Classifying Organizational Forms in the Field of Higher Education *

Martin Ruef
Department of Sociology
Princeton University

Manish Nag
Department of Sociology
Princeton University

November 2011

* This is a draft version of the paper – please do not circulate or cite without the first author's permission. Address correspondence to Martin Ruef, Department of Sociology, Princeton University, Princeton, NJ 08544. E-mail: mruef@princeton.edu.

Introduction

In 1973, the Carnegie Foundation for the Advancement of Teaching (CFAT) first published its basic classification of degree-granting colleges and universities in the United States. Building on a long history of earlier efforts to survey and evaluate the diverse organizational forms in American higher education, a commission under the leadership of Clark Kerr sought to differentiate these institutions into five broad categories, as well as a number of more nuanced sub-categories (Carnegie Commission on Higher Education 1973). Kerr's own philosophy, adopted from John F. Kennedy, was to create an "aristocracy of achievement arising out of a democracy of opportunity" (Lagemann 1992: 230). Practically speaking, this meant that the classification offered by the commission would continue to distinguish the traditional "elite" universities, such as Harvard, Princeton, Stanford, Berkeley, and the like, while encouraging systemic expansion – and greater access – at lower levels of post-secondary education.

The Carnegie Classification emerged at a time when scholars of institutions and organizations had come to appreciate the increasing complexity and profound change that was evident in the field of higher education (Hodgkinson 1971; Parsons and Platt 1973; Clark 1972). The population of colleges and universities expanded rapidly over the preceding century, with merely 250 schools in the United States at the time of the Civil War and roughly ten times that number by 1970. The growth of the academic profession was especially pronounced in the period leading up to the commission's activities, doubling between 1960 (260,000 faculty members) and 1970 (530,000, including 383,000 full-time instructors) (Oakely 1997: 47; Thelin 2004). More subtle changes in the culture of the American university were also evident. While students in 19th and early 20th century institutions of higher learning were relatively insulated from broader societal developments, the social movements of the 1960s and the decline of *in loco parentis* norms created far more permeable organizational boundaries (Aldrich and Ruef 2006: 128). A proliferation of coursework and academic units in the social sciences, natural sciences, and applied fields undermined the traditional emphasis on humanities as the academic core of the university (Frank and Gabler 2006). Changes in admission policies produced a more diverse

student body (in terms of gender, ethnicity, and class), even at elite institutions (Karabel 2005). The CFAT's classificatory schema could thus be seen as one concerted effort to impose order on an expanding and increasingly heterogeneous array of campus settings.

As the growth of American higher education has continued unabated over the succeeding four decades, the Carnegie Foundation has repeatedly issued new classifications. The most recent system (issued in 2010) represents the sixth update to the original schema and features 33 categories in its "basic" classification. The benefits of these evolving categories for understanding higher education have been decidedly mixed. One historian of education, John Thelin, has commented that the CFAT's "attempt at creating order actually increased the chaos among institutions", insofar as a descriptive device for analyzing the field of higher education was converted – both by the public and some university administrators – into a "hierarchical ranking scheme" (2004: 320). The heuristic distinctions drawn by the Foundation became an invitation to game the classification, especially for some institutions that appeared in the lower rungs of the hierarchy and sought to pursue a more prestigious status. This dynamic was especially pernicious given the early impetus of the Carnegie Commission to promote diversity in post-secondary education by encouraging the founding of more accessible community and comprehensive colleges (McCormick and Zhao 2005).

Another important challenge for the classification of organizational forms in higher education involves the social scientific validity of these efforts. Beginning in the 1960s, a substantial literature in organizational studies has developed methods to elicit taxonomies of organizational types and practices. Many of these approaches have been *a posteriori*, allowing salient categories to emerge from detailed information on activities, structures, membership, and expressed identities within organizations. By contrast, an early critique of taxonomies of administrative structures (Pugh et al., 1969: 115) lamented that typologies up to that point had been "*a priori* classifications, based on wide generalizations derived from common knowledge and common sense, the only concession to empirical complexities being the admission that they are in some sense pure, ideal, or archetypal" (see also McKelvey 1982).

While the classification efforts of the CFAT – as well as similar schemata issued by the Southern Regional Education Board and American Association of University Professors (AAUP) -- have been resolutely empirical, they continue to rely on the *a priori* approach, in which an analyst or commission comes up with mutually-exclusive categories that structure distinctions among universities and colleges, rather than allowing the data to drive those categories. This raises a number of concerns. First, classification in higher education has become decoupled from recent organizational scholarship, which offers a range of theoretical perspectives and inductive tools for understanding the landscape of American colleges and universities. Second, the top-down, *a priori* imposition of categories may be particularly ill-suited to capture new or emergent organizational forms (e.g., alternative medical schools, distance learning colleges, for-profit universities, work colleges), owing to the institutionalization and taken-for-grantedness of existing classification systems. Third, an important development in recent work on organizations has been to recognize that membership in categories is often fuzzy and partial (Hannan 2010), rather than conforming to the crisp boundaries proposed by traditional approaches to classification. This holds true especially when organizational fields are in flux and audiences struggle to make sense of new organizations. Finally, existing approaches to classification in higher education are based on the intuitions of experts, rather than rigorous statistical models. A crucial goal of the Carnegie Classification, as stated by Kerr, was to generate categories that were "relatively homogeneous with respect to the function of the institutions as well as with respect to characteristics of students and faculty members" (McCormick and Zhao 2005: 52). Only a quantitative model can systematically assess the homogeneity of underlying categories or themes that are applied across several thousand organizations and, possibly, several hundred attributes.

