	43	5
Job-Work	146	Financial	43	5
Online Course	34	Job-Work	146	5
Faculty Negative	48	Personal-Other	301	4
Info Technology	14	Financial	43	4
Job-Work	146	Family	54	4
Job-Work	146	Info Technology	14	4
Job-Work	146	Faculty Negative	48	4
Online Course	34	Info Technology	14	3
Online Course	34	Financial	43	3
Family	54	Health	30	2
Financial	43	Family	54	2
Job-Work	146	Health	30	2
Family	54	Course Negative	43	1
Financial	43	Health	30	1
Info Technology	14	Health	30	1
Job-Work	146	Federal Service	11	1
Online Course	34	Federal Service	11	1

Table 2

Comparison of Text Mining Sample Counts, Ranks, and Proportions to Full Fall 2010 Term by Course

Withdrawals Text Analysis Sample (n = 616)					Withdrawals Fall 2010 Full Term (n = 84,083)				
Course	Number	Rank	%	Cumulative %	Course	Number	Rank	%	Cumulative %
MAT1033	29	1	4.71%	4.71%	MAC1105	324	1	6.94%	6.94%
MAC1105	28	2	4.55%	9.25%	MAT1033	208	2	4.45%	11.39%
ENC1101	21	3	3.41%	12.66%	ENC1101	204	3	4.37%	15.76%
BSC2085C	18	4	2.92%	15.58%	BSC2085C	178	4	3.81%	19.57%
ENC1102	18	5	2.92%	18.51%	MAT0024	172	5	3.68%	23.25%
MAT0024	15	6	2.44%	20.94%	ENC1102	147	6	3.15%	26.40%

*Pearson's r = 0.838, p < .01 (± .917 critical value .01, two-tail)

Table 3
Category Web Table Spring 2011

Category 1		Category 2		Shared Records (Both Categories)
Node Title	Record Count	Node Title	Record Count	
Personal-Other	364	Time-Schedule	349	171
Job-Work	155	Time-Schedule	349	99
Job-Work	155	Personal-Other	364	61
Course Negative	58	Time-Schedule	349	42
Course Negative	58	Personal-Other	364	30
Online Course	48	Time-Schedule	349	28
Online Course	48	Personal-Other	364	24
Financial	52	Personal-Other	364	23
Time-Schedule	349	Family	64	23
Financial	52	Time-Schedule	349	21
Personal-Other	364	Family	64	21
Job-Work	155	Family	64	19
Personal-Other	364	Faculty Negative	47	19
Faculty Negative	47	Time-Schedule	349	16
Course Negative	58	Online Course	48	14
Course Negative	58	Faculty Negative	47	13
Job-Work	155	Financial	52	11
Time-Schedule	349	Health	29	10
Online Course	48	Job-Work	155	9
Personal-Other	364	Health	29	8
Course Negative	58	Job-Work	155	7
Time-Schedule	349	Info Technology	10	7
Family	64	Health	29	6
Personal-Other	364	Info Technology	10	5
Financial	52	Family	64	4
Job-Work	155	Faculty Negative	47	4
Online Course	48	Faculty Negative	47	4
Family	64	Course Negative	58	3
Financial	52	Online Course	48	3
Financial	52	Course Negative	58	3
Job-Work	155	Info Technology	10	3
Online Course	48	Info Technology	10	3
Faculty Negative	47	Health	29	2
Family	64	Info Technology	10	2
Info Technology	10	Course Negative	58	2
Job-Work	155	Health	29	2
Online Course	48	Family	64	2
Course Negative	58	Health	29	1
Financial	52	Faculty Negative	47	1
Financial	52	Health	29	1
Financial	52	Federal Service	9	1
Online Course	48	Health	29	1
Personal-Other	364	Federal Service	9	1

Table 4

Fall 2010 Record Classification Counts by Model Node

Category Node	Time-Schedule	Personal-Other	Job-Work	Family	Course Negative	Faculty Negative	Financial	Online Course	Health	Info Tech	Federal Service
Time-Schedule	331	151	79	21	32	16	28	20	14	11	7
Personal-Other	151	301	62	34	18	9	17	16	16	8	1
Job-Work	79	62	146	10	8	5	9	7	4	4	1
Family	21	34	10	54	1	0	2	0	4	0	0
Course Negative	32	18	8	1	43	16	6	8	0	6	0
Faculty Negative	16	9	5	0	16	48	7	3	0	6	0
Financial	28	17	9	2	6	7	43	3	1	4	0
Online Course	20	16	7	0	8	3	3	34	0	4	1
Health	14	16	4	4	0	0	1	0	30	1	0
Info Tech	11	8	4	0	6	6	4	4	1	14	0
Federal Service	7	1	1	0	0	0	0	1	0	0	11

Table 5

Fall 2010 Internode Rank Correlations

Category[a]		Time-Schedule	Personal-Other	Job-Work	Family	Course Negative	Faculty Negative	Financial	Online Course	Health	Info Technology	Federal Service
Time-Schedule	Spearman's rho (ρ)	1.000	-.070	.004	-.092*	.114**	-.119**	.063	.025	-.032	.076	.027
	Sig. (2-tailed)	.	.083	.917	.022	.005	.003	.121	.541	.427	.060	.507
Personal-Other	Spearman's rho (ρ)	-.070	1.000	-.071	.087*	-.038	-.175**	-.051	-.009	.020	.025	-.107**
	Sig. (2-tailed)	.083	.	.077	.030	.342	.000	.205	.829	.616	.532	.008
Job-Work	Spearman's rho (ρ)	.004	-.071	1.000	-.038	-.033	-.091*	-.018	-.018	-.055	.017	-.046
	Sig. (2-tailed)	.917	.077	.	.349	.416	.024	.658	.661	.171	.665	.251
Family	Spearman's rho (ρ)	-.092*	.087*	-.038	1.000	-.062	-.090*	-.040	-.075	.037	-.047	-.042
	Sig. (2-tailed)	.022	.030	.349	.	.122	.025	.323	.063	.365	.241	.300
Course Negative	Spearman's rho (ρ)	.114**	-.038	-.033	-.062	1.000	.301**	.075	.157**	-.062	.215**	-.037
	Sig. (2-tailed)	.005	.342	.416	.122	.	.000	.063	.000	.124	.000	.360
Faculty Negative	Spearman's rho (ρ)	-.119**	-.175**	-.091*	-.090*	.301**	1.000	.087*	.009	-.066	.199**	-.039
	Sig. (2-tailed)	.003	.000	.024	.025	.000	.	.031	.818	.103	.000	.331
Financial	Spearman's rho (ρ)	.063	-.051	-.018	-.040	.075	.087*	1.000	.017	-.032	.129**	-.037
	Sig. (2-tailed)	.121	.205	.658	.323	.063	.031	.	.665	.422	.001	.360
Online Course	Spearman's rho (ρ)	.025	-.009	-.018	-.075	.157**	.009	.017	1.000	-.055	.154**	.021
	Sig. (2-tailed)	.541	.829	.661	.063	.000	.818	.665	.	.175	.000	.601
Health	Spearman's rho (ρ)	-.032	.020	-.055	.037	-.062	-.066	-.032	-.055	1.000	.016	-.031
	Sig. (2-tailed)	.427	.616	.171	.365	.124	.103	.422	.175	.	.690	.450
Info Technology	Spearman's rho (ρ)	.076	.025	.017	-.047	.215**	.199**	.129**	.154**	.016	1.000	-.021
	Sig. (2-tailed)	.060	.532	.665	.241	.000	.000	.001	.000	.690	.	.610
Federal Service	Spearman's rho (ρ)	.027	-.107**	-.046	-.042	-.037	-.039	-.037	.021	-.031	-.021	1.000
	Sig. (2-tailed)	.507	.008	.251	.300	.360	.331	.360	.601	.450	.610	.

a. Academic Term = fall 2010 (n=616)

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

Table 6

Internode Rank Correlations All Records (fall 2010 and spring 2011 combined)

Category[a]		Time-Schedule	Personal-Other	Job-Work	Family	Course Negative	Faculty Negative	Financial	Online Course	Health	Info Technology	Federal Service
Time-Schedule	Spearman's rho (r)	1.000	-.084**	.073**	-.097**	.121**	-.106**	-.005	.031	-.052	.062*	-.038
	Sig. (2-tailed)	.	.002	.009	.001	.000	.000	.850	.259	.063	.026	.169
Personal-Other	Spearman's rho (r)	-.084**	1.000	-.115**	-.030	-.022	-.123**	-.052	-.013	-.047	.008	-.100**
	Sig. (2-tailed)	.002	.	.000	.280	.423	.000	.061	.631	.093	.781	.000
Job-Work	Spearman's rho (r)	.073**	-.115**	1.000	.010	-.058*	-.092**	-.015	-.023	-.068*	.019	-.052
	Sig. (2-tailed)	.009	.000	.	.719	.038	.001	.600	.409	.015	.488	.062
Family	Spearman's rho (r)	-.097**	-.030	.010	1.000	-.052	-.089**	-.027	-.060*	.060*	-.004	-.039
	Sig. (2-tailed)	.001	.280	.719	.	.061	.001	.326	.030	.032	.894	.165
Course Negative	Spearman's rho (r)	.121**	-.022	-.058*	-.052	1.000	.238**	.018	.185**	-.050	.131**	-.035
	Sig. (2-tailed)	.000	.423	.038	.061	.	.000	.528	.000	.074	.000	.202
Faculty Negative	Spearman's rho (r)	-.106**	-.123**	-.092**	-.089**	.238**	1.000	.012	.012	-.033	.093**	-.034
	Sig. (2-tailed)	.000	.000	.001	.001	.000	.	.674	.667	.234	.001	.217
Financial	Spearman's rho (r)	-.005	-.052	-.015	-.027	.018	.012	1.000	.000	-.033	.049	-.010
	Sig. (2-tailed)	.850	.061	.600	.326	.528	.674	.	.995	.234	.077	.727
Online Course	Spearman's rho (r)	.031	-.013	-.023	-.060*	.185**	.012	.000	1.000	-.042	.129**	-.005
	Sig. (2-tailed)	.259	.631	.409	.030	.000	.667	.995	.	.135	.000	.847
Health	Spearman's rho (r)	-.052	-.047	-.068*	.060*	-.050	-.033	-.033	-.042	1.000	-.003	-.027
	Sig. (2-tailed)	.063	.093	.015	.032	.074	.234	.234	.135	.	.927	.338
Info Technology	Spearman's rho (r)	.062*	.008	.019	-.004	.131**	.093**	.049	.129**	-.003	1.000	-.017
	Sig. (2-tailed)	.026	.781	.488	.894	.000	.001	.077	.000	.927	.	.547
Federal Service	Spearman's rho (r)	-.038	-.100**	-.052	-.039	-.035	-.034	-.010	-.005	-.027	-.017	1.000
	Sig. (2-tailed)	.169	.000	.062	.165	.202	.217	.727	.847	.338	.547	.

a. Academic Term = fall 2010 and spring 2011 combined (n=1,295)

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 7

Principal Components Analysis Rotated Component Matrix (fall 2010)

Rotated Component Matrix[a,b]

Node	Component		
	1	2	3
Info Technology	.670	.100	-.028
Course Negative	.666	-.093	-.192
Online Course	.474	.021	.241
Financial	.350	-.063	-.010
Personal-Other	.047	.746	.174
Federal Service	-.187	-.482	.096
Family	-.209	.457	-.181
Health	-.109	.287	-.084
Faculty Negative	.402	-.278	-.709
Time-Schedule	.291	-.195	.630
Job-Work	-.057	-.187	.325

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Academic Term = fall 2010
b. Rotation converged in 5 iterations.

