Understanding Persistence of At-Risk Students in Higher Education Enrollment Management Using Multiple Linear Regression and Network Analysis

Donald L. Gilstrap – University of Alabama

Citation of published version:

Gilstrap, D. (2019): Understanding Persistence of At-Risk Students in Higher Education Enrollment Management Using Multiple Linear Regression and Network Analysis. *The Journal of Experimental Education*, 88(3).

DOI: https://doi.org/10.1080/00220973.2019.1659217

Understanding Persistence of At-Risk Students in Higher Education Enrollment Management Using Multiple Linear Regression and Network Analysis

Abstract: Since complex systems theory takes into account anomalies that emerge in linear models, this article presents research on persistence among at-risk students using network analysis following Multiple Linear Regression (MLR). Data taken from an institutional research center on an entire population of enrolled students at an urban serving university over several years (P=35,239) is tested using Multiple Linear Regression. Variables interacting at different dimensions of the model are then analyzed using IBM SPSS Neural Networks (2017) and Cytoscape (2018) rendering to show network linkages between interacting variables. Analysis of the data shows that, while variables may not be found to be significant in MLR models where anomalies are ruled out, network analysis takes these anomalies into account and further reveals complex layers of interactions between and among variables. Findings show that loans of any kind contribute to attrition while financial aid of any kind contributes to persistence and offsets attrition from loans.

KEYWORDS: persistence, financial aid, at-risk, complex systems, network analysis, higher education

Introduction

Since complex systems theory takes into account anomalies that emerge in linear models, this article presents research to identify persistence among at-risk students using network analysis following Multiple Linear Regression (MLR). As colleges and universities strive to increase persistence and retention rates among undergraduate students, attempting to determine what factors and influences contribute to being an at-risk student have received particular attention. At-risk students can be generally defined as students who “are at risk for educational failure—either by failure to learn in school or by dropping out of school altogether” (Kaufman, Bradbury, & Owings, 1992, p. 1). The research literature often generalizes these students along low-socioeconomic backgrounds, minority status, or lack of involvement by parents in the educational process (Kaufman, Bradbury, & Owings, 1992). When we discuss “persistence” in higher education, a working definition used in this paper is that “students continue their enrollment from the time of their matriculation to their graduation” (Hossler, 2002, p. 1). The term “retention” is often used interchangeably with “persistence,” although there is a slight variation in meaning. For this article, the definition of retention “occurs when a student enrolls each semester until graduation, studies full-time, and graduates in about four years” (Bean, 2003). The term “enrollment management” can be defined as “an organizational concept and a systematic set of activities designed to enable educational institutions to exert more influence over their student enrollments… [and] to guide institutional practices in the areas of new student recruitment and financial aid, student support services, curriculum development, and other academic areas that affect enrollments, student persistence, and student outcomes from college" (Hossler & Bean, 1990, p. 5).

There has been a plethora of predictive modeling that has helped us understand these students in more detail and subsequently tailor our recruitment, financial aid, and persistence efforts. Common variables used by researchers in enrollment management to determine at-risk status and persistence include pre-university training, completion of studies, coexistence, adaptation, and performance; however, the most well researched variables fall into the general categories of academic background and support, financial aid, and socio-economic status (Bernardo, et al., 2017). The difficulty in exploring this phenomenon involves the often extremely large number of variables under investigation and which variables a scholar should include in the research process.

Accordingly, strategic enrollment managers would be well served by including methods from complex systems theory, since the concepts of nonlinearity, chaos, and uncertainty relate uniquely to the fluid and somewhat unpredictable environment in university enrollments. As Hossler (2008) has pointed out, "current trends suggest that the next fifteen to twenty years will not be a period characterized by stability for enrollment managers" (p. 13). Complicating typical strategic enrollment management research approaches, linear models can sometimes suffer from boundary conditions that reduce variable sets to a workable number, excluding other variables of interest that might be concurrently important while discounting the role of randomness and variance in the research endeavor (Cohen, 2013; Cohen, Cohen, West & Aiken, 2003; Cohen & Cohen, 1983). As one example, poor construct validity is frequently used as an argument to perform variable reduction due to interaction effects. However, Nimon and Henson (2015) showed that construct validity actually becomes subjective to the relationships between covariates within a linear model, making a strong case for a careful examination of how and when we perform data extraction and transformation in order to reduce a model to a non-messy state and whether this is always an appropriate approach. Equally, it is not the purpose of linear models to look specifically at interactions between variables as the main point of research, whereas network analysis incorporates complexity in models by focusing on "complex simultaneous interaction between microlevel and macrolevel processes" (Gilbert, 1999; Frantz & Carley, 2009, p. 482) or rather interactions at the discrete and individual levels that are intrinsically connected to the boundary parameters of the larger system or systems.

We know that current predictive modeling for student persistence is vast in its scope, and it encompasses hundreds of variables that are commonly used, adding to the strength of this approach when studying at-risk students. However, this research attempts to investigate how variable interaction can also be an important aspect when looking at enrollment management and persistence of at-risk students. What do we miss when we exclude variables from Multiple Linear Regression (MLR) models? How important are non-significant and outlier variables when looking at enrollment management models from a comprehensive lens? How can nonlinear and complex systems models such as network analysis contribute to the research literature on enrollment management? The purpose of this research, therefore, is to first confirm variables commonly found in the literature as contributing to persistence and to next show how network analysis can be used as a compliment to linear models such as MLR, exploring variables and their interactions that may contribute to the persistence of at-risk students that may have previously been overlooked.

Literature Review

When observing educational systems, we are often drawn toward an ideal of stable system states or near-equilibrium. We can think of this through the analogy of a quiet, structured classroom where we are able to control outcomes and where students behave as we expect them to. In quantitative research methods, we have a proclivity to dampen noise or distortion, remove messiness that makes the system difficult to observe, and to make the system model as clean as possible. However, we frequently see situations where students do not behave as anticipated, where quiet and structured classrooms do not facilitate robust learning, and where controlling outcomes serves as a negative feedback loop, dampening the potential for student creativity and innovation in the learning process. In higher education, these conditions can sometimes be amplified in negative ways by financial and social pressures on students, competing priorities between work and school, and new generations of students have technological competencies that outpace those used in the classroom. In effect, we often find educational systems in unbounded chaos, or far-from-equilibrium conditions, which are unpredictable, and their outcomes have the potential to be either positive or negative. However, there is an area in between where we find complex systems. They exclusively exhibit neither near-equilibrium nor far-from-equilibrium conditions. They are not chaotic, but they are also unstable (Mitchell, 2009). And in as much as we might try, complexity is not easily defined but is rather a description of a system state (Waldrop, 1992). As Bertuglia and Vaio (2005) note:

> Complexity is characterized by the breakdown of the symmetry, both that of the perfect order, and that of the total disorder, due to the fact that no part of the system is able to

> provide sufficient information to predict, even statistically, the properties of the other parts… Complexity can be characterized by means of the two dimension of *differentiation* and *connection*. Differentiation means variety, heterogeneity, it means that the different parts of the system behave in different ways. Connection, on the other hand, concerns the constraints that link the component parts to one another, it concerns the fact that different parts are not independent from one another, but also the fact that the knowledge of one part can enable us to determine the characteristics of other parts." (Bertuglia & Vaio, 2005, p. 289, 287).