To confront these shortcomings, this paper offers a new approach to classification in higher education which is grounded in contemporary organizational theory. We begin by surveying the literature on the development of organizational taxonomies, considering four distinct perspectives on the empirical basis of categories – (a) internal functions, routines, and structures;

(b) resource niches that support an organizational form; (c) the identity claims advanced by organizational leaders and members; and (d) the external attributions applied to organizations by field participants and the general public. For each perspective, we consider both how it has been applied to organizations in general, as well as how it has been used more specifically to understand developments in the field of education. The latter half of the paper then introduces a statistical model that provides a formal basis for implementing some insights from these perspectives on organizational classification. The intuition behind the model, termed Latent Dirichlet Allocation (LDA), is that organizations may have partial membership in a number of different categories, that those categories are not observed directly, and that there is a generative process whereby the observed attributes of organizations are produced by their membership in categories (Blei et al. 2003).

To bring our inductive model into dialogue with the Carnegie Classification of institutions of higher learning, we rely on the same data set that informs those efforts, the Integrated Postsecondary Education Data System (IPEDS). The IPEDS now provides a directory of over 6,800 post-secondary schools from a survey of Title IV institutions (and roughly 200 non-title IV voluntary submitters) and is collected annually by congressional order. Drawing on the IPEDS, we illustrate how our inductive model can be used to derive new sets of categories for the population of American colleges and universities and how those categories vary depending on the theoretical perspective used to understand differences among these institutions. We conclude by contrasting the classification systems derived inductively with the Carnegie Classification itself.

Approaches to Defining Organizational Forms

While common labels for organizational forms in the field of education – such as 'Ivy Leagues', 'community colleges', or 'state universities' -- suggest a well-established and intuitive understanding of the ways that higher education is structured and the distinct student populations that it caters to, the history of the Carnegie Classification reveals considerable contestation

around the basis for differentiating colleges and universities. The history of classification in organizational theory is no different. When researchers in the 1960s first proposed empirical approaches to deriving taxonomies of organizational forms (Haas et al. 1966; Pugh et al. 1969), they confronted an older tradition that had primarily been oriented toward understanding organizations in terms of ideal-types. The newer empirical approaches to studying organizational forms soon manifested their own points of divergence. Following Aldrich and Ruef (2006: Chapter 6), these approaches can be distinguished along two dimensions. The vertical dimension shown in Table 1 addresses the role of perception and considers whether a theoretical perspective treats organizational attributes as objective features or subjective interpretations on the part of observers. The horizontal dimension addresses the analyst's focus with respect to organizational boundaries, considering whether a perspective primarily employs a 'closed system' approach, emphasizing attributes that are internal to an organization, or an 'open system' approach, emphasizing the relationship of the organization to its broader environment (see also Scott and Davis, 2007).

[Insert Table 1 About Here]

Organizational Forms as Blueprints

The earliest approaches to defining organizational forms inductively can be placed in the upper left-hand cell of the table. Drawing on interviews with managers of 52 enterprises near Birmingham, England, the British Aston group sought to sort organizations based on features of their internal human resource practices, especially those related to the concentration of authority, the degree of formal structure in activities, and the line control of workflow (Pugh et al. 1969). Analyzing clusters of these features, the Aston group identified seven distinct categories of workplace structures, many of them deviating from the Weberian ideal-type of formal bureaucracy. Using a somewhat broader sample of organizations, a similar research effort was undertaken by Richard Hall and his colleagues in the United States (e.g., Haas et al. 1966). An emerging method with respect to organizational taxonomy thus appeared, emphasizing the

inductive derivation of categories based on surveys of internal practices across samples of organizations and multivariate analysis (see McKelvey 1975 for an overview and critique)

In a highly influential paper, Michael Hannan and John Freeman (1977) provided a theoretical rationale for this taxonomic approach. Arguing that organizational theorists had focused for too long on the adaptation of individual organizations, they called for a shift in the unit of analysis to organizational populations. The shift required that scholars "identify classes of organizations which are relatively homogeneous in terms of environment vulnerability" (ibid: 934). In an analogue to the study of genetic structure among population biologists, they suggested that the key to identifying these classes of organizations was to look inside organizations and study empirical differences in organizational form. For Hannan and Freeman, "an organizational form is a *blueprint* for organizational action, for transforming inputs into outputs" (ibid: 935). They went on to identify various internal features of organizations that might allow analysts to infer blueprints, including an organization's formal structure, routines, and normative order, where the latter feature was thought to be encoded in claims regarding the history of an organization, its politics, and the like.