Table 8

Principal Components Analysis Rotated Component Matrix (spring 2011)

Rotated Component Matrix[a,b]

Node	Component				
	1	2	3	4	5
Job-Work	.707	-.211	.025	-.080	-.004
Time-Schedule	.642	.065	-.275	.118	-.278
Personal-Other	-.552	-.297	-.485	.005	-.367
Faculty Negative	-.129	.757	.033	-.130	-.039
Course Negative	.053	.604	-.135	.417	-.094
Family	.041	-.243	.679	.127	-.036
Health	-.156	.135	.635	-.120	-.165
Info Technology	-.003	-.227	.116	.703	-.042
Online Course	-.009	.206	-.082	.683	.035
Federal Service	-.130	.020	.003	.010	.717
Financial	.028	-.095	-.128	-.023	.589

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Academic Term = spring 2011
b. Rotation converged in 7 iterations.

Table 9

Principal Components Analysis for Both Terms Combined

Total Variance Explained

			Initial Eigenvalues		Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	Parallel Analysis[1]	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.486	1.1487 ***	13.512	13.512	1.486	13.512	13.512	1.404	12.761	12.761
2	1.212	1.1066 ***	11.022	24.535	1.212	11.022	24.535	1.206	10.962	23.723
3	1.136	1.0743 ***	10.329	34.863	1.136	10.329	34.863	1.176	10.689	34.412
4	1.078	1.0479 ***	9.798	44.661	1.078	9.798	44.661	1.127	10.248	44.661
5	1.020	1.0213	9.269	53.929						
6	1.002	0.9971	9.113	63.043						
7	.973	0.9745	8.844	71.886						
8	.893	0.9508	8.120	80.006						
9	.876	0.9234	7.966	87.973						
10	.742	0.8962	6.746	94.718						
11	.581	0.8593	5.282	100.000						

Extraction Method: Principal Component Analysis.

1. Randomly Generated Parallel Analysis Eigenvalues for 11 variables, n=1,295 subjects, 100 replications (Watkins, 2006) *** indicates component should be retained

Table 10

Principal Components Analysis Rotated Component Matrix for Both Terms Combined

Rotated Component Matrix[a]

Node	Component			
	1	2	3	4
Course Negative	.683	-.121	.172	-.121
Info Technology	.565	.092	.024	.114
Online Course	.512	.034	-.056	-.174
Job-Work	-.052	.695	.141	.099
Time-Schedule	.237	.602	.025	-.277
Faculty Negative	.390	-.471	.459	.037
Federal Service	-.369	-.168	.442	-.356
Financial	.046	.007	.226	-.059
Personal-Other	.032	-.254	-.811	-.196
Family	-.069	.063	-.081	.687
Health	-.042	-.133	.031	.583

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 5 iterations.

Table 11

Case Processing Summary from Hierarchical Cluster Analysis

Case Processing Summary[a]

Cases							
		Rejected					
Valid		Missing Value		Negative Value		Total	
N	Percent	N	Percent	N	Percent	N	Percent
1295	100.0%	0	.0%	0	.0%	1295	100.0%

a. Chi-square between Sets of Frequencies used

Table 12

Agglomeration Schedule from Hierarchical Cluster Analysis

Agglomeration Schedule

	Cluster Combined			Stage Cluster First Appears		
Stage	Cluster 1	Cluster 2	Coefficients	Cluster 1	Cluster 2	Next Stage
1	10	11	6.557	0	0	2
2	9	10	7.264	0	1	3
3	8	9	8.048	0	2	4
4	5	8	8.310	0	3	5
5	5	6	8.144	4	0	6
6	5	7	9.045	5	0	7
7	4	5	9.963	0	6	8
8	3	4	14.063	0	7	9
9	1	3	18.154	0	8	10
10	1	2	18.310	9	0	0

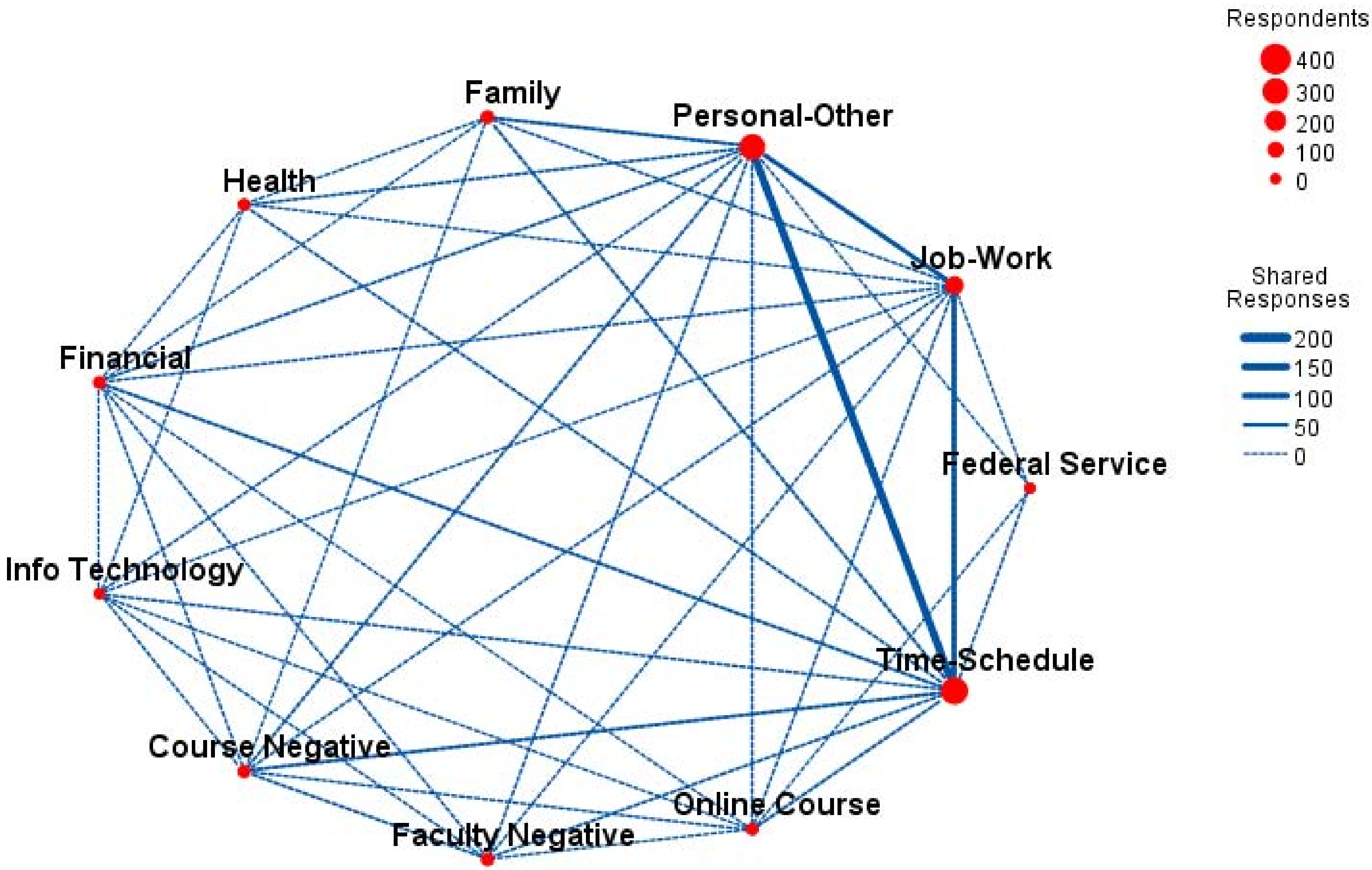

Figure 1. Category web diagram of all responses from fall 2010 data. A total of 512 responses were categorized into 11 nodes. The overall model categorized 96.1% of cases.

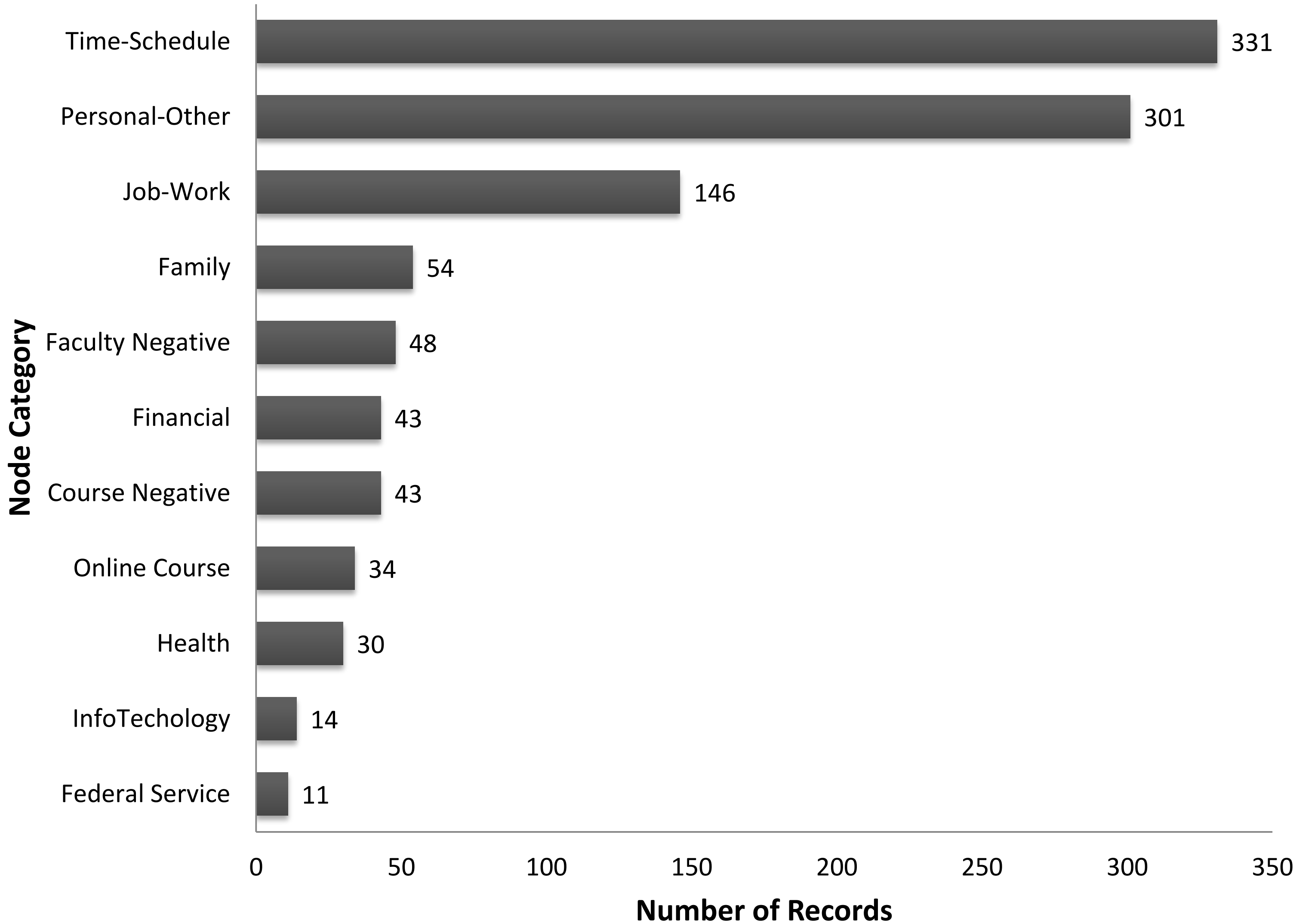

Figure 2. Category bar chart showing the number of responses coded into each development model category using fall term data.