A major characteristic of a complex system is that it exhibits emergence during its evolution. Emergence can be defined as "the processes whereby the global behavior of a system results from the actions and interactions of agents… In emergence, patterns, structures or properties emerge at the global system level that are difficult to explain in terms of the system's components and their interactions" (Sawyer, 2007, p. 318).

Strategic enrollment management quite often falls within this same framework. Predictive modeling is very helpful, but it attempts to move enrollment data as a system toward a near-equilibrium state. Many enrollment managers, however, will argue that predicting enrollment retention and persistence among at-risk students through the use of data sets is often unstable (Chapman & Jackson, 1981; Hossler, 2008; Nelson, Malone, & Nelson, 200; Delen, 2011). "Differentiation" between at-risk students is disparate and broad in scope, yet there is also a key element of "connection" (Bertuglia & Vaio, p. 289, 287) between these students and the institution within which they are enrolled. In effect, the retention and persistence of at-risk students in studies of strategic enrollment management quite frequently resembles a complex system: there is the individual at the discrete and micro level of the system interacting within the macro level system that includes other students, parent involvement, student loans, scholarships, academic rigor, and the like.

Network analysis is a tool of complex systems theory that has been used for many years as an application to investigate system phenomena at concurrent macro and micro levels, or rather looking at the system as a whole or focusing in on specific parts of the system (Gilbert & Troitzsch, 1999; Franz & Carley, 2009). Network analysis focuses on connections of nodes. These nodes are important areas of interest or variables where the system tends to gravitate to, and these nodes also show how close these areas or variables are to one another. Several decades ago, network analyses were primarily used in engineering, mathematics, and computer science to investigate networks of data, yet this strategy has matured to account for complex interactions among and between system agents. Eckles and Stradley (2012) first proposed the importance of looking at variables emerging from the social networks of college students that help them persist from the first and second year. Their research, which drew from both sociology and student retention theories, utilized the concepts of density and centrality among variables. The results of their study showed that if a particular student's friends had variables leading to persistence, then, consequently, those friends had a large influence on that student equally persisting. Davidson, Beck, and Grisaffe (2015) expanded on this framework of thinking, by confirming variables of affect in social networks but also by confirming interacting variables in data networks such as test scores by using an "nomological network of variables" (Davidson, Beck, & Grisaffe, 2015, p. 165) through the use of factor analysis.

Forsman, et al. (2014), who also used factor analysis, found that models coming from complexity theory, primarily network theory, uncovered variables of influence that were significant in determining student retention through multidimensional scaling. Delen (2011) integrated neural networks, decision trees, and logistic regression as three main models of persistence, showing that network analysis had "an overall accuracy rate of 81.19%" in determining significant variables while accurately classifying first year freshmen (Delen, 2011, p. 28). Moreover, their research showed that variables relating to finances and education were the strongest of the predictor variables. Additionally, Raju and Schumacker (2015) used logistic regression, as well as neural networks, as data mining tools in order to determine that GPA was a significant predictor of student persistence.

As Dowling (1998) showed many years ago, the director of financial aid plays a critical role in developing the enrollment policies and practices at institutions of higher education through the use of entrepreneurialism and marketing strategies. Moreover, Leslie and Brinkman (1987) poignantly note that "per dollar of subsidy, aid programs, if carefully administered, should be more effective than low tuition policies if the goal is to improve access" (Leslie & Brinkman, 1987, p. 197). "Aid offers could play an important role in affecting students' choice among schools" (Manski & Wise, 1983, p. 23). And "surveys of institutions concerning their financial aid policies suggest that the number of colleges awarding some aid on a no-need or academic basis, without regard to financial need considerations, is growing" (Chapman & Jackson, 1987, p. 7)." We have seen simultaneously, "decreasing levels of subsidies whether in the form of governmental appropriations, private gifts, or endowment earnings result in students having to pay a higher proportion of the total costs of their educations" (Brown, 2008, p. 40) while at the same time, "public sentiment against soaring tuition rates and resulting legislative caps on tuition increasingly threaten its controllability by state-governed institutions" (Brown, 2008, p. 46). In recent years, we have focused on financial aid stacking and tuition discounting as strategies for increasing enrollments at our institutions. It may well be that these strategies play a significant role in helping solve the problem of persistence among at-risk students.

When looking at MLR models, we can see how predictive modeling has limitations. Oftentimes, we are forced to remove variables within the model by nature of the statistical package we use, as is the case with backward stepwise logistic regression in SPSS (Nelson, Malone, & Nelson, 2001) Likewise, in a typical MLR model, multicollinearity diagnostics are also used to remove variables interacting with one another as a means to reduce the model to a level that supports the most significant and well defined variables (Cohen & Cohen, 1983, 2003; Lewis-Beck, Bryman, & Futing Liao, 2004). Multicollinearity can be defined as:

> "In a vector model, in which variables are represented as vectors, exact collinearity would mean that two vectors lie on the same line or, more generally, for *k* variables, that the vectors lie in a subspace of a dimension less than *k*… In practice, an exact linear relationship rarely occurs, but the interdependence of social phenomena may result in 'approximate' linear relationships. This phenomenon is known as multicollinearity" (Lewis-Beck, Bryman, & Futing Liao, 2004, p. 668).

Furthermore, the use of more than ten variables in an MLR model is uncommon, because most researchers realize high multicollinearity scores will inevitably result. However, interaction effects between variables can serve as catalysts for understanding complex systems phenomena through deeper investigation of multicollinearity tools (Author, XXX). Moreover, Cohen and Cohen (1983) argued that high multicollinearity scores should not be used as an omnibus test for rejection of the MLR model, even though this practice is quite common. Additionally, the research of Wolff Smith and Beretvas (2015) has shown that applied researchers tend to extract respondents whose data include variables of mobility when assessing student achievement, and therefore contributing to bias effects among methods used in contemporary research.

As Miller (2010) shows in his research on student attrition through the use of the College Student Expectations Questionnaire, retention could be predicted at between 80-87% based on the variables under scrutiny within the logistic regression model of prediction. This helps reinforce that linear models are indeed critical to understanding the at-risk student. However, Miller (2010) also points out that there are several variables that contribute to the roughly 20% of students that cannot be explained. In the case of his research studies, as well as Santos (2004), mentoring and intervention programs helped explain unaccounted variance in these regression models (Miller, 2007, 2010).