The idea of classifying organizations in terms of internal, objective features has been carried forth under various labels, including the study of 'dominant competencies' (McKelvey 1982), 'grammars of action' (Pentland and Reuter 1994), and 'organizational genealogies' (Phillips 2002).[1] In the field of higher education, the application of such perspectives to classification is of a relatively recent vintage. Steven Brint and his colleagues (2006) launched an effort to map the "objective structure" of American colleges and universities, employing a cluster analysis to identify relatively homogeneous categories. Drawing from the Institutional Data Archive on American Higher Education (IDA), a survey of four-year university presidents, they considered

[1] In referring to such features as "internal", it is perhaps important to acknowledge that they may nevertheless be transferred from one organization to another. Indeed, an early critique of analogies between biological species and organizational populations was that the blueprints used to define the latter lacked the property of heritability (Betton and Dess 1985). A rich literature has subsequently developed to tackle the question of how personnel flows may transfer formal structure and routines from older to newer organizations (e.g., Phillips 2002).

such internal characteristics as the form of institutional control (e.g., public, nonprofit, religiously affiliated, independent), student selectivity, tuition, operating budget, and the extent of vocational training (% occupational or professional degrees). An analysis of these features yielded a classification schema with seven institutional clusters – ranging from elite private colleges and universities to relatively nonselective, religiously-affiliated baccalaureate-granting colleges (ibid: 235). Notably, Brint and his colleagues found that these inductively-derived clusters corresponded only loosely to the Carnegie Classification.

Organizational Forms as Resource Niches

In the 1980s, organizational theorists began to move away from the conception of organizational forms as internal structures and routines.[2] A number of methodological critiques had been raised with respect earlier attempts at inductive taxonomy. Replications of the Aston studies (e.g., Child 1972) raised questions about organizational sampling and the structural dimensions used to differentiate organizational forms. McKelvey (1975) argued, moreover, that the attributes selected in such studies tended to be too narrow – often deriving from a Weberian view of organizations as 'closed' bureaucracies – and that the observers selected to report on those attributes tended to be top administrators. A more inclusive effort at organizational taxonomy would also need to consider attributes reported by low-ranked members of organizations, even extending to external stakeholders, such as suppliers, customers, or clients (ibid: 517).

A shift in conceptualization was also evident in the ecological perspective that had provided much of the theoretical impetus for studying organizational forms as internal blueprints. Writing only a few years after Hannan and Freeman's initial statement on the population ecology of organizations, Miller McPherson sought to describe organizational forms in terms of their niches, "location[s] in multidimensional space defined by the resources in the environment" (1983: 520). McPherson eschewed an emphasis on internal features of organizational forms – such as size and

[2] One telling marker, in this respect, was the title of a 1979 article in *Administrative Science Quarterly*, which called for the *resurrection* of taxonomic approaches to organizational analysis (Pinder and Moore 1979), as pioneered in the 1960s by the Aston group and Haas and colleagues (1966).

the structural dimensions (formalization, centralization, etc.) that had come to be associated with it in the literature. Instead, he suggested that the ecology of organizations be understood in terms of a duality between forms and their demographic niches (see Mohr and Guerra-Pearson 2010 for an overview). This duality, in which "niches define forms and forms define niches", was soon picked up by Hannan and Freeman (1986: 57), who abandoned their earlier emphasis on internal organizational blueprints.

McPherson recognized that the boundaries of organizations were porous, noting that "individuals may be members of multiple organizations, or may enter or leave them repeatedly" (ibid: 519). In the face of such fluidity, an emphasis on internal structures and routines made less sense in defining organizational forms than an emphasis on the demographic profile of members that different forms might draw from. Applied to the field of higher education, for instance, this conception might seek to identify categories of universities and colleges based on the gender, age, ethnic, geographic, and class composition of their student body or applicant pool.

The conception of niches in ecology has also been broadened beyond the demographic composition of organizational forms. By the mid-1990s, for instance, ecological theorists were defining resource niches in terms of the "social, economic, and political conditions that can sustain the functioning of organizations that embody a particular form" (Hannan and Carroll 1995: 34). Other scholars, such as Paul DiMaggio, noted that resource dependencies could be captured in the network relationships of organizations and, as such, analysts could rely on "an operational definition of niche and form as mutually defined by observable patterns of relations among sets of actors" (1986: 360).