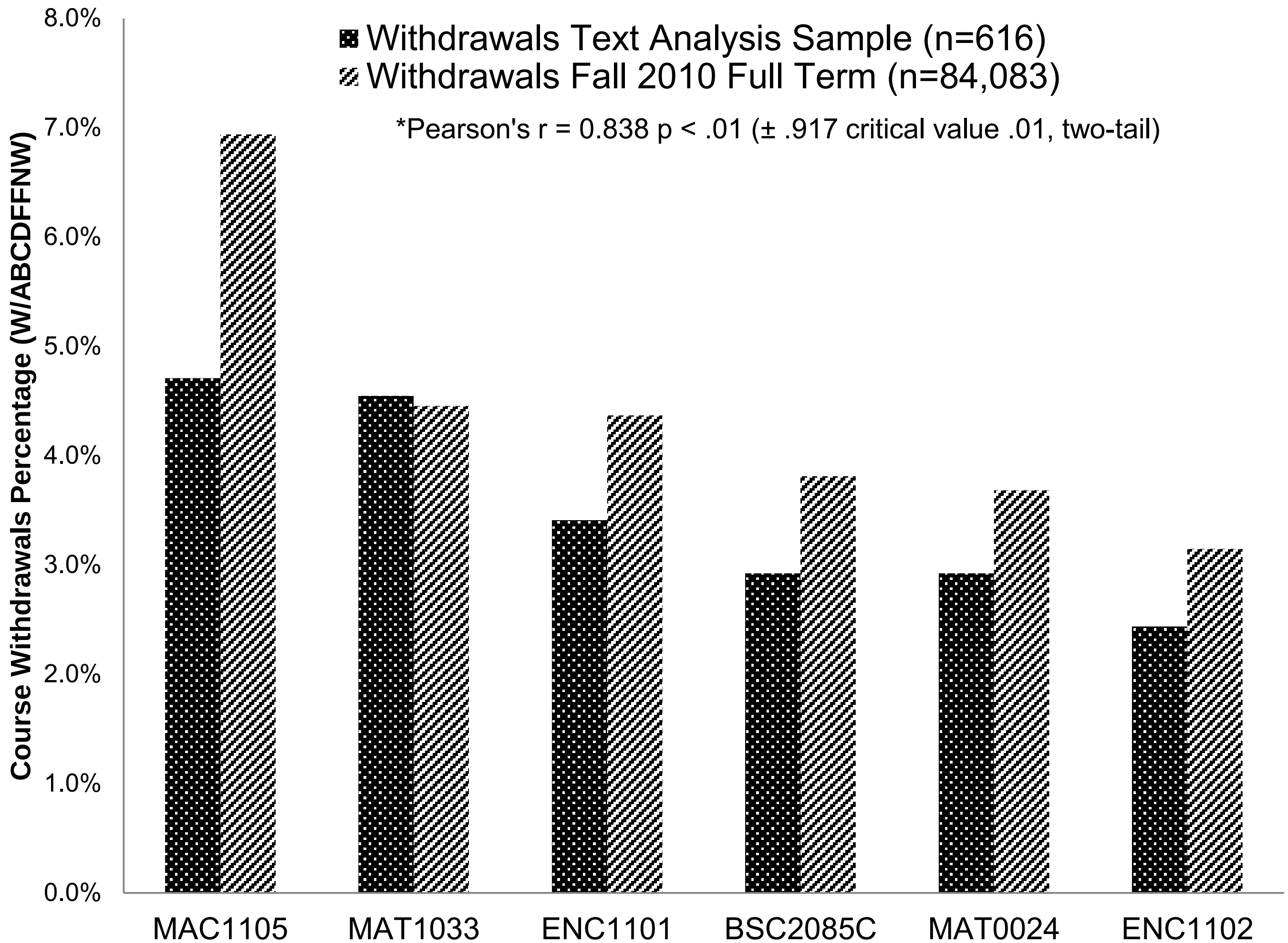

Figure 3. Withdrawal comparison of top six courses in text analysis data (n = 616) and full fall 2010 term (n = 84,083). The same subset of six courses was present in both the text analysis and full term grade set. The Pearson correlation between the two was positive and significant (0.84, p < 0.01).

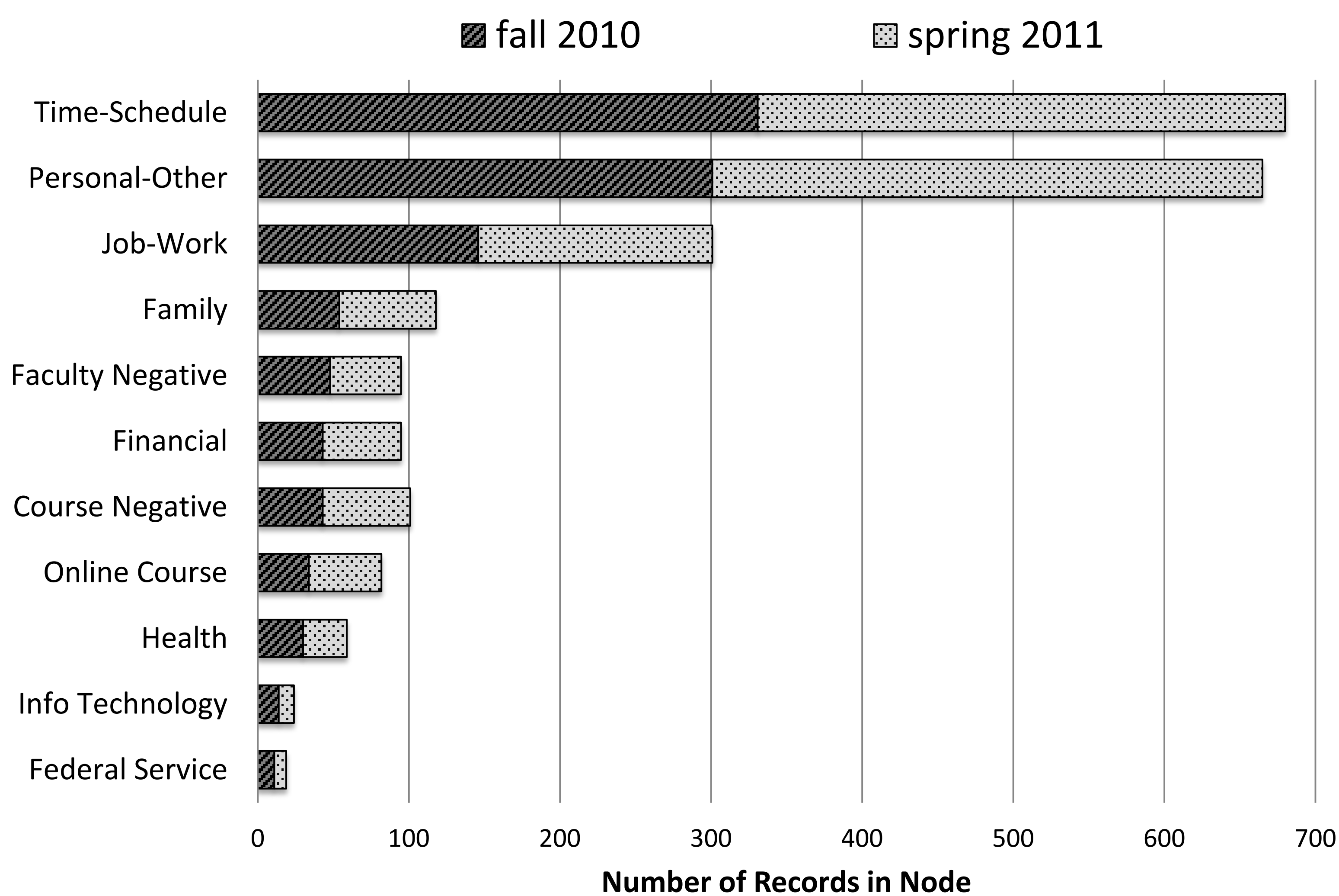

Figure 4. Category bar chart showing total number of withdrawal reason responses coded into each main category by term.

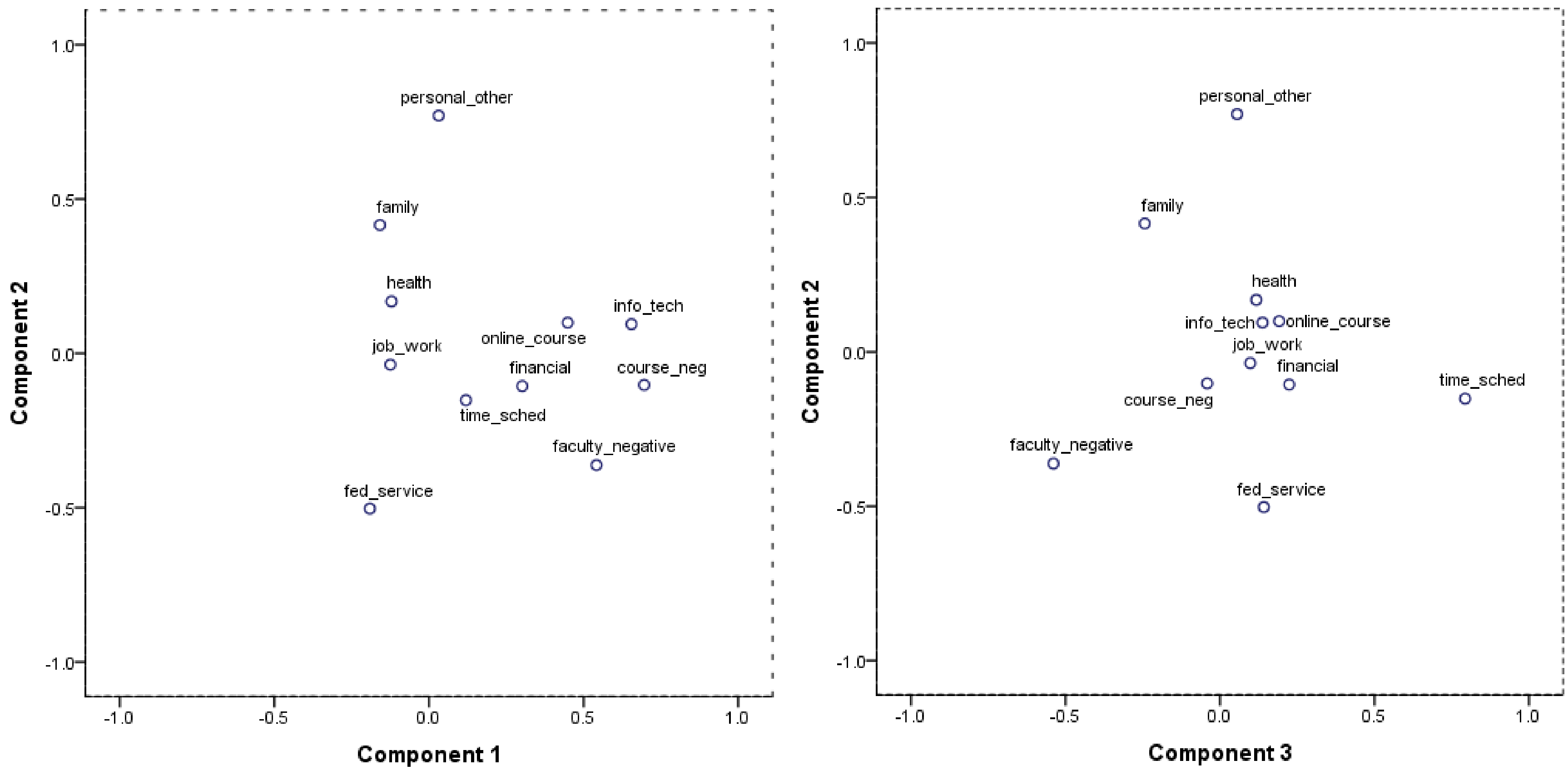

Figure 5. Fall 2010 Principal Components Analysis Rotation Views.