In Nelson, Malone, and Nelson's (2001) critique of Madaus and Walsh's (1965) study of persistence among graduate students showed that "differing Beta weights for regression equations were determined for students in different areas. Later investigations found that the area of study was itself a predictor of success for at risk students in graduate work" (Nelson, Malone, & Nelson, 2001, p. 3). Moreover, these authors also argued that, although GPA is a strong predictor of academic success, there are cases where students with low GPA scores excel within graduate programs (Nelson, Malone, & Nelson, 2001). When studying persistence among graduate students, therefore, the area of study and GPA are both concurrent predictor variables and non-significant variables, depending on the methods chosen when conducting these types of research studies. And since these variables are not exclusive disjunctions, removing them from the MLR model either by intent or through stepwise regression takes away a great deal of information and observation about complex system agents and how they interact with one another when analyzing factors that contribute to students being at-risk.

Theoretical Framework

This article draws primarily from White and Johansen's (2005) research on network analysis in ethnographic anthropology to develop a theoretical framework for expanding upon the MLR model used as the preliminary foundation for this research of emergent phenomena within data sets. Estrada's (2012) research on complex networks also provides a macro system perspective of network analysis. What the reader will find most germane to the theoretical framework in this paper is the role of *centrality* within network structures, a concept equally noted as important by Eckles and Stradley (2012). For the purposes of this paper, the "system network" is an analogy to the data that are interacting in ways that were unpredictable in linear models. Data themselves can be a network as we see frequently in a network analysis that is conducted in computer science and engineering. These are not real world networks, they are theoretical networks of data sets we see commonly explored in mathematics.

White and Johansen (2005) propose two significant challenges to the research community when looking specifically at data sets as networks. Data collection and analysis has historically passed through the filter of the researcher in order to normalize discrete units of measurement about individuals within groups. Once this normalization of data takes place, the researcher can then move forward on analysis of clean data elements. The authors contend that, "behavior itself, however, is an instantiation of a symbolic system: like any other sign or symbol, behavior can be read or interpreted (White & Johansen, 2005, p. 7)." As Miller (2010, 2007) shows in his research on enrollment management, variables that contribute to student retention can be predicted; however, several variables related to retention cannot be explained. Consequently, our current understanding of how enrollment management phenomena actually emerge within data set networks is insufficient at the present time.

In complex systems, phase transitions are discrete snapshots in time where phenonoma cannot be predicted. However, their location in a dimensional model can be estimated to take place in given areas based on probability scenarios. In typical linear models, descriptive statistics show that the mean, median, and mode are equal or close to equal in the normal curve through measures of centrality (Cohen, 2013). In network analysis methods, however, "macro properties of networks may or may not be directly linked to micro properties but in either case macro properties alter the context of everyone in a network, and may affect how people interact" (White & Johansen, 2005, p. 9-10). After a phase transition, "the global network is almost certain to have a connected component that is giant relative to the others, containing most of the nodes" (White & Johansen, 2005, p. 9-10). This very large "connected component" is the attractor state that serves to influence centrality. In the same ways, we see this global network emerge in ethnographic studies of humans, we see similar phase transitions in data sets about human beings. Although adjacency of system agents plays a critical role in any network analysis, or as Frantz and Carley (2009, p. 490) describe as "closeness" and "betweeness," it is the aspect of centrality that helps define the system as a network (Eckles & Stradley, 2012). The closeness and betweeness of nodes within data sets also highlights the role of centrality in network analysis of simultaneous conflicting and complementing phenomena that emerge within data sets of several variables. In a linear model, we might describe this as a nonlinear relationship between predicted outcomes and residuals when posited beside a regression line.

White and Johansen (2005) equally propose an additional challenge of scale independence. Scalability at the micro and macro levels of a system is one aspect of research methods that differentiates network analysis from linear models much like Forsman, et al. (2014) described through multi-dimensional scaling. Macro and micro level variables, or rather general system-wide variables compared to specific variables within the system, have the potential to be equally mutually exclusive while the macro level still influences the emergence of phenomena at the micro level and vice versa. Therefore, there is a feedback loop taking place in network analyses where the micro level variables affect the macro level variables and subsequently changes the initial conditions of the micro level. What makes this concept important for persistence of at-risk students in higher education is that network analysis places those variables in relation to each other through centrality.

This theoretical framework, therefore, focuses on data sets as theoretical networks rather than exploring real world networks of individuals. Both MLR and network analysis models serve to advance knowledge about a phenomenon, yet both methods have different frames of reference with subsequently different results. Both methods also reveal a paradox which, at face value, suggests an exclusive disjunction (*XOR*) in where the “truth” resides: only one method and subsequent result can be correct. This paradox reinforces the proposition that these methods are not mutually exclusive but are rather concurrently and, moreover, multi-dimensionally correct. It is rather the questions we ask about the phenomenon that lead us to different conclusions, and, in the case of research on persistence of at-risk students in higher education, using both MLR and network analysis shows us how complex this issue is and, at the same time, provides us with multiple lenses for understanding persistence.

Methodology and Procedures

Sample

For this research, data sources used as a sample from the larger U.S. population of at-risk students for this paper are drawn from several years of data collected at an urban-serving Midwest Carnegie High Research institution. Since this is an urban-serving institution, the scope of at-risk students falling within this population is robust with many commonly generalized variables being accounted for within the group, such as GPA, ACT scores, first generation student status, transfer status, engagement by parents in the educational process, and underrepresented minority status. This cross sectional data set was composed of the population (*P*=35,239) of undergraduate students who were enrolled at the university and were tracked over a period of six years. Since this researcher had access to the entire population of undergraduate students, there was no need to perform random sampling.

Measures

In the original variable set prepared and analyzed by the university's office of institutional research, eighty-one variables were assigned to each student to help determine what factors contribute to at-risk status, based upon entering the university or after having completed one or more semesters at the university. Predictive modeling was then conducted by the office of institutional research, and twenty-seven independent variables were identified as contributing in a connected fashion to at-risk status determined by the dependent variable *currently on probation*, which was categorized by the office of institutional research as having a GPA of 2.0 or less, receiving and incompletes in classes, falling below full-time status due to dropped classes, and/or a combination of these attributes. However, it is not the purpose of this research to explore that study but rather use the significantly identified independent variables as a springboard for new research.

Since this research is exploratory in nature, MLR was chosen as an appropriate test to analyze continuous and categorical variables through stepwise regression, since this model is most appropriate for exploratory research of both categorical and continuous variables while also providing a framework for understanding variable interactions through collinearity diagnostics (Cohen, Cohen, West, & Aiken, 2003). VIF (Variance Inflation Factor) and Tolerance scores were computed, as well as a condition index, to determine if and when interaction effects between variables were taking place at different dimensions of the collinearity diagnostics. It is important to note here, however, that interaction effects are not meant as the traditional covariance that leads to variable exclusion due to faulty constructs of variables. These are interactions that reveal relationships between variables that are symbiotic in nature for which variance is accounted. The interactions emerge through multiple dimensions of the condition index, and, by different dimensions of this index, the reader is challenged to consider the impossibility of being able to graphically represent a model that has 27 dimensions. Multicollinearity diagnostics were then analyzed for these interaction effects at different dimensions of the model [Tables 1 and 2]. When reviewing multicollinearity diagnostics, keep in mind that collinearity is a result of computing Eigen values within one or more Eigen vectors. As a result, consider that each dimension, one through twenty-seven, is a multi-dimensional construction in 27D rather than 3D form.