In recent years, these insights have begun to be deployed in the educational field. Linda Renzulli (2005) examines the emergence of the charter school form as a function of environmental conditions between 1991 and 1998. Analyzing the number of charter school applications across school districts in U.S. states with charter school legislation, Renzulli finds that this organizational form has thrived in niches with high levels of urbanization, supportive state laws,

a critical mass of nonreligious private schools, and a large proportion of nonwhite students. Extending such analyses of resource niches to higher education (with simultaneous consideration of multiple organizational forms) would represent a novel approach to classification.

Organizational Forms as Identities

Even as some scholars moved the definition of organizational forms outside the organization during the 1980s, others continued to privilege an internal perspective, but increasingly couched it in terms of culture and the subjective perceptions of members. In one widely-cited article, Albert and Whetten (1985) proposed a view of organizational forms as identities, revolving around the sense of members as to "who 'we' are"? In their formulation, such identities were rooted in features of the organization that were seen as central, enduring, and distinctive. Despite the durability of organizational identities, they were conceptualized as subject to claims-making and contestation. In the realm of higher education, for instance, some stakeholders characterize the mission of the university in reverent terms as a "church" of knowledge, while others view it more mundanely, as a "business" or system of vocational training (Frank and Gabler 2006).

At first glance, the claims of uniqueness that are implicit in organizational identities may appear to clash with efforts at classification, especially when those efforts are directed toward the identification of relatively homogeneous classes of organizations. But empirical investigations of organizational culture have found that assertions of uniqueness tend to be paradoxical, as notions of identity draw on standardized cultural templates or narratives that are widely rehearsed in society. For instance, an early study of narratives by Joanne Martin and colleagues (1983) found that seven stories used to highlight uniqueness could be found in a large variety of organizational contexts. Scholars of organizational identity now readily acknowledge that identities are hierarchical, with higher-order categories and organizational forms that are more central, more enduring, and more constraining than lower-order identities, which may offer greater uniqueness (Whetten 2006). The higher-order identities (e.g., Notre Dame's mission as a

Catholic university) impose the greatest switching costs and thus offer a suitable basis for organizational classification (ibid: 226).

A recent application of this perspective to education again focuses on charter schools, considering the emergence of this organizational form in Arizona between 1996 and 2001. Drawing on annual school report cards, Brayden King and his colleagues (2011) analyze how newly-founded schools construct their identities in this novel category. They note that "administrators craft the report cards to create public identities for their schools, broadcasting the schools' defining practices and policies and distinguishing the schools from their peers" (ibid: 557). A textual analysis of these mission statements reveals that schools commonly highlight social values, learning processes, aspects of curricular structure, and resources; they tend to downplay the demographics of their student body, including issues of ethnic identity. Based on the co-occurrence of these elements, King et al. find that the Arizona schools could be differentiated into two clusters. One cluster corresponds to an emerging organizational form that emphasizes vocational and social service programs; another cluster emphasizes creative and artistic learning that represents an alternative to conventional public school curricula. Like the early Aston group studies, the focus on the administrator statements in the report cards thus highlighted internal features of these organizations; but, because the statements were defined by the administrators themselves (rather than being elicited by social scientists), they offer a link between subjective identity and organizational form.

Organizational Forms as Cultural Codes

A final perspective on organizational forms continues to privilege the understandings of participants in the field, but moves the locus of perception from organizational insiders to include broader audiences. In a major revision of earlier theories of organizational forms, Hannan and his colleagues (2007) highlight the role of "audience segments", particularly where these segments achieve some consensus on *cultural codes* that allow them to classify organizations and sanction deviance from categorical schema. The description of audience segments in the theory

includes "insiders – the actual and potential members or employees of producer organizations – as well as various kinds of outsiders: buyers and suppliers, investors, critics, regulators" (ibid: 36). While the range of observers treated by the theory is thus quite encompassing, empirical analyses following this approach have tended to focus on external audiences of organizations, who are in the strongest position to evaluate and critique organizational behaviors that may not conform to their expectations.

An important aspect of the research on cultural codes involves the recognition that organizational membership in categories may be partial. Rather than examining classification systems with crisp boundaries, organizational theorists "usually study worlds in flux, with categories that emerge, transmute, and decay" (Hannan 2010: 160). As Michael Hannan has emphasized, the field of higher education is a particularly relevant example of a domain were categories are evolving and the mapping of universities and colleges to those categories ought to involve considerations of partiality. For instance, we might conceptualize the category of "university" itself as containing full-fledged, prototypical members, such as Stanford University and the University of North Carolina, but also consider other organizations that do not match the dominant conception of the category, such as Britain's Open University, which only offers distance learning (ibid). Audiences in higher education are especially likely to assign partial membership to newer forms of broad-access education, such as for-profit universities or "no frills" colleges, though that seems likely to change as perceptions and folk categories evolve.