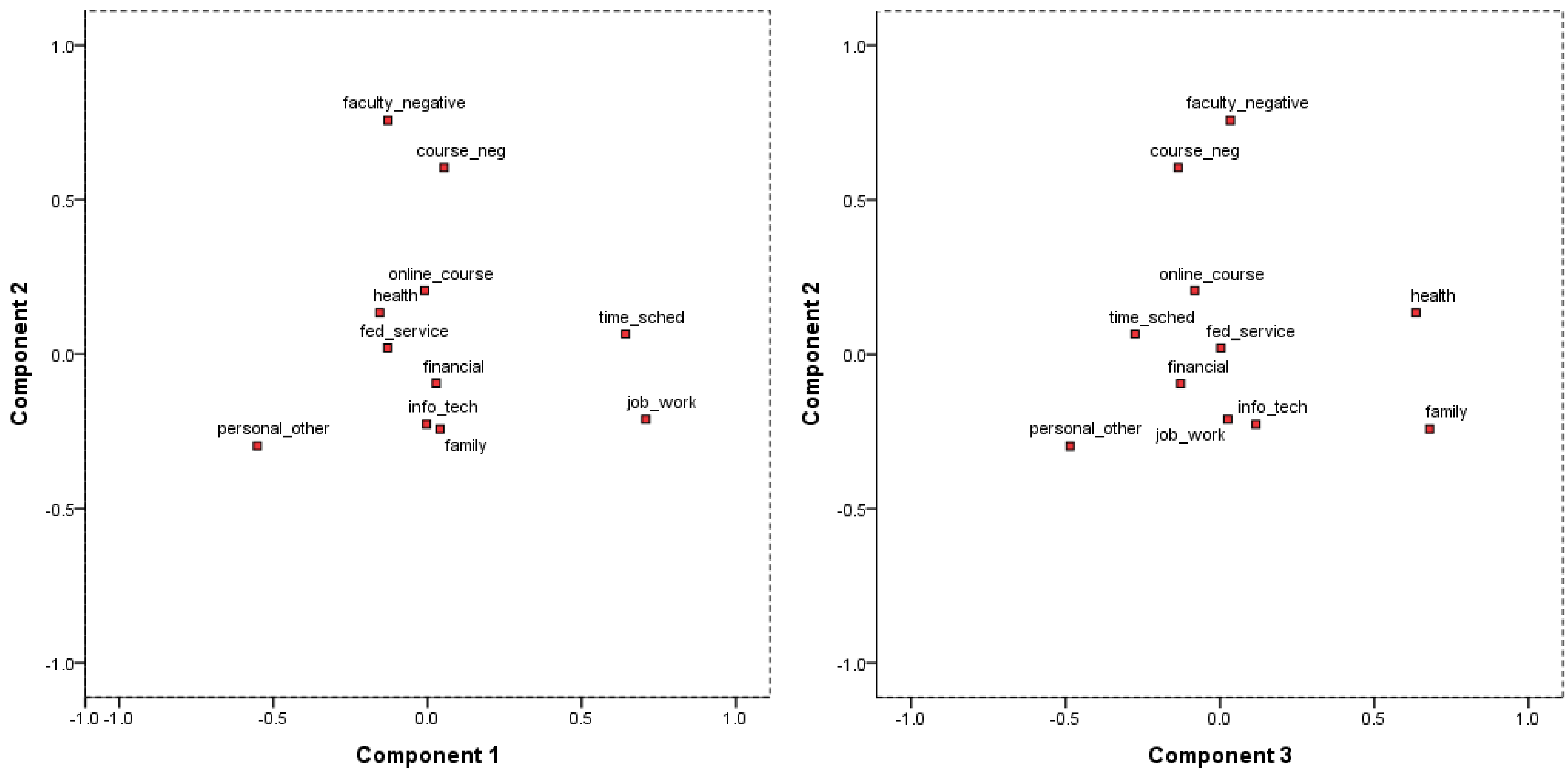

Figure 6. Spring 2011 Principal Components Analysis Rotation Views.

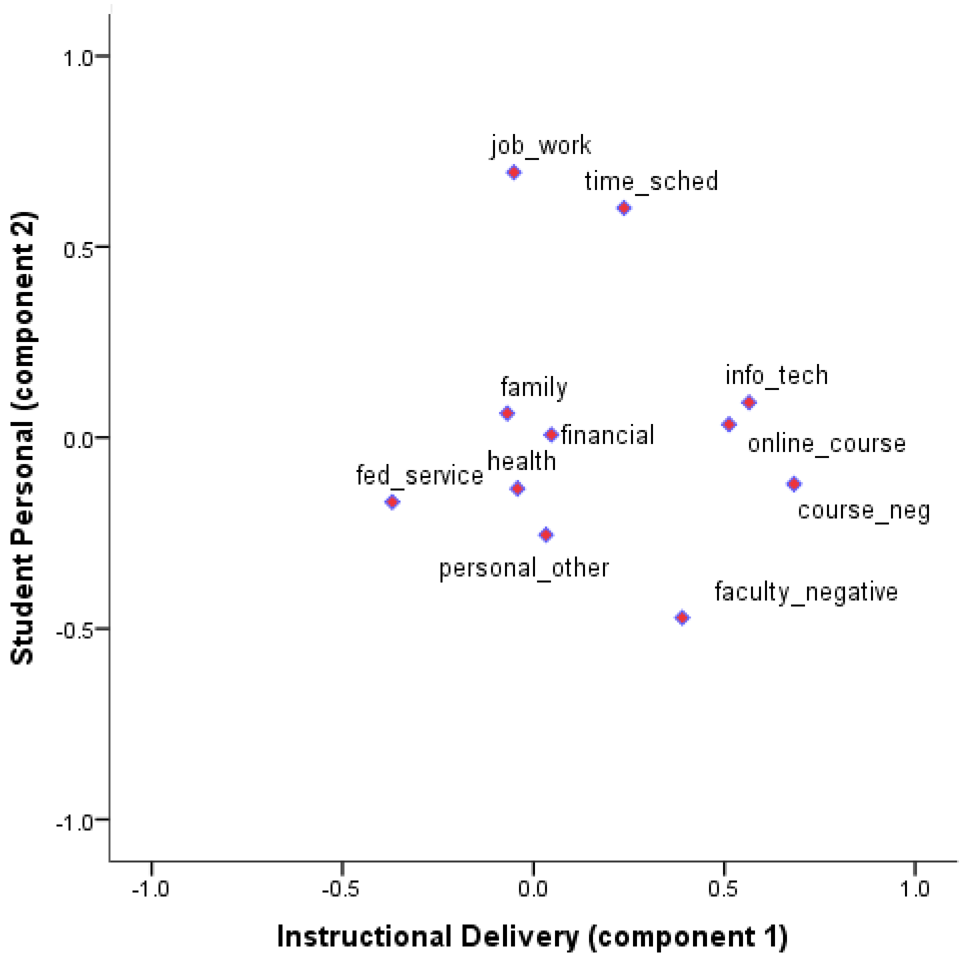

Figure 7. Labeled Component View of fall 2010 and spring 2011 Terms Combined (n = 1,295).

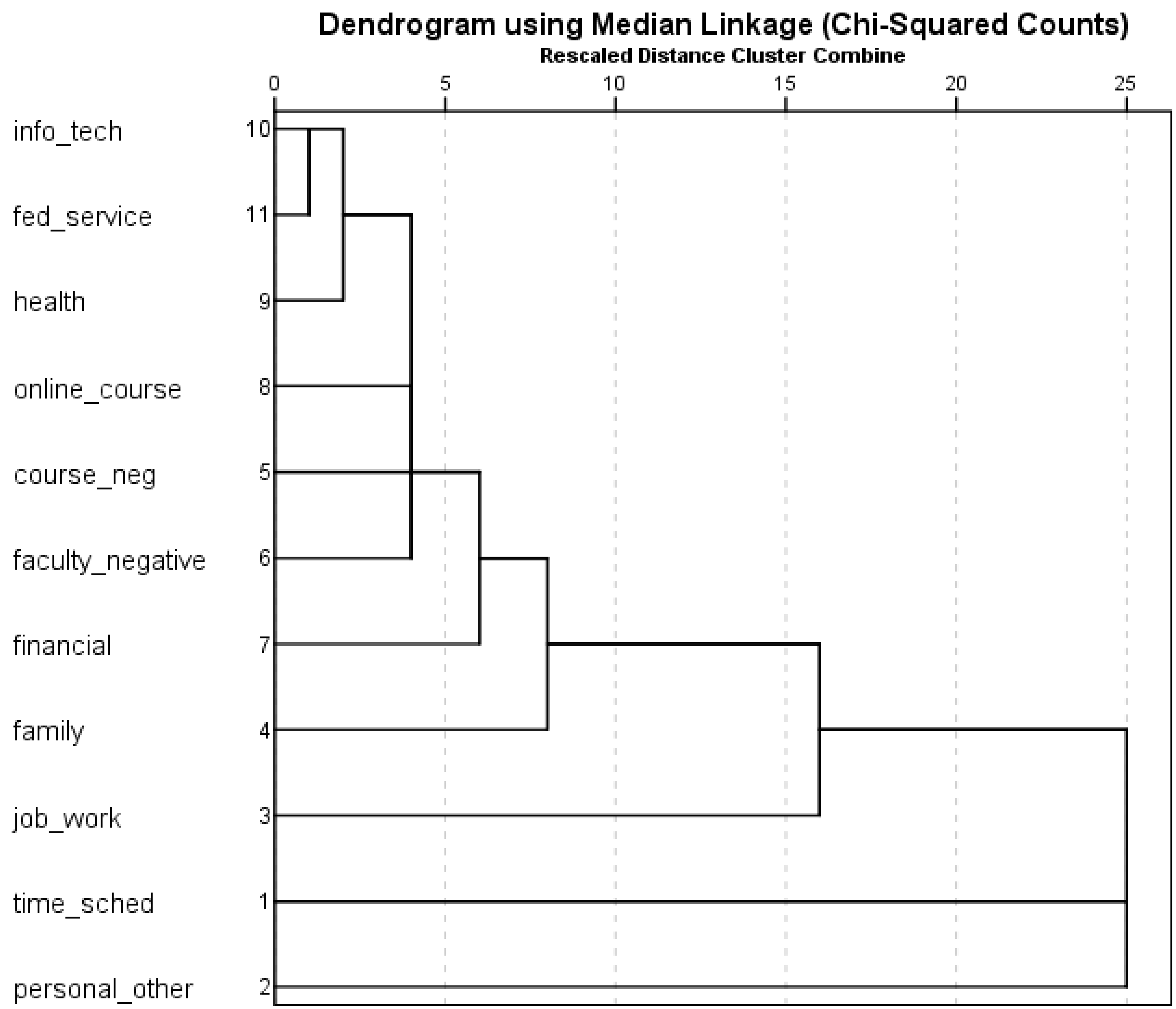

Figure 8. Dendrogram from Hierarchical Agglomerative Cluster Analysis.