[INSERT TABLE 1 AND TABLE 2 HERE]

The multi-dimensionality of the Condition Index then helps the reader understand where exactly interaction effects are taking place. When looking at Table 2, the Collinearity Diagnostics serve as an Adjacency Matrix. The furthest left column "Model" denotes each dimension of the model where a variable is added in. In effect, 1=*Constant*; 2= *College division of major* in relation to the *Constant (1)*; 3=*undecided college major* in relation to the previous model numbers *College* and the *Constant* (1-2); 4=*new fulltime freshman* in relation to the previous model numbers (1-3), and so on.

Variables that were interacting at different dimensions of the condition index were then analyzed using IBM SPSS Neural Networks (2017), using both Multilayer

Perceptron and Radial Basis Function. White and Johansen's (2005) first proposition stated that "networks have structural properties (local and global) (White & Johansen, 2005, p. 9)" and that power law coefficients affect "when… a phase transition begins from a network having tiny 'islands' of successively connected pairs in a disconnected 'sea' to one having larger connected 'islands' (p. 10, 17). Neural networks is uniquely analogous to this research proposition, since Multilayer Perceptron looks at scalability from the macro to micro levels of the network of variables, and Radial Basis Function looks at the distances between nodes within the model. In the case of this research, variables found to be significant in the MLR model can be associated with the smaller "islands" while the nodes to which they are attracted can be viewed as the "larger connected islands." Since the output of the network map derived from Multilayer Perceptron on all variables included in the MLR model was extremely large, containing thousands of nodes and edges, network analysis of the multicollinearity adjacency matrix was also conducted using Cytoscape.

Summary of Post-hoc Regression Network Analysis

A working summary of this post-hoc method can help researchers working on enrollment management of at-risk students in many ways, and the following suggestions are given to the reader. First, by utilizing variables at the departmental, college, or institutional level that have been identified as contributing to at-risk student categorization, we are able to form a general foundation for network analysis. The researcher should perform Multiple Linear Regression (MLR) on all of these variables collectively, adding VIF and Tolerance Scores into the regression model, as well as computing collinearity diagnostics. Analysis of collinearity diagnostics enables the researcher to observe emergent phenomena through multi-dimensional scaling of interaction effects [Tables 1 and 2]. Although the field has not reached a formal conclusion on the level of multicollinearity, as a general rule of thumb, variables with collinearity scores of variance explained >.30 are the focus of interest for this study (Cohen, Cohen, West, & Aiken, 2003). This collinearity table is dimensional at different stages of the MLR test, particularly when one has performed either stepwise or zero-order regression. After analyzing VIF and Tolerance score ranges in relation to the Condition Index, one needs to identify which variables interact with each other and at which dimensions of the model [Tables 1 and 2]. These variables then become the area of focus for network analysis.

After completing these steps, the research then takes the entire set of variables originally included in the MLR model. In the case of this research, all 27 independent variables are then analyzed through SPSS Neural Networks Multilayer Perceptron and Radial Basis Function to highlight nodes of centrality and adjacency. Analysis of the neural network maps produced by Cytoscape [Figures 1 and 2] then helps the researcher to identify variables of adjacency and centrality while equally confirming or refuting those significant variables (p<+/-.05) previously perceived to be interacting at different dimensions of the MLR model. Equally, the researcher should be looking for variables that were not found to be significant in the MLR model but that show high levels of centrality and/or adjacency found in SPSS Neural Networks maps. The collinearity diagnostics adjacency matrix can also be analyzed through network analysis, in the case of this research using Cytoscape. The combination of both insignificant and significant variables that show high levels of centrality and/or adjacency in network analysis should then be accounted for in future predictive modeling on persistence of at-risk students. In effect, MLR provides a linear model for identifying statistically significant variables. Network analysis provides evidence of important variables that might not be significant in a linear model but appear to influence the system as a whole through their interactions with other variables.

Data Analysis

Multiple Linear Regression was conducted on the 27 independent variables in relation to the dependent variable *currently on probation*. An ANOVA test of the strength of significance of the regression model was found to be highly significant [$p < 0.001$], protecting against the likelihood of Type I errors, with a moderate effect size ($R^2 = 0.329$). Results of the MLR test showed significance for twenty-one of the twenty-seven variables when analyzed collectively [Table 1]. Not surprisingly, several variables confirmed already existing predictor variables available in the extant research so far on at-risk students which were the basis of the original data set prepared by the office of institutional research such as GPA, ACT scores, transfer status, first generation student status, age, and underrepresented minority. Analysis of VIF and Tolerance scores, as well as collinearity diagnostics, did not show variables were interacting at different dimensions of the model; however, analysis of the collinearity matrix does show these interactions. Normally, these variables would be removed in a typical MLR setting either using backwards stepwise regression or through direct exclusion (Nelson, Malone, & Nelson, 2001; Cohen & Cohen, 1983). However, the emergence of phenomena that are unable to be observed in a reductionist method is perhaps the defining characteristic of complex systems, from the pure and applied sciences to the social and behavioral sciences (Waldrop, 1992; Bak, 1996; Kauffman, 1995; Bertuglio & Vaio, 2005; Holland, 1998; Byrne, 1998; MacIntosh & McLean, 1999, 2001; Mitleton-Kelly, 2003; Author, XXX; Guastello, 2003; Stacey, 2003; Simon, 1985; Bar-Yam, 1997; Prigogine, 1967, 1980).

Emergence, as defined previously by Sawyer (2007), can be observed in the MLR model through analysis of the high Condition Index Score found in the multicollinearity adjacency matrix. As shown in Table 1, in addition to zero-order, part, and partial correlations, VIF and Tolerance indicators were computed. VIF "provides an index of the amount that the variance of each regression coefficient is increased relative to a situation in which all of the predictor variables are uncorrelated" (Cohen, Cohen, West & Aiken, 2003, p. 423). Although there is disagreement in the field on an index level for VIF scores, "a commonly used rule of thumb is that any VIF of 10 or more provides evidence of serious multicollinearity involving the corresponding IV… Tolerance is the reciprocal of the VIF and therefore tells us how much of the variance in X_i is independent of the other IVs" (Cohen, Cohen, West & Aiken, 2003, p. 423). In other words, Tolerance can be computed by taking the whole integer 1 and dividing it by the VIF score. These additional tests compliment the calculation of the Condition Number *(K),* which "is defined as the square root of the ratio of the largest eigenvalue to the smallest eigenvalue," or rather the square root of the largest divided by the smallest (Cohen, Cohen, West & Aiken, 2003, p. 424). As can be seen, the VIF and Tolerance levels are in acceptable ranges for almost all of the independent variables when each variable is analyzed separately. However, when each variable is analyzed systemically, the MLR model shows a very high Condition Index (CI=55.658) in the multicollinearity diagnostics adjacency matrix (Table 2), suggesting interaction effects at different dimensions of the model. What makes this most important to the researcher is that it is an exception that a variable set with this large of a Condition Index would still show VIF and Tolerance scores within normal ranges (Table 1). One would expect to find many of those scores out of range, but that is not the case!