The implications of partial category membership for organizations are well-documented in recent empirical scholarship. One consistent finding is that many audiences sanction those organizations that do not conform to cultural codes. For instance, firms that do not fall within standard industry categories and, therefore, are not followed consistently by a homogenous set of analysts suffer an 'illegitimacy' discount (Zuckerman 1999). Exceptions to this rule tend to obtain in contexts where categories have yet to be institutionalized, as reflected in a lack of trained observers, taken-for-granted systems of classification, and / or organizational routines and infrastructure for assigning organizations to categories (Ruef and Patterson 2009). Partiality may

also be beneficial to newly emerging forms, which must simultaneously signal some differentiation from existing organizational arrangements – or risk being subsumed by them – while also drawing on the legitimacy of established categories (Ruef 2000).

Another insightful aspect of the literature on cultural codes is that audience segments may include other organizations in a field: "producers themselves are [an] audience to each other" (Hannan et al. 2007: 36). Brint and his colleagues (2006) deploy this intuition to map the perceived structure of the field of higher education, drawing on the reference sets of universities that college presidents either believe to be similar to their own ("current reference set") or aspire to become ("aspiration reference set"). In the aggregate, the first set of comparisons thus allow us to view the cultural codes and boundaries that structure subjective categories formed by leaders in higher education, while the second set of comparisons address how well the aspirational identities of their institutions map onto those categories. Comparing the classifications that result from the current reference set of college presidents with other schema, the correspondence to institutions identified inductively through a cluster analysis of "objective" features in the IDA survey is high – e.g., 85% of presidential choices in a cluster of large research universities reference other universities within the same objective category. This statistic falls, however, for more peripheral categories in field of higher education. For instance, 56% of presidential choices in a cluster of nonselective baccalaureate-granting colleges reference other colleges within that category. Moreover, the ability of *a priori* typologies, such as the Carnegie Classification, to capture the cultural boundaries drawn by university presidents seems to be modest. When Brint et al. (2006: Table 2) apply the 2000 Carnegie codes to 270 institutions in their sample, they find that only 54% of the reference choices fall within the Carnegie categories on average. The fit is especially poor for less prototypical schools, such as the comprehensive colleges and universities in the MA II category (a mere 8% match with reference choices).

Summary

Our review of the literature on organizational forms suggests a rich array of perspectives and inductive tools for classifying organizations into categories, many of which have been applied in educational contexts. It also reveals dissatisfaction with a previous generation of *a priori* typologies, such as the Carnegie Classification, which sorted organizations into mutually-exclusive categories based on heuristic rules, rather than statistical criteria or theoretical considerations.

Nevertheless, our understanding of classification in higher education remains incomplete. Partially, this is a problem of data. Surveys of institutions of higher education are often limited to four-year colleges and universities, excluding two-year colleges, for-profit schools, and many specialized institutions (e.g., Brint et al. 2006). This inevitably leads to the exclusion of many newer organizational forms in the field, especially those devoted to broad-access education. Moreover, the attributes chosen to guide any particular classification schema tend to be small in number, often limited to one audience of organizational observers or otherwise constrained by one of the perspectives shown in Table 1. Following McKelvey (1975), we argue instead that the data used to inform organizational classification ought: (a) to sample from the broadest possible population of colleges and universities; (b) analyze institutional attributes and identity claims as inclusively as possible; and (c) address the viewpoints of multiple observers, including those internal to university and college administrations, as well as external stakeholders, such as prospective students and third-party evaluators (e.g., *U.S. News and World Report* rankings, *Princeton Review*, AAHE, CHEA, etc.).

The other problem with existing inductive approaches to organizational classification is one of modeling. We expect that the categories applied to institutions of higher learning will be relatively homogeneous, with a firm empirical foundation for the boundaries drawn between them. Early efforts at inducing taxonomies of organizations, like those of the Aston group, continued to rely on rules-of-thumb (e.g., means of dimensions in a factor analysis), rather than

statistical criteria for distinguishing among categories (McKelvey 1975). As noted above, the recent literature has also recognized that category membership may be partial, with 'hybrid' organizations that may be mapped to multiple categories (Albert and Whetten 1985; Hannan 2010). Traditional models of organizational classification, which emphasize discrete, mutually-exclusive categories, are ill-suited to represent such hybridity. Finally, the existing inductive models tend to fit categories closely to the clusters of organizational features that are observed in specific data sets, leading to problems of "overfitting" and a poor ability to extrapolate classification to new organizations. Given the rapid evolution of the field of higher education, it seems critical that any existing system of classification be able to accommodate new colleges and universities without redrawing category boundaries in an ad hoc fashion.

These considerations lead to three additional criteria for classification in higher education, wherein inductive models ought to: (d) identify relatively homogeneous categories of colleges and universities on a statistical basis; (e) allow some of these institutions to exhibit partial membership in multiple categories; and (f) permit analysts to systematically infer the classification of new kinds of colleges and universities, even when the data on those institutions were not available when the original system of classification was developed. We now turn to the preliminary development of some tools for organizational classification, with these criteria in mind.