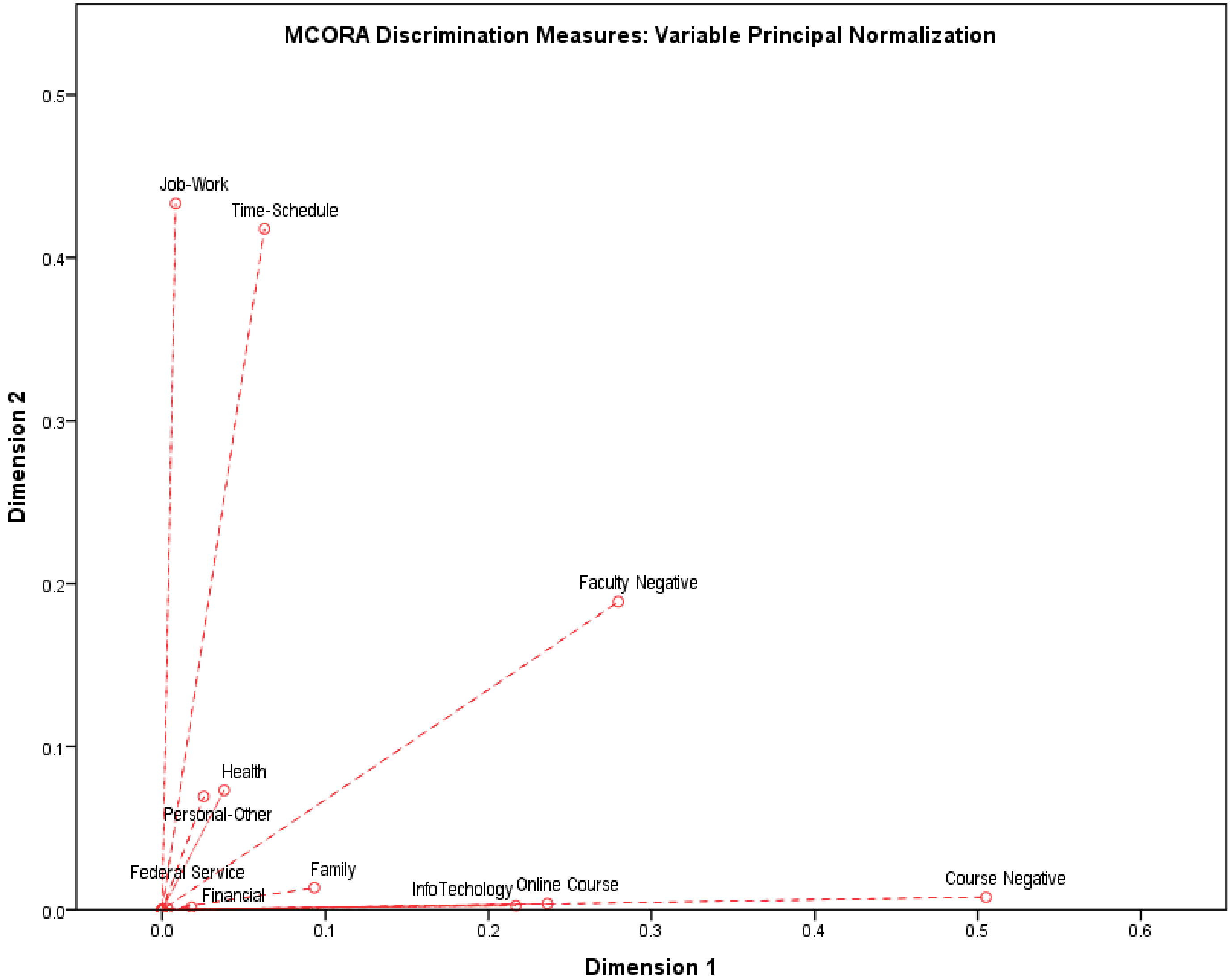

Figure 9. MCORA Discrimination Measures: Variable Principal Normalization (n = 1,295).

Appendix A: Additional Model Diagrams by Primary Node Category (fall 2010)

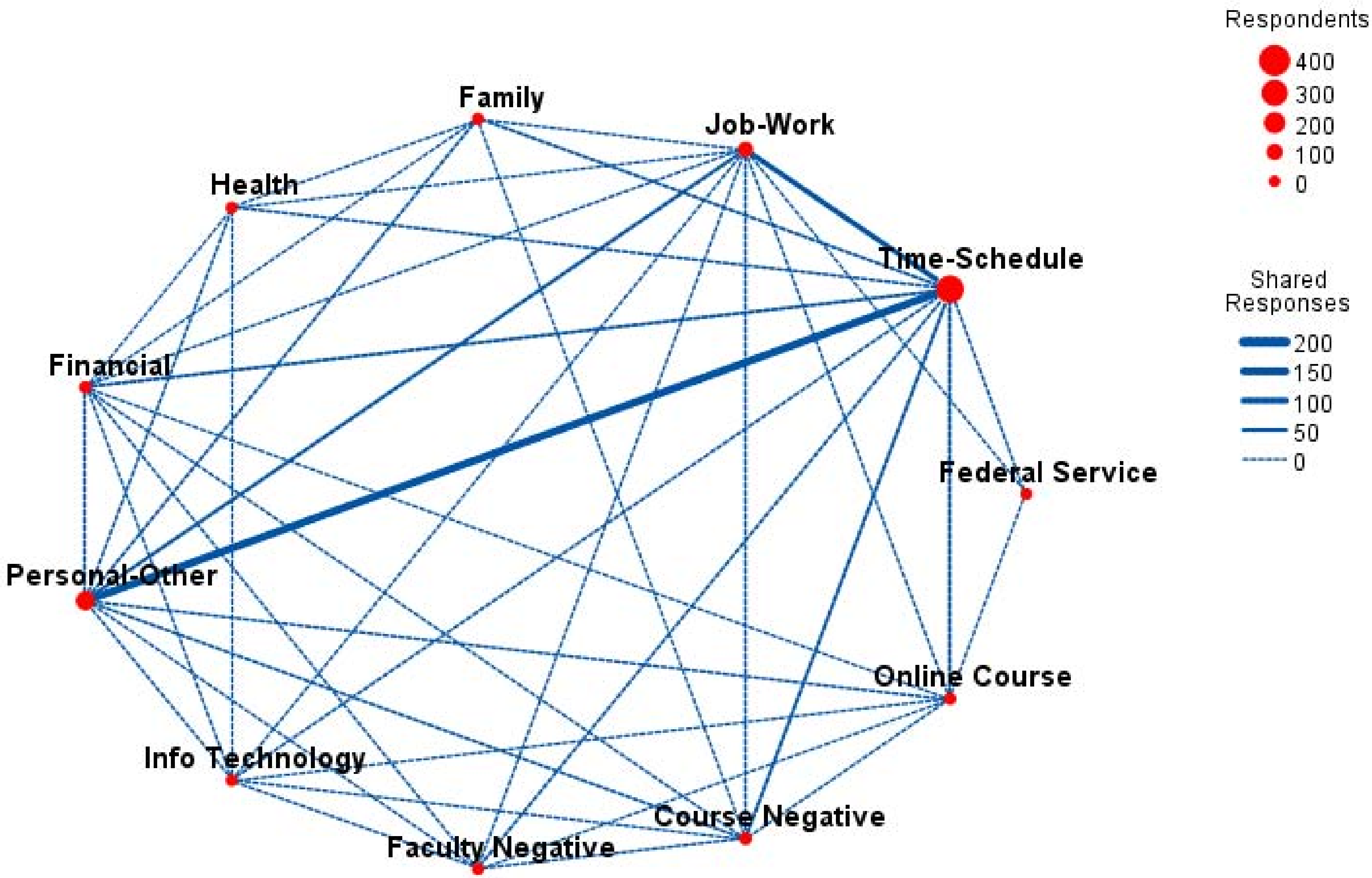

Figure A1. Category web diagram of “Time-Schedule” category (n = 331) with shared responses.

Appendix A: Additional Model Diagrams (continued)

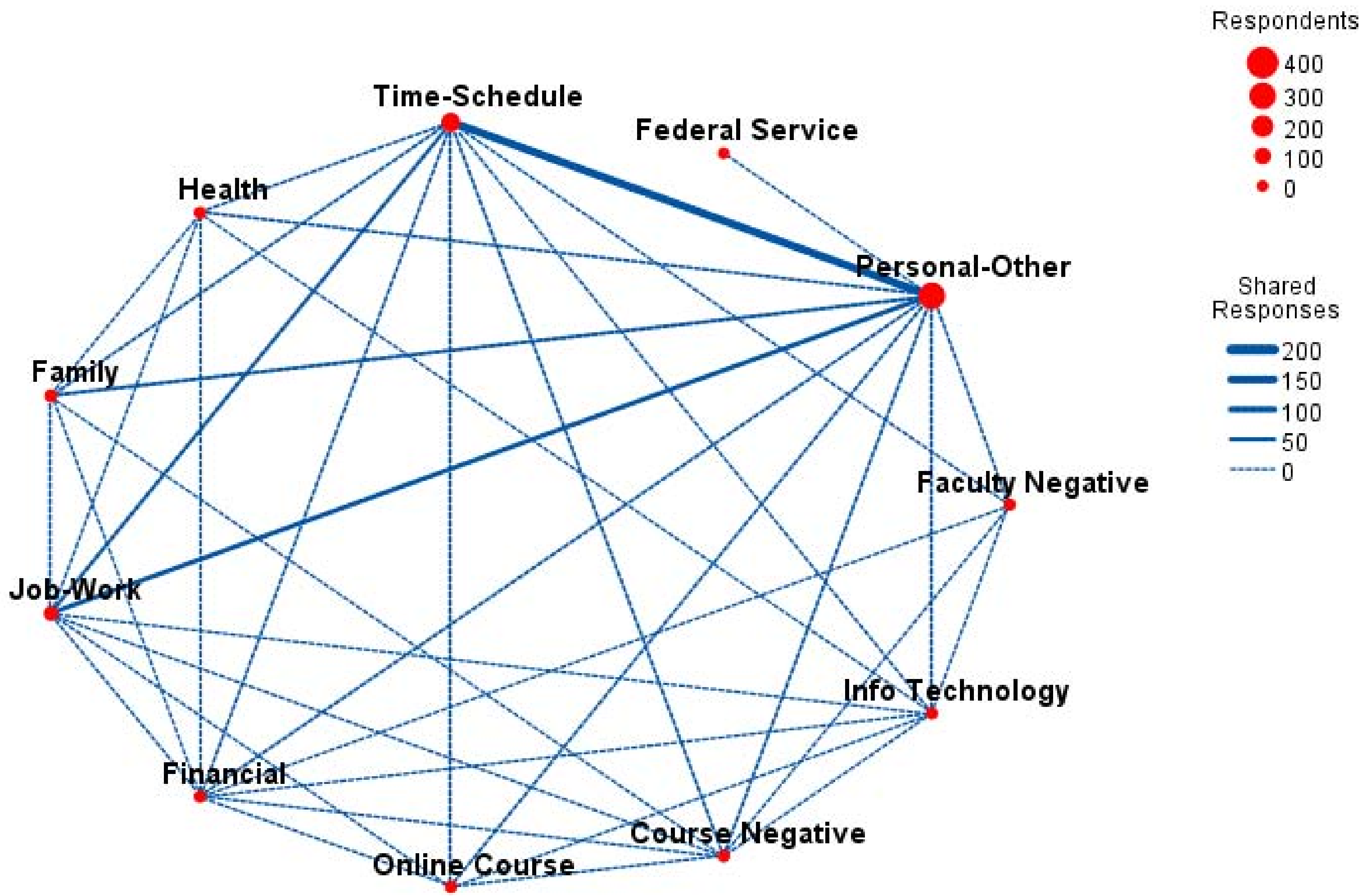

Figure A2. Category web diagram of “Personal-Other” category (n = 301) with shared responses.