Some researchers might argue, at this point, a better or different regression model would have been more appropriate due to the high multicollinearity involved, such as principal components, ridge, or ordinary least squares. However, as Cohen, Cohen, West, and Aiken (2003) have strongly argued, it makes no sense to test with a regression model other than MLR when multicollinearity has been observed, and, "instead, the researcher's focus should be on attempting to understand the nature and the source of the multicollinearity" (Cohen, Cohen, West, & Aiken, 2003, p. 429) Further observation of the Condition Index in this study led this researcher to examine the variables that were interacting at different dimensions of the MLR model and incorporate them into SPSS Neural Networks analysis, using both Multilayer Perceptron, which looks at scalability from the macro to micro levels, and Radial Basis Function, which looks at distance between nodes within the model. These two tools of analysis are analogous to White and Johansen's (2005) first proposition, that "networks have structural properties (local and global) (p.9)" and that power law coefficients affect "when… a phase transition begins from a network having tiny 'islands' of successively connected pairs in a disconnected 'sea' to one having larger connected 'islands' (p. 10, 17).

Findings

Further analysis of the adjacency matrix of multicollinearity diagnostics revealed interaction effects between and among several different variables within the MLR model [Table 2]. Approximated variance proportions between and among the dimensions of the model were analyzed by evaluating standardized coefficients in relation to variables found to be significant (Table 1), the independent variables involved, and where the interactions within different dimensions of the collinearity table took place (Table 2). Returning to the concept of multi-dimensionality of the collinearity diagnostics adjacency matrix discussed previously, when looking at Table 2, we can view the model dimensions at the far left of the table as rows of an X axis and the variables as columns as a Y axis. The variable *new freshman transfer student* is at the 5th dimension of the model (Y axis) and shows a high variance proportion at the 7th dimension of the model (X axis) which is in relation to *new junior transfer student* (Y axis). Likewise, this variable (7th) shows high variance proportion at the 8th dimension of the model which is in relation to *new senior transfer student*. This variable (8th), along with *new sophomore transfer student* (6th*),* show high variance proportions at the 5th dimension of the model which is in relation to *new freshman transfer*. As a result, there was a great deal of interaction between *new freshman, new sophomore, new junior,* and *new senior transfer student* as variables. In a typical MLR model, one could make a solid argument that class status was not strong enough to distinguish between the variables, and, rather, simply *new transfer student* would have been a more robust variable for reduction. However, interactions between *department major (2nd)* and *race (17th)*, or between *first generation* *student (19th)* and *cumulative GPA (11th)* do not fit easily within this way of thinking when performing data reduction in MLR models.

The use of network analysis equally upheld White and Johansen's (2005) fourth proposition that "emergents may be local or non-local, depending on whether they have micro-macro linkages" (White & Johansen's, 2005, p. 27). The network map created by SPSS Multilayer Perceptron cannot be included in the confines of this article since it contains thousands of nodes and edges. However, as just one example, Multilayer Perceptron revealed a network model, showing there are several micro nodes of variables, such as *total charges to student, fall semester award amount,* and *fall semester loan amount*, each scaled at different levels and contributing in significant ways when analyzing variable interactions in relation to the macro level variable *currently on probation*. Additionally, network analysis was performed on the multicollinearity diagnostics adjacency matrix using Cytoscape, and Figure 1 shows a summary network map of the interacting variables.

[INSERT FIGURE 1 HERE]

Interestingly, the node which identified students receiving *fall semester grants* served as a node of centrality for the other variables through micro linkages in moving students away from being on academic probation. In the MLR model, this variable was found to be significant ($p<.001$). Conversely, the node which identified students receiving *fall semester loans* served as a node of centrality for the other variables through micro linkages of adjacency in moving students toward being on academic probation. Understandably, in the collinearity diagnostics adjacency matrix, *fall semester loans* showed interaction with *maximum ACT score*, implying that if one has a high ACT score, the amount of loans s/he receives will be smaller compared to someone with a low ACT score. In the MLR model, conversely, *fall semester loans* was not significant (p>.400) supporting the findings of Forsman, et al. (2014); however, it proved to be a highly linked node within the network analysis model.

As White and Johansen (2005), as well as Estrada (2012), have argued, the roles of centrality and adjacency both provide strength for network analysis models in descriptions of observations and in differentiating outcomes from typical observations in linear models. Both of the nodes for the variables *fall semester loans* and *fall semester grants* show the power of centrality in network analysis in the nodes' adjacencies to the other variables. The closer the micro linkages, or edges, of each node were adjacent to the node *fall semester loans*, the stronger their relationship to the node of centrality that led to academic probation. So the node *fall semester loans* served as a node of centrality, where those nodes that were adjacent to it tended to move. Conversely, nodes that were adjacent through micro linkages to *fall semester grants* had a stronger relationship to the node of centrality that led students away from being on academic probation. Moreover, nodes of adjacency through micro linkages to *fall semester grants* were not adjacent and did not follow patterns of centrality as those nodes that were adjacent to *fall semester grants* and vice versa. These findings support the research of Nelson, Malone, and Nelson (2001) who pointed out, "health, financial, or family situations may influence students' decisions to complete their degrees" (Nelson, Malone, & Nelson, 2001, p. 6). Moreover, Multilayer Perceptron revealed that individual students exhibited traits of "smaller islands" within the "larger islands" of the dollar amounts of *fall semester loans* (White and Johansen, 2005). Although not linear in nature due to the influence of additional variables, network analysis showed that the larger the dollar amount of *fall semester loans*, the closer adjacency students had to the node being on academic probation (Figure 2). Conversely, in similar fashion, the smaller the dollar amount of *fall semester grants*, the closer adjacency students had to the node being on academic probation.

[INSERT FIGURE 2 HERE]

Discussion and Conclusion

This study helps to confirm Nelson, Malone, and Nelson's (2001) findings that finances influence a student's decision to persist. Equally, this study supports Delen's (2011) findings that the use of neural networks models are important in determining student persistence. In the case of this research, financial aid packages offsetting tuition and fee structures could indeed play a significant role in persistence among undergraduate students who might be viewed as at-risk. In effect, the phenomena that emerged through this method serve as nodes of centrality and adjacency in neural networks, again adding strength to the complementarity of complex systems theory used alongside linear models. The nodes *fall semester loans* and *fall semester grants* are connected through adjacency to other variables that lead students toward or away from being on academic probation, and both nodes operate within the same systemic boundary conditions.