Data and Model

Data Sources

Our sample of colleges and universities is drawn from the U.S. Department of Education's Integrated Postsecondary Education System (IPEDS), which is also employed by the Carnegie Foundation. IPEDS has a number of desirable attributes for purposes of developing systems of classification in higher education. On an annual basis, it collects data from every U.S. university, college, and vocational school that participates in federal student financial aid programs, as well

as a smaller number of schools that do not. While IPEDS does emphasize degree-granting institutions, the sample is extremely broad, covering organizations that range from research universities and state colleges to technical schools, for-profit universities, tribal colleges, and schools of cosmetology. Under Title IV of the Higher Education Act (1965), data reporting is mandatory for any institution where students may receive federal funding.

The set of school attributes reported for the IPEDS surveys is also extremely broad, covering institutional characteristics, demographics of enrolled students, faculty and staff composition and compensation, student financial aid, admission and test scores, graduation rates, and, in some years, mission statements. For purposes of exploratory analysis, we emphasize three clusters of variables that map closely onto the theoretical distinctions shown in Table 1. With respect to internal, institutional characteristics, we consider (1) institutional control (public, for-profit, secular NPO, religious NPO), degrees offered, forms of instruction (e.g., occupational, academic, continuing professional, etc.), special learning opportunities (distance learning, ROTC, study abroad, etc.), and whether a school accepts various forms of transfer credits. With respect to the resource niche of each school, we consider (2) the gender, race, age, international, and in-state demographics of the student body (fall enrollment), with each dimension differentiated by percentage quintiles across the IPEDS sample. Finally, for subjective claims of identity, we consider (3) the mission statements that were issued by each school, as reflected in statements either provided directly to IPEDS (up to 2,000 characters) or in school web pages. Table 2 illustrates these characteristics for a typical broad-access institution.

[Insert Table 2 About Here]

We impose some limitations on the scope of attributes used for organizational classification in the exploratory analyses. First, we do not consider any of the numerous performance metrics reported in IPEDS (esp., student test scores and completion rates), since we seek to separate the classification of institutions involved in higher education from efforts to evaluate them. Second, we do not consider subjective, external classifications of schools and universities by third parties.

While various observers within a school may be asked to respond to IPEDS surveys, this data collection effort does not ask for external attributions from third parties. Below, we consider how future research might incorporate such statements.

The following analyses focus on IPEDS data for schools in 2007, the most recent year when surveys collected mission statements from school administrators. For schools where mission data was not provided directly, mission statements were retrieved via a Google query that searched for "mission" or "about" in an institution's web pages. Harvesting data from the Internet was done using a web-crawling program for 1,100 schools. For data obtained from sites outside of IPEDS, data cleaning was required to remove HTML code and as much header, footer and navigation text as possible, in order to focus data entries on mission statements themselves. Data cleaning was performed both in an automated fashion and by hand.[3]

Model

We model the assignment of organizations to categories using a suite of algorithms termed *probabilistic topic models*, focusing in particular on Latent Dirichlet Allocation (LDA), the simplest kind of topic model (Blei et al. 2003). To motivate this approach, we begin by assuming that the categories in a classification schema are defined as a probability distribution over a set of attributes or words used to describe organizations. For example, a category of "medical schools" in higher education might be associated with objective features, such as having a hospital, and identity claims regarding professional competence, each with a high probability. The same category might also have a very low probability of being linked to other attributes, such as remedial adult education or identity claims regarding environmental stewardship.

Following the intuition of Blei (2011), we then assume that a description of a specific organization sampled from a population is produced in a three-stage process: (1) the

[3] Following listwise deletion, we have data on the institutional characteristics of 6,902 schools, on the fall enrollment demographics of 6,761 schools, and on the mission statements of 4,359 schools (including identity claims retrieved from school web pages).

organization itself is characterized as being distributed over categories (which may involve exclusive membership in a single category or partial membership in multiple categories); (2) for each attribute or identity claim involving the organization, a relevant membership category is chosen at random (subject to the distribution in [1]) and, then, (3) a specific feature is chosen at random from the category's vocabulary of attributes (subject to the category selected in [2] and the pre-existing distribution of attributes or words linked to the classification schema). So, if the Southwest College of Naturopathic Medicine has a partial membership in the category of medical schools (e.g., 0.9) and a (much smaller) partial membership in the category of environmental and naturopathic programs (0.1), then there is a 0.9 probability that an identity claim in the college's mission statement will be selected from those that are typical of other medical schools.