Appendix A: Additional Model Diagrams (continued)

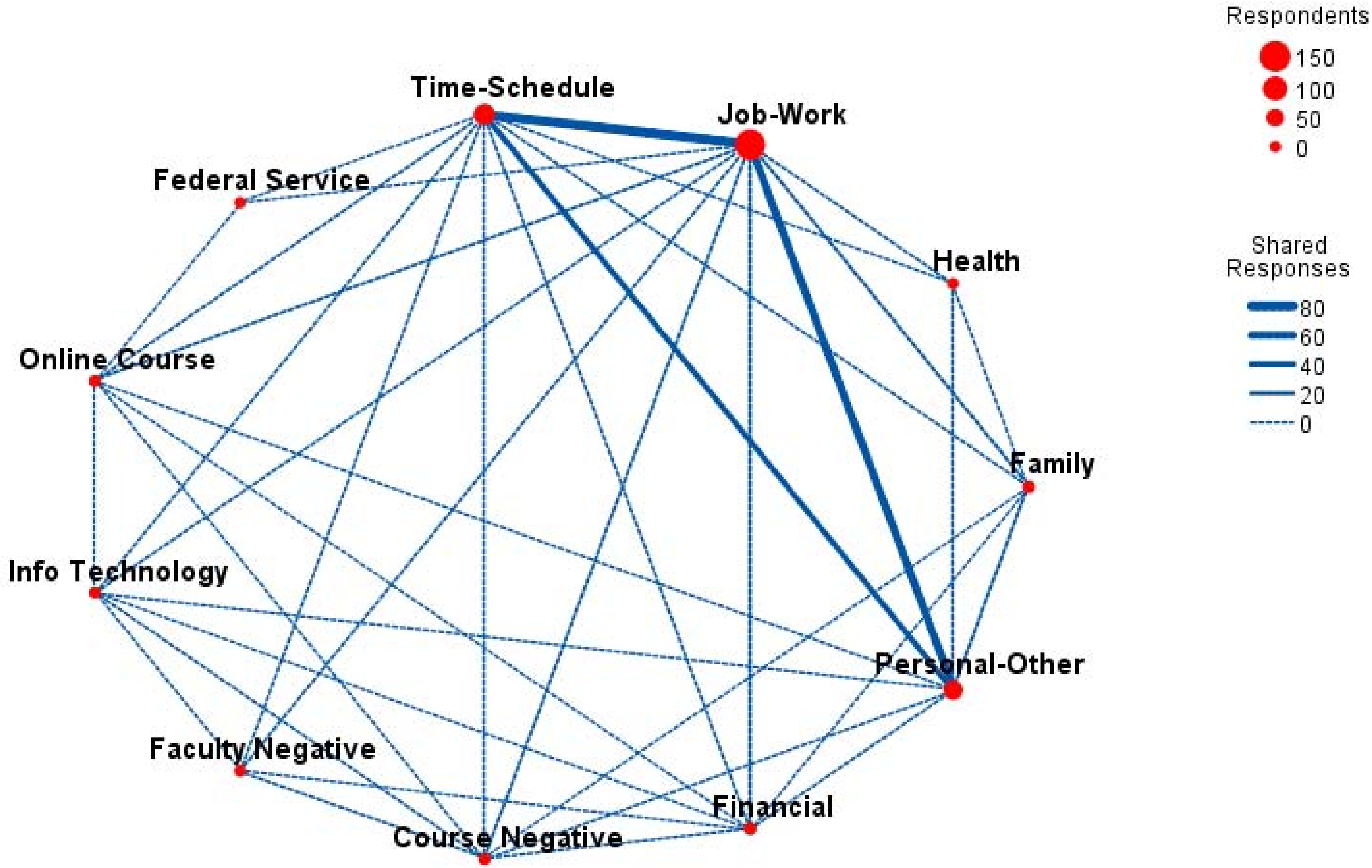

Figure A3. Category web diagram of “Job-Work” category (n = 146) with shared responses.

Appendix A: Additional Model Diagrams (continued)

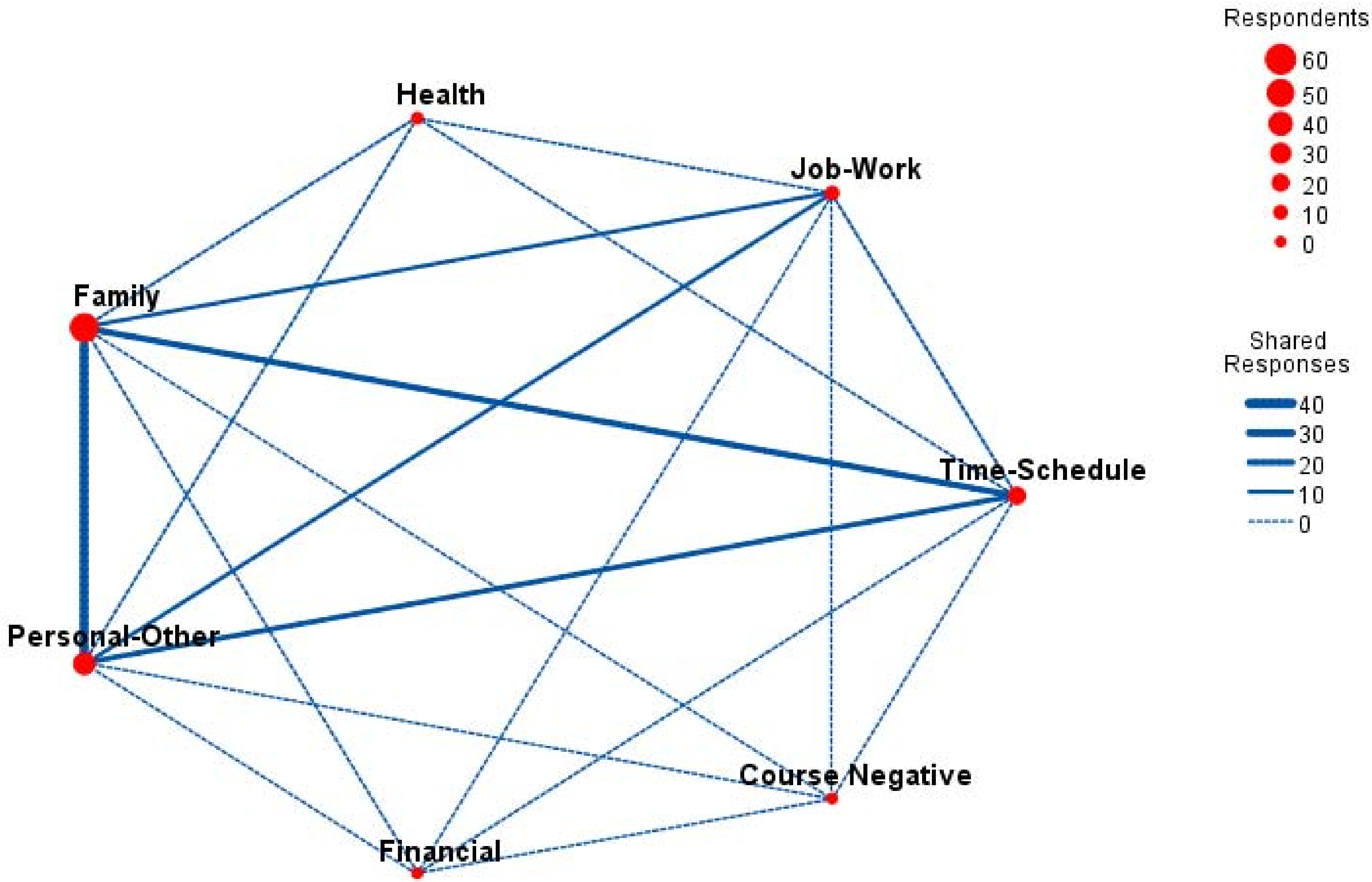

Figure A4. Category web diagram of “Family” category (n = 54) with shared responses.

Appendix A: Additional Model Diagrams (continued)

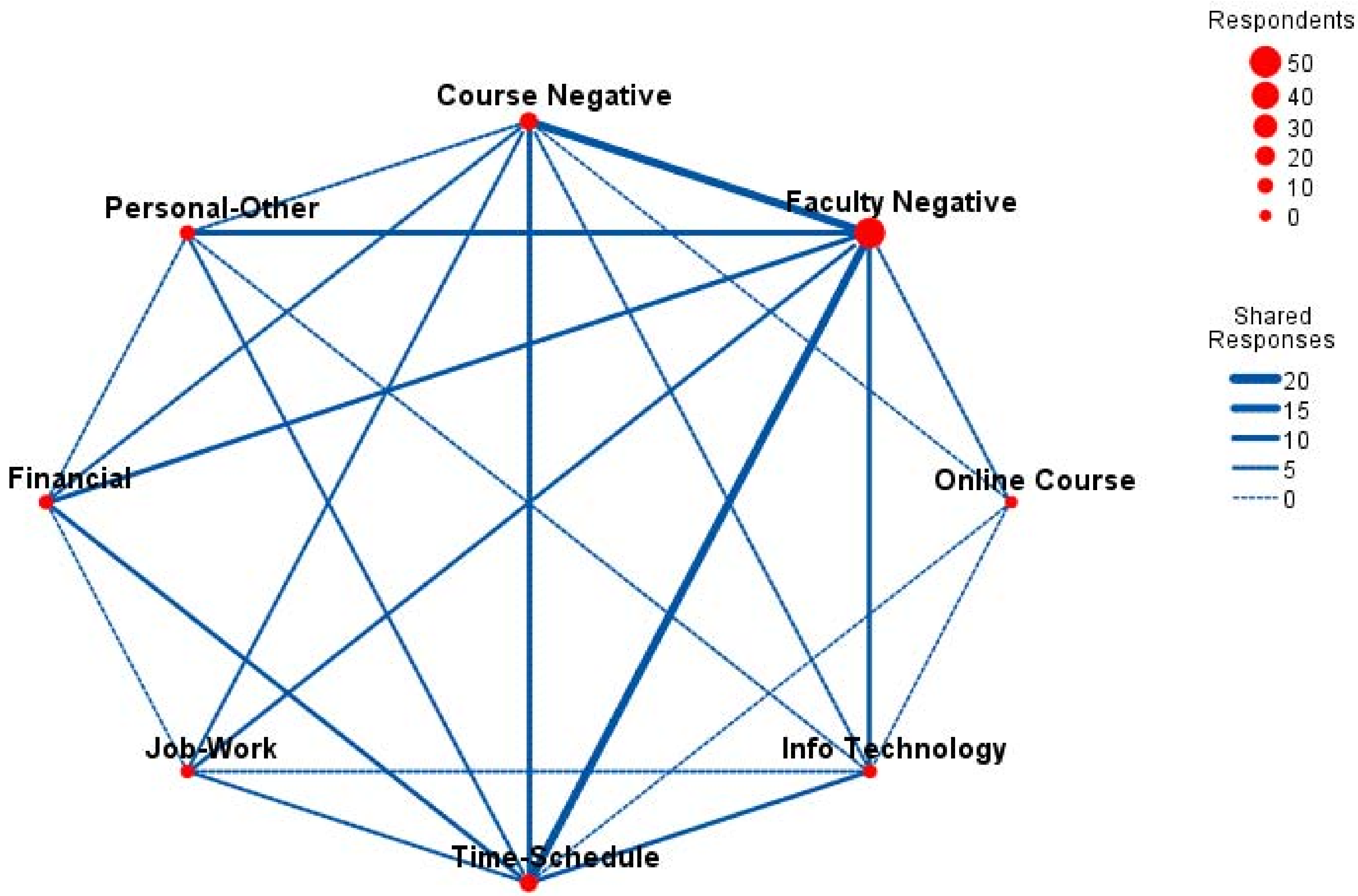

Figure A5. Category web diagram of “Faculty-Negative” category (n = 48) with shared responses.