This research represents a snapshot of one particular phase space within a larger trajectory of generalized predictive modeling on persistence of at-risk students in higher education. Detailed analysis of these data sets shows that, while variables may not be found to be significant in linear models where anomalies are ruled out, centrality and adjacency in network analysis takes these anomalies into account and further reveals complex layers of interactions between and among variables. *Findings show that loans of any kind contribute to attrition while financial aid of any kind contributes to persistence and offsets attrition from loans*. As a result, a complimentary method to MLR is introduced in at-risk student persistence research through additionally using network analysis post-hoc. This research is exploratory in nature, and more studies on student persistence data sets would need to be conducted to triangulate the complimentary roles of network analysis models in understanding at-risk students. Therefore, additional investigation of similar phenomena by other researchers on the methods and outcomes described in this research study is encouraged. Equally, researchers are strongly encouraged to explore this phenomenon using probability scenarios such as Monte Carlo, which provides probability distributions, Bayesian, which provides iterative probability distributions over time, and Lyapunov Exponent analyses, which explain probabilities in phase spaces, in order to expand the use of complexity theory in student persistence research.

Complex systems phenomena at the micro level highlight the interdependence among and between variables at different scales of focus in order to help describe the emergent phenomena at the macro level. As examples, there were several interaction effects between freshmen, sophomore, junior, and senior transfer students. However, these interactions were not equidistant dimensionally within the MLR model, and they were not equally bounded between all four variables. Conversely, a linear analysis of the MLR model showed that any new freshman, regardless of being a transfer student, was significant for predicting being on academic probation, while being a transfer student at any semester standing was significant regardless of being a new freshman. This linear observation, consequently, presents a significant and seemingly contradictory paradox when using data reduction on large data sets with many variables. Alternatively, using network analysis as a method for understanding these phenomena actually helps to explain why the paradox in the MLR model is taking place. Equally, this approach accounts for and describes anomalies at both the micro and macro levels, whereas only using the linear approach does not.

The results of this paper show an area of research that needs to be explored further. Although there has been much research on at-risk students within higher education, variable reduction in predictive modeling can lead to constraints in understanding interactions between variables and how they relate to at-risk students. This method explores these students through concurrent individual and group categorization, as well as both micro and macro levels of interaction within a network of observation. It is anticipated that the results of this study will show how critical it is to include research on at-risk students using additional forms of network analyses to create a broader picture of how we can contribute not only to persistence but also to the successes of at-risk students in higher education. Equally, the use of the method introduced in this article provides researchers of college student persistence and financial aid with a powerful approach to synthesizing the General Linear Model and complex systems theory in order to understand a much broader and more encompassing description of enrollment management processes.

References:

Bak, P. (1996). *How Nature Works: The Science of Self-Organized Criticality*. New York: Copernicus.

Bar-Yam, Y. (1997). *Dynamics of Complex System*. Reading, MA: Addison-Wesley.

Bean, J. P. (2003). College student retention. In J. W. Guthrie (Ed.), *Encyclopedia of Education* (2nd ed.). New York: Macmillan Reference USA.

Bertuglia, C. S., & Vaio, F. (2005). *Nonlinearity, Chaos, and Complexity: The Dynamics of Natural and Social Systems*. Oxford, UK: Oxford UP.

Brown, G. (2008). Higher education costs and the role of tuition. In Bob Bontrager (Ed.), *SEM and institutional success: Integrating enrollment, finance and student access* (pp. 33-46). Washington, D. C.: American Association of College Registrars and Admissions Officers.

Byrne, D. (1998). *Complexity Theory and the Social Sciences*. New York: Routledge.

Chapman, R. G. & Jackson, R. (1987). College choices of academically able students: The influence of no-need financial aid and other factors. *Research Monograph No. 10*. New York: College Entrance Examination, Board.

Cohen, B. H. (2013). *Explaining Psychological Statistics* (4th ed.). Hoboken, NJ: Wiley.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). London: Lawrence Erlbaum Associates.

Cytoscape 3.7.1. [Computer software]. (2019). Retrieved from https://cytoscape.org/

Davidson, W. B., Hall, P. B., & Grisaffe, D. B. (2015). Enhanced measurement and test of a nomological model. *Journal of College Student Retention: Research, Theory & Practice, 17*(2), 162-185.

Delen, D. (2011). Predicting student attrition with data mining methods. *Journal of College Student Retention: Research, Theory & Practice, 13*(1), 17-35.

Dowling, E. E. (1998). The student financial aid director's most difficult challenge: Un-learning financial aid 101. *College & University Journal,* (Summer/Fall), 11-15.

Eckles, J. E. & Stradley, E. G. (2012). A social network analysis of student retention using archival data. *Social Psychology Education, 15*, 165-180.

Estrada, E. (2012). *The structure of complex networks: Theory and applications*. Oxford, UK: Oxford University Press.Forsman, J., Linder, C., Moll, R., Fraser, D., & Andersson, S. (2014). A new approach to modeling student retention through an application of complexity thinking. *Studies in Higher Education 39*(1), 68-86.

Frantz, T. L. and Carley, K. M. (2009). Agent-based modeling within a dynamic network. In S. J. Guastello, M. Koopmans, & D. Pincus (Eds.), *Chaos and Complexity in Psychology* (pp. 475-505). Cambridge, UK: Cambridge UP.

Gilbert, N., & Troitzsch, K. G. (1999). *Simulation for the Social Scientist*. Berkshire, UK: Open University Press.

Gilstrap, D. L. (2013). Quantitative research methods in chaos and complexity: From probability to post hoc regression analyses. *Complicity: An International Journal of Complexity and Education 10*, 1. https://journals.library.ualberta.ca/complicity/index.php/complicity/article/view/20400

Gilstrap, D. L. (2007). Dissipative structures in educational change: Prigogine and the academy.

International Journal of Leadership in Education, 10(1): 49-69.

Guastello, S. J. (2009). Group dynamics: Adaptation, coordination, and the emergence of leaders. In S. J. Guastello, M. Koopmans, & D. Pincus (Eds.), *Chaos and Complexity in Psychology* (pp. 402-433). Cambridge, UK: Cambridge UP.

Holland, J. H. (1998). *Emergence: From Chaos to Order*. Reading, MA: Perseus.

Hossler, D. (2008). The public policy landscape: Financing higher education in America. Bob Bontrager (Ed.). *SEM and institutional success: Integrating enrollment, finance and student access*, (pp. 1-13). Washington, D. C.: American Association of College Registrars and Admissions Officers.

Hossler, D. (2002). Enrollment management in higher education: Defining enrollment management, key offices and tasks in enrollment management, organizational models. *Education Encyclopedia*, Bloomington, IN: StateUniversity.com.

Hossler, D., & Bean, J. P. (1990). *The Strategic Management of College Enrollments* (First ed.). San Franciso: Jossey-Bass.