The methodological challenge for LDA is that only the attributes or identity claims linked to organizations are observed in any given sample, while the underlying categories are latent (i.e., hidden) and must be inferred from those associations. To formalize the model, we let $\beta_{1:K}$ correspond to the K latent categories (where β_k is the distribution over a vocabulary in a category), $\theta_{1:M}$ correspond to the category memberships for the M organizations in a sample (where θ_m identifies the category membership for the *m*th organization), $z_{1:M}$ enumerate the categories assigned to individual attributes used to describe the organizations (where $z_{m,n}$ is the category for the *n*th attribute and the *m*th organization), and $w_{1:M}$ enumerate the words that are actually observed in the descriptions of the organizations (where $w_{m,n}$ is the word given to the *n*th attribute and the *m*th organization). With this notation, Blei (2011) notes that the generative process for LDA is given by the following joint distribution:

$$p(\beta_{1:K}, \theta_{1:M}, z_{1:M}, w_{1:M}) = \prod_{k=1}^{K} p(\beta_k) \prod_{m=1}^{M} p(\theta_m) \left(\prod_{n=1}^{N} p(z_{m,n} \mid \theta_m) p(w_{m,n} \mid \beta_{1:K}, z_{m,n}) \right) \quad (1)$$

The LDA procedure relies on hierarchical Bayesian modeling to fit categories to the observed attributes or identity claims of organizations. Bayesian modeling attempts to calculate a *posterior distribution* of the parameters that might generate the data observed. In LDA, the two

key parameters are *Dirichlet* distributions (an extension of the beta distribution often used as priors in lower dimensional Bayesian models). The first key Dirichlet distribution is θ, the distribution of categories over organizations, which is sampled from in order to assign how much of an organizational description is devoted to a certain category. The second Dirichlet distribution is β, which is sampled from to assign the likelihood that an attribute or identity claim is devoted to a certain category. The interaction of these two parameters with the other multinomial distributions (z and w) shown in Equation 1 results in the assignment of each attribute in an organizational description to a single category. The assignment of each attribute to a category still allows for organizations to be assigned to multiple categories. It also allows for polysemy -- multiple instances of the same attribute in an organizational description that are assigned to different categories when distinct meanings are expressed.

Calculating posterior distributions in Bayesian modeling involves calculating integrals. However, the high dimension integrals involved in hierarchical models like LDA are not directly calculable and must instead be approximated. The use of Markov Chain Monte Carlo (MCMC), and especially the Gibbs sampler, has provided a necessary tool for the proliferation of Bayesian methods. In MCMC methods, repeated samples are taken from a given complex distribution, and the values of a previous sample draw determine subsequent sample draws, with the process continuing until convergence is found. Though our study utilizes the MCMC method of Gibbs sampling, variational inference methods have also been applied to LDA (Blei and Jordan 2005).

Regardless of the methods used to approximate posterior distributions, what LDA provides in terms of tangible data to organizational researchers are *category attribute assignments*. In essence every attribute in an organizational description is assigned to a category at a given probability. This category assignment helps determine the *categorical membership* for each organization. From category attribute assignments, terms (unique attributes) can be scored in regard to their relevance to a category. From this scoring, term lists can be analyzed by researchers to analyze whether a category is meaningful. The following discussion of the application of LDA to the IPEDS data set provides an example of the outputs mentioned above.

Results

For the sake of comparison, we set the number of categories (K) to eighteen forms of organizations involved in post-secondary education, equal to that used in the 2000 Carnegie Classification system.[4] Tables 3 through 5 summarize the preliminary results of categories derived through probabilistic topic modeling, focusing on internal institutional characteristics, student demographics, and mission statements, respectively.

[Insert Tables 3 through 5 About Here]

Considering the classification of schools by internal institutional characteristics (Table 3), we find that there is considerable homogeneity among the elite research universities, coupled with great diversity among lower-tier, broad-access institutions. In contrast to the Carnegie Classification, which draws fine-grained distinctions among research universities (extensive and intensive), the LDA-derived schema places these institutions in a single category. Among the next tier of institutions, the LDA schema employs approximately as many distinctions as the Carnegie system. For instance, comprehensive colleges and universities are divided into two categories, with another category for religious liberal arts colleges. The LDA schema adds a new category of "professional schools", which subsumes a variety of specialized institutions emphasizing graduate-level education in the Carnegie classification (e.g., schools of law, graduate schools of business and management, etc.). At the Baccalaureate level, the LDA schema distinguishes three categories, like the Carnegie Classification, albeit with a stronger emphasis on technical schools (Technical and Art Institutes I and II, Liberal Arts Schools).[5]

[4] Although the number of categories in an LDA classification is essentially arbitrary, calibrating it with an existing classification of colleges and universities offers two methodological advantages. First, it allows analysts to evaluate whether an inductively-derived system of categories explains more variance in some outcome than an *a priori* system, such as the Carnegie Classification, without adjusting model fit for the number of categories. Second, it allows differences between an inductively-derived set of categories an *a priori* system to be evaluated directly, based on either the "meaning" attached to categories or the mapping of colleges and universities to them.

[5] One distinction between the two technical school categories appears to hinge on the fact that the first group tends to highlight distance and on-line learning opportunities.