Appendix A: Additional Model Diagrams (continued)

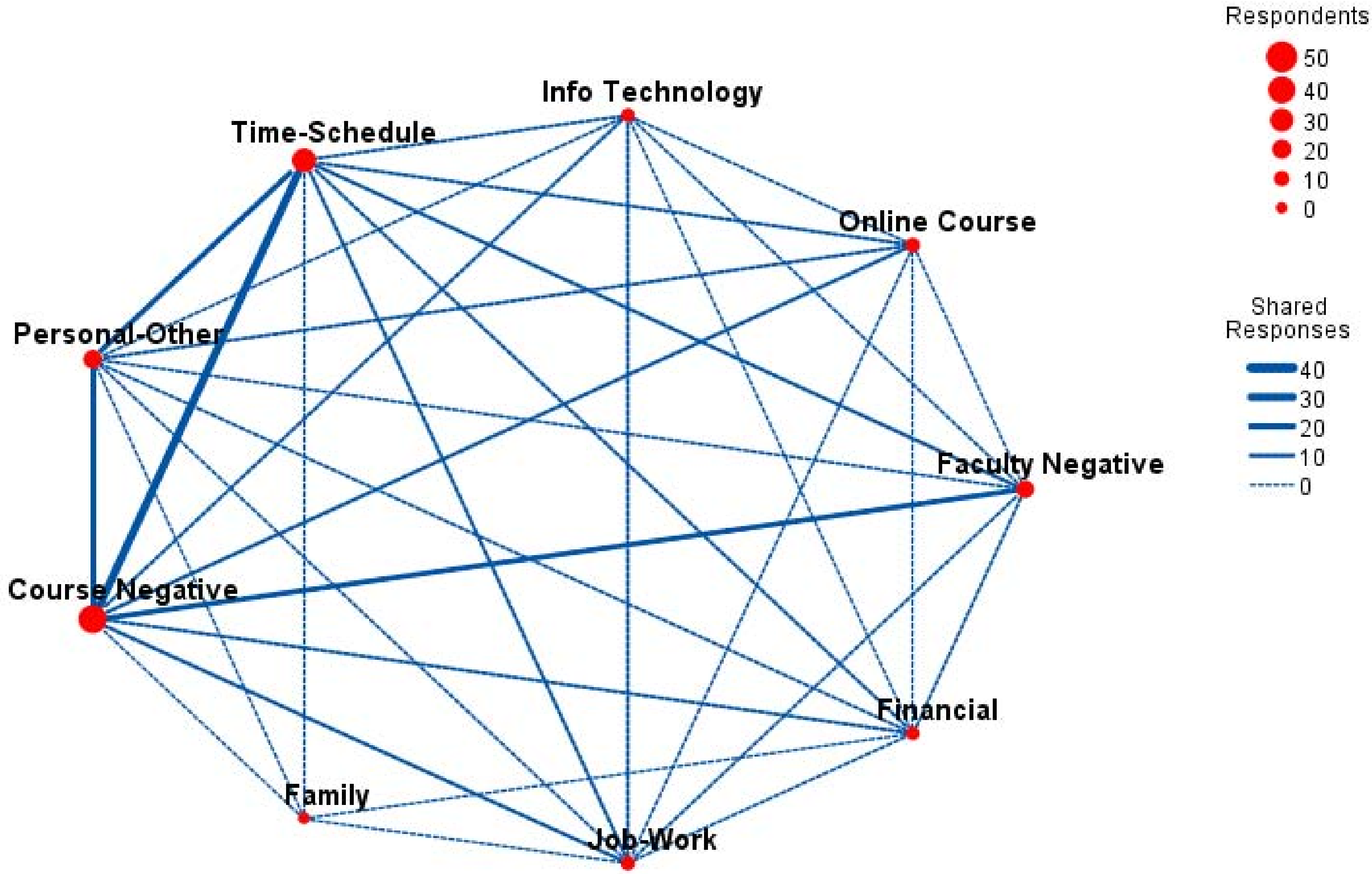

Figure A6. Category web diagram of “Course-Negative” category (n = 43) with shared responses.

Appendix A: Additional Model Diagrams (continued)

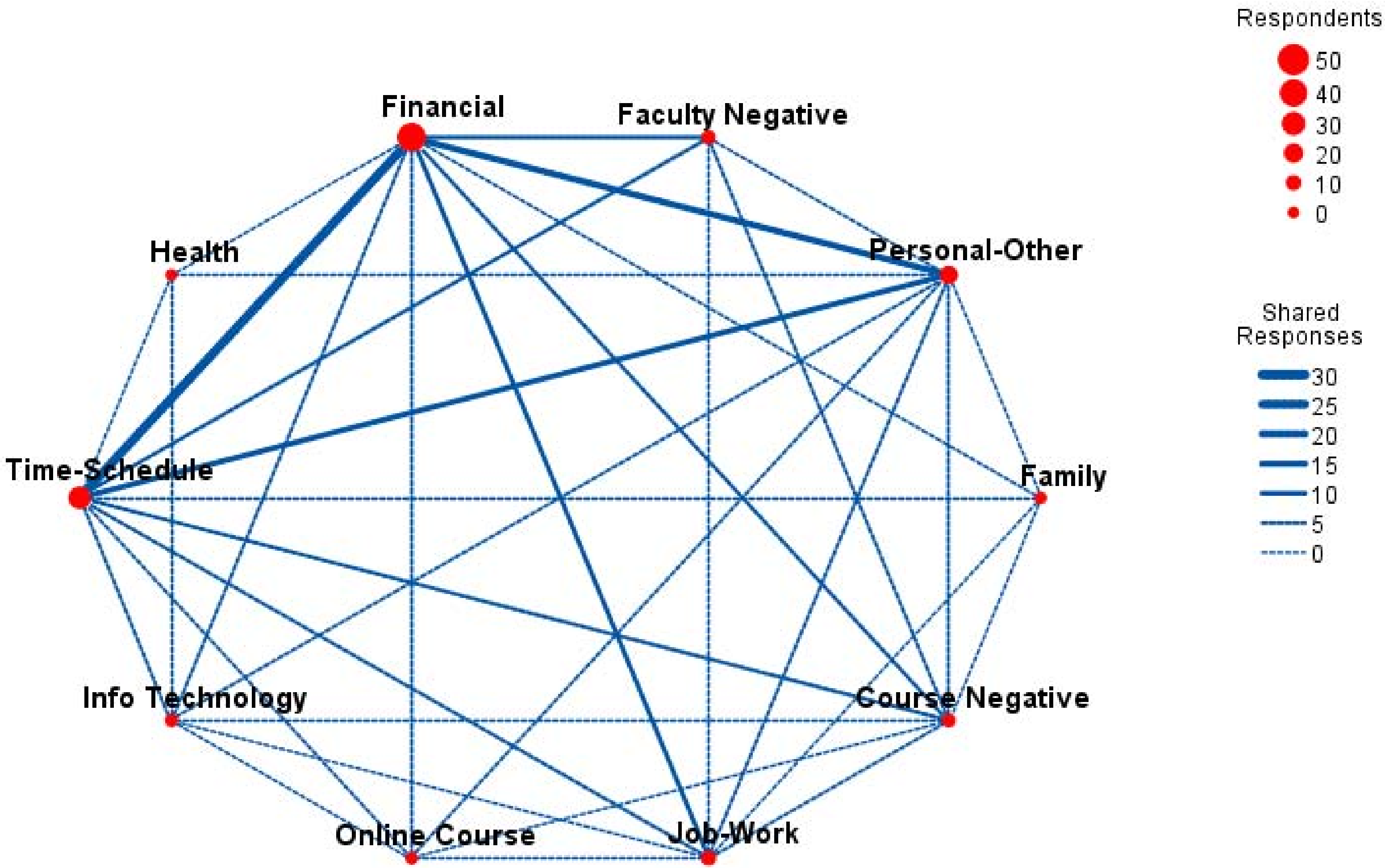

Figure A7. Category web diagram of "Financial" category (n = 43) with shared responses.

Appendix A: Additional Model Diagrams (continued)

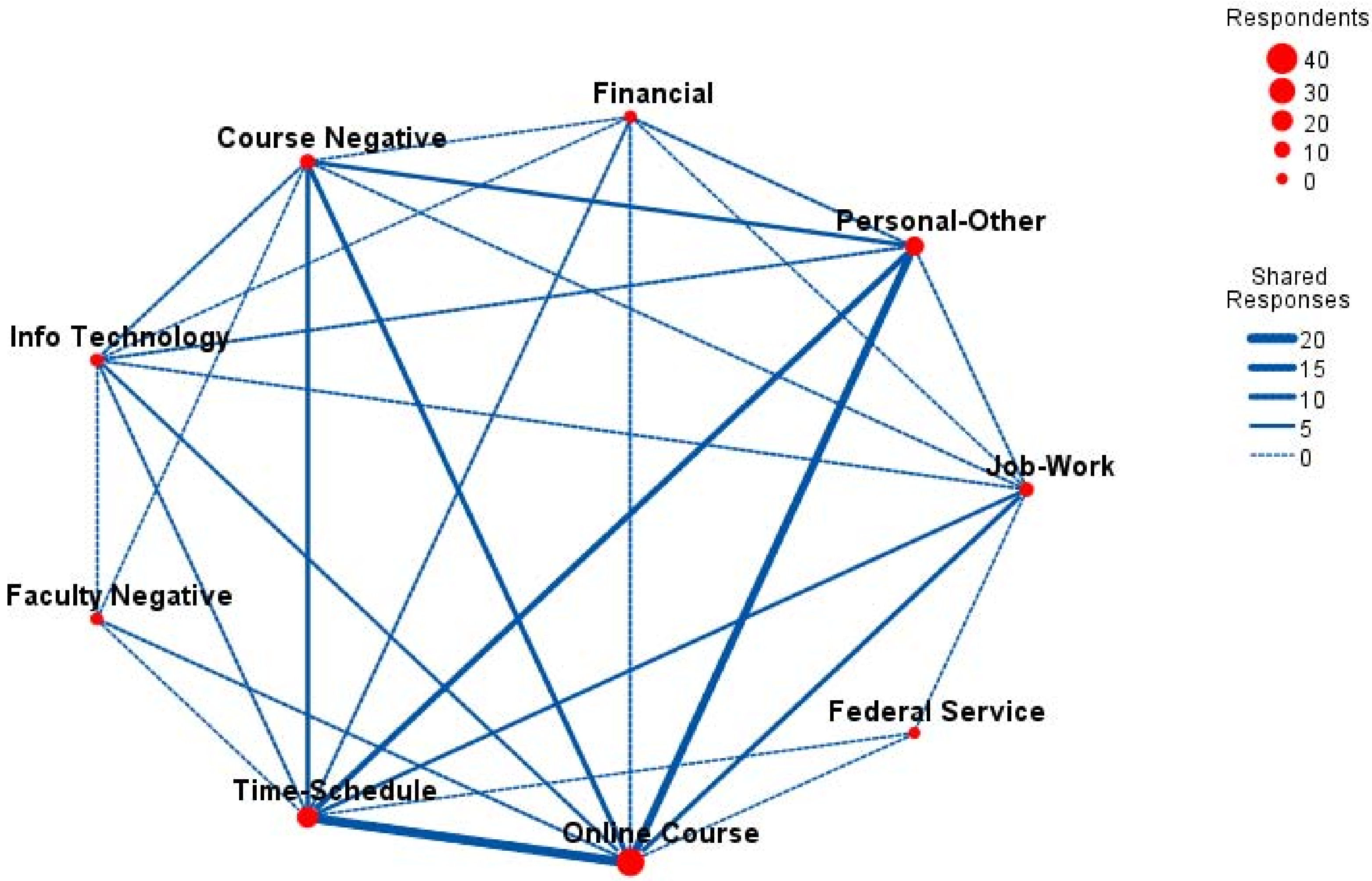

Figure A8. Category web diagram of "Online Course" category (n = 34) with shared responses.