IBM Corporation. Released 2017. IBM SPSS Statistics for Windows, Version 25.0, SPSS Neural Networks. Armonk, NY: IBM Corporation.

Kaufman, P., Bradbury, D., & Owings, J. (1992). *Characteristics of At-Risk Students in NELS:88*. Washington, DC: National Center for Educational Statistics.

Kauffman, S. A. (1995). *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity*. Oxford, UK: Oxford University Press.

Leslie, L. L. & Brinkman, P. T. (1987). Student price response in higher education: The student demand studies. *Journal of Higher Education, 58(2)*, 181-204.

Lewis-Beck, M. S., Bryman, A., & Futing Liao, T. (2004). *The SAGE encyclopedia of social science research methods.* Thousand Oaks, CA: Sage Publications, Inc.

MacIntosh, R., & MacLean, D. (1999). Conditioned emergence: A dissipative structures approach to transformation. *Strategic Management Journal, 20*(4), 297-316.

MacIntosh, R., & MacLean, D. (2001). Conditioned emergence: Researching change and changing research. *International Journal of Operations and Production Management, 21*(10), 1343-1357.

Manski, C. F., & Wise, D. A. (1983). *College choice in America*. Cambridge, MA: Harvard University Press.

Miller, T. E. (2010). Results of the use of a model that predicts individual student attrition to intervene with those who are most at risk. *College & University Journal, 85*(3), 12-19.

Miller, T. E. (2007). Will they stay or will they go: Predicting the risk of attrition at a large public university. *College & Research Journal, 83*(2), 2-15.

Mitchell, M. (2009). *Complexity: A Guided Tour*. Oxford, UK: Oxford UP.

Mitleton-Kelly, E. (Ed.) (2003). *Complex Systems and Evolutionary Perspectives on Organisations: The Application of Complexity Theory to Organizations*. Kidlington, Oxford UK: Elsevier Science.

Nelson, C. Van, Malone, B. G., & Nelson, J. S. (2001). Admission models for at-risk graduate students in different academic disciplines. *College & University Journal*, *77(2)*, 3-7.

Nimon, K. & Henson, R.K. (2015). Validity of residualized dependent variable after pretest covariance adjustments: Still the same variable? *The Journal of Experimental Education, 83*(3), 405-422.

Prigogine, I. (1967). *Thermodynamics of Irreversible Processes*. New York: John Wiley & Sons.

Prigogine, I. (1980). *From Being to Becoming: Time and Complexity in the Physical Sciences*. San Francisco: W. H. Freeman and Company.

Raju, D. & Schumacker, R. (2015). Exploring student characteristics of retention that lead to graduation in higher education using data mining models. *Journal of College Student Retention: Research, Theory & Practice, 16*(4), 563-591.

Santos, S. J. & Reigadas, E. T. (2004). Understanding the student-faculty mentoring process: Its effects on at-risk university students. *Journal of College Student Retention, 6*(3), 337-357.

Sawyer, R. K. (2007). Simulating Complexity. In W. Outhwaite & S. P. Turner (Eds.), *The SAGE Handbook of Social Science Methodology*. Thousand Oaks, CA: SAGE.

Simon, H. A. (1985). *The Sciences of the Artificial* (Second ed.). Cambridge, MA: MIT Press.

Waldrop, M. M. (1992). *Complexity: The Emerging Science at the Edge of Order and Chaos.* New York: Simon & Schuster.

White, D. R., & Johansen, U. (2005). *Network analysis and ethnographic problems: Process models of a Turkish nomad clan*. Oxford, UK: Lexington Books.

Wolff Smith, L. J. & Beretvas, S.N. (2015). A comparison of techniques for handling and assessing the influence of mobility on student achievement. *The Journal of Experimental Education, 85*(1), 3-23.

Table 1
Multiple Linear Regression Results

Variables	Variable Type	Beta	t	Sig	Zero Order	Partial	Part	Tolerance	VIF
Major by College	numeric(8)	0.031	4.172	**0.000**	0.093	0.036	0.030	0.894	1.118
Undecided major	numeric(1)	0.033	4.344	**0.000**	0.105	0.038	0.031	0.866	1.154
New fulltime freshman	numeric(1)	0.023	2.951	**0.003**	-0.057	0.260	0.021	0.874	1.144
New freshman transfer	numeric(1)	-0.024	-3.376	**0.001**	-0.030	-0.029	-0.024	0.969	1.032
New sophomore transfer	numeric(1)	-0.015	-2.087	**0.037**	-0.046	-0.018	-0.015	0.973	1.028
New junior transfer	numeric(1)	-0.019	-2.597	**0.009**	-0.046	-0.023	-0.019	0.988	1.010
New senior transfer	numeric(1)	-0.022	-3.097	**0.002**	-0.024	-0.027	-0.022	0.995	1.005
Enrolled hours	numeric(2)	-0.010	-0.533	0.594	-0.190	-0.005	-0.004	0.145	6.884
Full time enrollment	numeric(1)	-0.065	-5.338	**0.000**	-0.196	-0.046	-0.038	0.338	2.956
Student cumulative hours	scale	0.114	12.658	**0.000**	0.078	0.110	0.090	0.624	1.603

Cumulative GPA	scale	-0.223	-22.321	**0.000**	-0.420	-0.191	-0.159	0.510	1.961
Currently on probation	numeric(1)	0.234	28.589	**0.000**	0.352	0.242	0.204	757.000	1.322
Maximum ACT score	scale	-0.035	-4.374	**0.000**	-0.192	-0.038	-0.031	0.786	1.272
Maximum Adavanced Placement	scale	0.203	23.429	**0.000**	0.383	0.200	0.167	0.676	1.480
Age at time of 20th day into semester	scale	0.125	13.741	**0.000**	0.246	0.119	0.098	0.611	1.638
Female student	numeric(1)	0.003	0.446	0.656	-0.068	0.004	0.003	0.884	1.131
Race ethnicity	numeric(9)	-0.001	-0.098	0.922	0.010	-0.001	-0.001	0.960	1.042
Underrepresente d minority	numeric(1)	-0.019	-2.549	**0.011**	0.060	-0.022	-0.018	0.882	1.133
First generation student	numeric(1)	-0.028	-3.755	**0.000**	-0.013	-0.033	-0.027	0.946	1.057
Finanicial aid award	numeric(1)	0.001	0.520	0.958	0.010	0.000	0.000	0.069	14.395
Grant award	numeric(1)	-0.055	-4.142	**0.000**	-0.066	-0.036	-0.030	0.289	3.462
Loan amount	scale	0.019	0.836	0.403	0.130	0.007	0.006	0.099	10.128
Student net cost	scale	-0.053	-3.359	**0.001**	0.009	-0.029	-0.024	0.207	4.841
Total charges to student	scale	0.022	0.890	0.373	-0.165	0.008	0.006	0.087	11.539
Residential housing	numeric(1)	0.027	1.765	**0.078**	-0.058	0.015	0.013	0.211	4.735
End of term GPA	scale	-0.054	-6.117	**0.000**	-0.228	-0.053	-0.044	0.652	1.533
End of term hours completed	scale	0.030	2.437	**0.015**	-0.193	0.021	0.017	0.346	2.893

Table 2
Multicollinearity Diagnostics

Model Dimension	Eigenvalue	Condition Index	(Constant)	college division of major	undecided college major	new FT freshmen	new freshmen transfer std	new sophomore transfer std	new jun transfer std	new senior transfer std	enrolled hours	full time enrl std	cumulative hours	cumulative gpa	currently on probation
1	14.346	1.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	1.501	3.091	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	1.129	3.565	0.00	0.00	0.25	0.01	0.06	0.00	0.04	0.03	0.00	0.00	0.00	0.00	0.17
4	1.043	3.709	0.00	0.00	0.00	0.38	0.05	0.11	0.01	0.04	0.00	0.00	0.00	0.00	0.02
5	1.002	3.784	0.00	0.00	0.00	0.00	0.02	0.18	0.07	0.69	0.00	0.00	0.00	0.00	0.00
6	1.000	3.788	0.00	0.00	0.00	0.00	0.17	0.37	0.36	0.04	0.00	0.00	0.00	0.00	0.00
7	0.994	3.799	0.00	0.00	0.00	0.01	0.59	0.06	0.06	0.00	0.00	0.00	0.00	0.00	0.15
8	0.963	3.861	0.00	0.00	0.11	0.07	0.00	0.15	0.42	0.16	0.00	0.00	0.00	0.00	0.01
9	0.818	4.187	0.00	0.00	0.37	0.00	0.04	0.07	0.02	0.03	0.00	0.00	0.00	0.00	0.13
10	0.772	4.312	0.00	0.00	0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.19
11	0.652	4.691	0.00	0.00	0.08	0.36	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.02
12	0.530	5.203	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03
13	0.452	5.634	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
14	0.442	5.695	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01
15	0.389	6.070	0.00	0.01	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.05	0.01	0.00	0.00
16	0.225	7.986	0.00	0.00	0.01	0.02	0.01	0.01	0.00	0.00	0.00	0.05	0.10	0.00	0.04
17	0.171	9.147	0.00	0.20	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00
18	0.162	9.399	0.00	0.51	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.02
19	0.117	11.063	0.00	0.07	0.00	0.02	0.01	0.00	0.00	0.00	0.00	0.02	0.58	0.00	0.00
20	0.083	13.133	0.01	0.14	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.04	0.08	0.02	0.02
21	0.074	13.908	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.24	0.00	0.01	0.00
22	0.057	15.916	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.27	0.00	0.00	0.00
23	0.024	24.444	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.05	0.13	0.01
24	0.022	25.629	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.16	0.02	0.00	0.01
25	0.018	27.860	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.69	0.10
26	0.008	42.585	0.07	0.02	0.00	0.01	0.00	0.00	0.00	0.00	0.69	0.10	0.01	0.03	0.01
27	0.005	51.256	0.92	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.19	0.04	0.01	0.11	0.06

Table 2 Continued
Multicollinearity Diagnostics

Model Dimension	max act sortest	max APAP sortest	age at time of 20th day	female std	race ethnicity	undr rep minority	first generation std	fa sem grant amount	fa sem loan amount	student net cost	total charges to std	housing flag binary	end of term term gpa	end of term hrs completed
1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.00	0.05	0.00	0.00
3	0.00	0.01	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.00
5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
9	0.00	0.00	0.00	0.00	0.00	0.15	0.01	0.00	0.00	0.02	0.00	0.01	0.00	0.00
10	0.00	0.00	0.00	0.00	0.02	0.54	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00
11	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.08	0.00	0.10	0.00	0.00
12	0.00	0.03	0.00	0.00	0.02	0.07	0.73	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13	0.00	0.08	0.00	0.32	0.31	0.03	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00
14	0.00	0.13	0.00	0.09	0.59	0.02	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.00
15	0.00	0.06	0.00	0.17	0.01	0.04	0.08	0.00	0.01	0.02	0.00	0.06	0.00	0.01
16	0.00	0.30	0.00	0.26	0.00	0.04	0.03	0.01	0.00	0.00	0.00	0.00	0.04	0.00
17	0.00	0.04	0.00	0.01	0.01	0.00	0.02	0.01	0.45	0.00	0.00	0.00	0.05	0.00
18	0.00	0.09	0.00	0.08	0.00	0.00	0.00	0.00	0.15	0.00	0.00	0.00	0.05	0.01
19	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.22	0.01	0.00	0.00	0.17	0.00
20	0.06	0.01	0.01	0.00	0.02	0.00	0.01	0.07	0.08	0.06	0.00	0.03	0.21	0.04
21	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.56	0.00	0.26	0.00	0.04	0.04	0.03
22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.14	0.01	0.00	0.25	0.41
23	0.60	0.04	0.17	0.04	0.00	0.01	0.00	0.03	0.01	0.02	0.02	0.01	0.01	0.01
24	0.00	0.04	0.19	0.00	0.00	0.00	0.00	0.06	0.01	0.07	0.24	0.16	0.07	0.40
25	0.01	0.10	0.29	0.00	0.00	0.00	0.00	0.07	0.04	0.06	0.04	0.01	0.09	0.04
26	0.04	0.01	0.06	0.00	0.00	0.01	0.00	0.02	0.00	0.01	0.63	0.47	0.01	0.02
27	0.25	0.02	0.27	0.01	0.02	0.04	0.01	0.03	0.01	0.03	0.06	0.04	0.00	0.02

Figure 1
Network Map of Variables Interacting within the MLR Model
https://cytoscape.org/

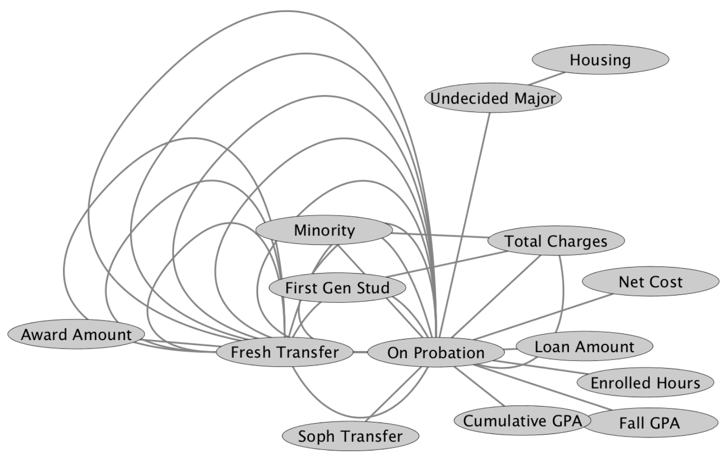

Figure 2
Network Map of Node Distances/Adjacency
https://cytoscape.org/