Appendix A: Additional Model Diagrams (continued)

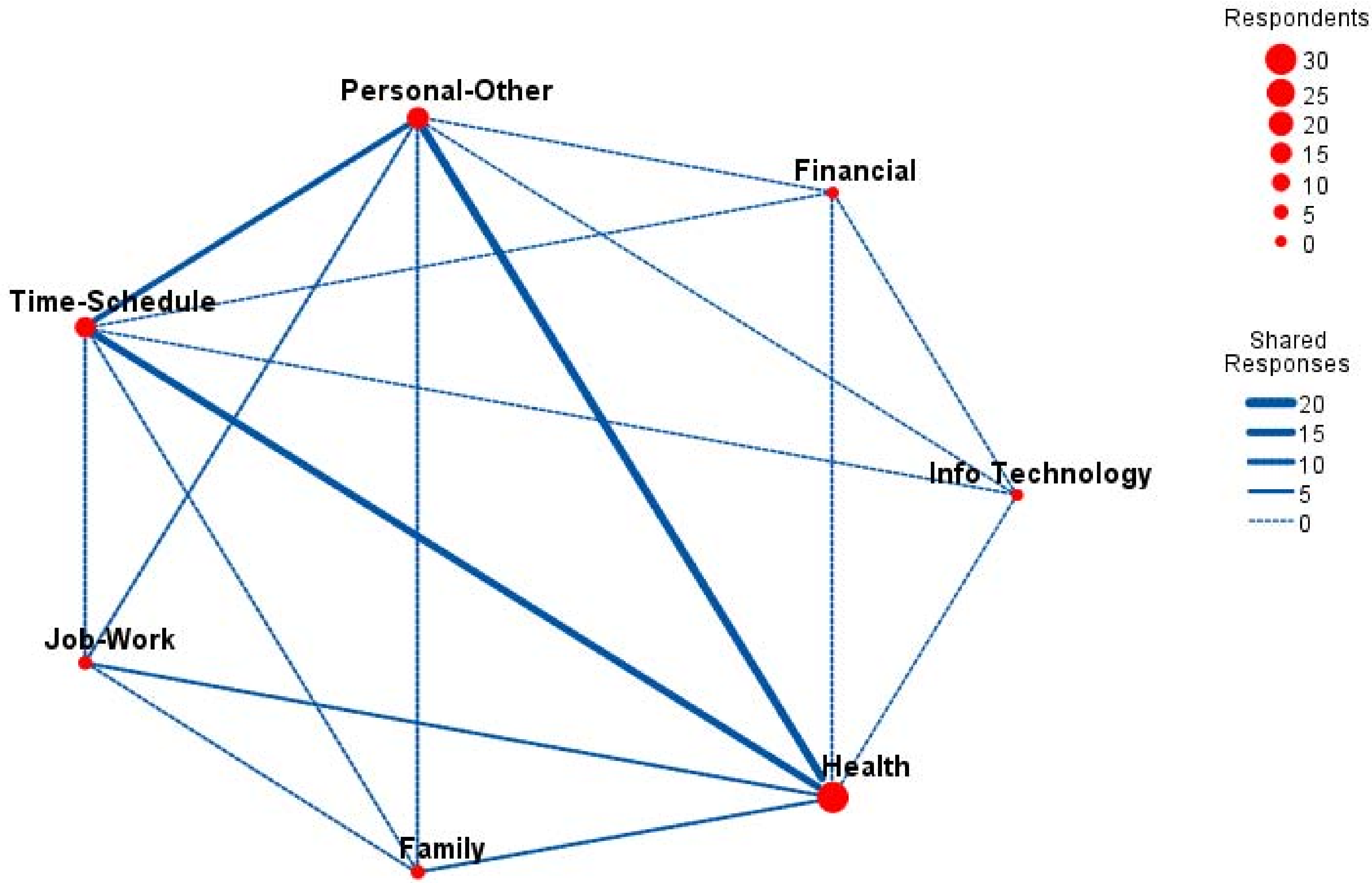

Figure A9. Category web diagram of “Health” category (n = 30) with shared responses.

Appendix A: Additional Model Diagrams (continued)

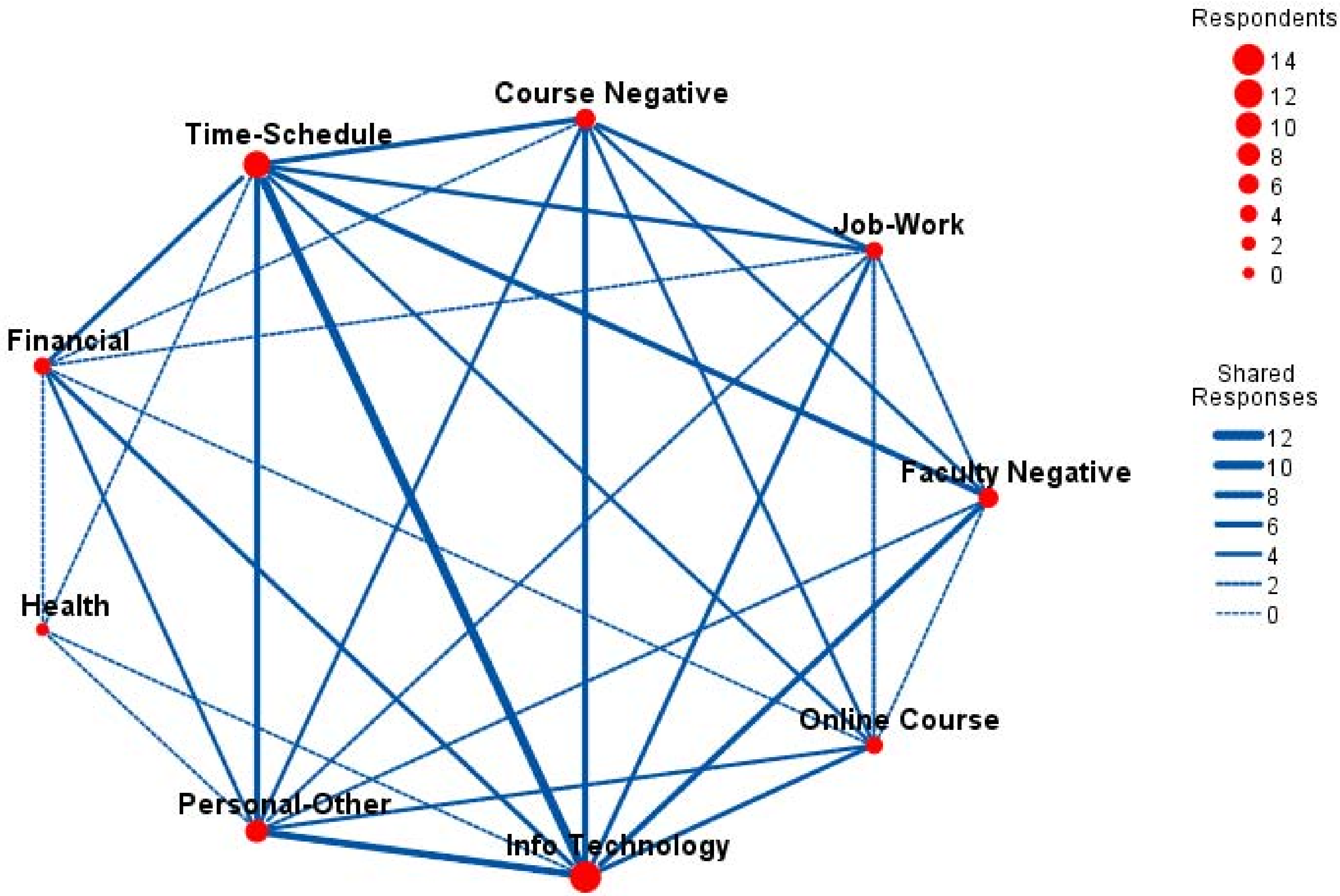

Figure A10. Category web diagram of "Information Technology" (n = 14) with shared responses.

Appendix A: Additional Model Diagrams (continued)

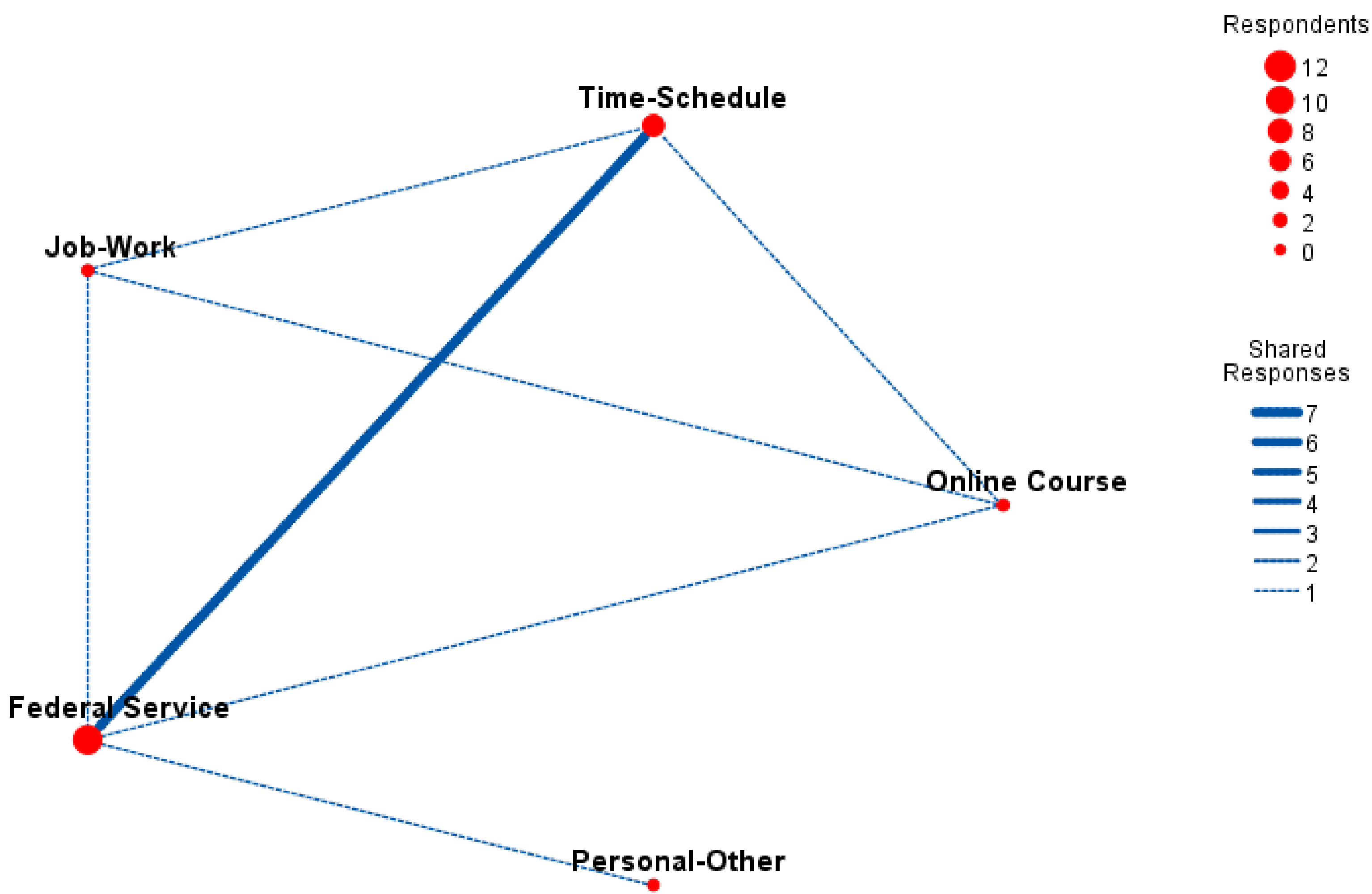

Figure A11. Category web diagram of “Federal Service” (n = 11) with shared responses.