The more striking divergence between the two classification systems occurs among institutions that do not offer at least a bachelor's degree. The 2000 Carnegie Classification applies a single category ("Associate's Colleges") to a diverse range of community, junior, and technical colleges, as well as other schools offering postsecondary vocational training. Yet in 2007, these organizations comprised nearly one quarter of the entire population of schools in American higher education. Based on internal variation in school structure and pedagogical routines, the LDA model breaks this group into nine categories. For instance, community colleges are distinguished as public institutions that offer basic adult (e.g., GED) and recreational education.[6] In the interest of brevity, we will not discuss the differentiating features of other categories in the LDA classification here. But the general inference is clear: *the statistical variation in internal structure and pedagogy among non-baccalaureate-granting institutions requires a more nuanced classification than that presumed by the older Carnegie Classifications*.

An LDA analysis of the demographics of student populations across U.S. campuses reveals some similarities and some differences from the classifications induced from institutional characteristics alone (see Table 4). The research and state universities are placed into two categories, both of which are characterized by their large size and relatively high enrollments of international students. A distinguishing feature between them is the extent to which they encourage the enrollment of part-time students. The classic liberal arts colleges have student bodies that are slightly smaller, younger, and less likely to originate from the same state as their schools. Along with men's military schools, these categories capture a good deal of the variation in the demographic niches among higher-tier institutions.

As was the case for institutional characteristics, the LDA model proposes a large number of categories to accommodate the heterogeneity in student demography among broad-access institutions. For instance, in a set of categories that we label as "career colleges", the students

[6] Of course, these broad-access institutions are known for their vocational programs as well. But this feature does not differentiate them very clearly from the other schools offering associate's degrees and postsecondary certificates.

are older, more likely to be female, and, often, more likely to be minorities than the students found in traditional institutions of higher education. Many of these colleges prepare their graduates for careers in health care, information technology, nursing, paralegal work, and the like. Another set of categories that we label as "men's vocational schools" are oriented toward male students, but also feature more minorities (category I) or older students (category II) than those found on traditional college campuses. The training in these institutions varies from esoteric pursuits – such as golf course management and cinematography -- to automotive technology, CAD drafting, and HVAC repair.

Our inductive analysis of mission statements (Table 5) yields the greatest number of categorical distinctions among traditional institutions of higher learning and the lowest number of categories for broad-access institutions. The mission statements of universities differentiate between a category of institutions that emphasize research and those that highlight the diversity and values of their students. The identity claims of liberal arts colleges fall into three categories: those that embrace a global mission, those that highlight a classic liberal arts curriculum, and those that advocate progressivism in spheres such as social justice or the advancement of women. Baccalaureate and post-graduate schools with a religious identity (particularly, Christian colleges and Talmudical seminaries), also stand out in analyzing these mission statements.

The broad-access institutions, by contrast, tend to emphasize more mundane and practical concerns in describing themselves, with a focus on careers, technology, and training. Based on the LDA analysis, these institutions fall into ten categories, including art schools, career colleges, cosmetology programs, institutes for massage therapy, medical technology programs, nursing schools, technical schools, trade schools, and (two forms of) community colleges. Compared to the inductive analyses of institutional characteristics and demographics, the mission statements tend to differentiate broad-access institutions by career tracks rather than institutional control, pedagogy, or student diversity.

Discussion

In the interest of stimulating new approaches to the classification of educational institutions, this conference paper has provided an overview of frameworks that address categorization in organizational theory, as well as a probabilistic model (LDA) that allows these frameworks to be applied to empirical data on trade schools, colleges, and universities. Preliminary results suggest that the inductive LDA model may be well-suited to categorize a variety of broad-access institutions. Nevertheless, optimism must be tempered by the exploratory nature of this research. Far more work needs to be done to assess the reliability of categories derived using probabilistic topic modeling, assessing category consistency in the face of changing sets of attributes, samples of educational institutions, and organizational observers. With respect to the latter, our analyses have relied exclusively on attributes reported by university and college administrators, as well as members of the student population. An important supplement to the IPEDS data base would consist of external (particularly, qualitative) assessments of postsecondary institutions, such as those offered by "college guides" and other third-party observers.

The construct validity of inductively-derived classifications must also be examined in greater detail. In comparing these approaches to *a priori* schema for the classification of institutions in higher education, the implicit claims are that LDA models will create more "meaningful" categories for purposes of peer comparison and more "explanatory" categories for purposes of analyzing educational outcomes. The first claim can be investigated by interrogating the folk taxonomies used by administrators themselves (i.e., who do they identify as their peer institutions). The second claim can be assessed by considering the ability of the inductive classifications to explain student admissions, completions, financial aid, and job placements, as well as organizational outcomes such as research productivity, graduation rates, fiscal integrity, and reputation. If inductive approaches tend to explain more variance in such outcomes, and prove more meaningful to university and college administrators, it may be time to jettison the Carnegie Classification in favor of alternative perspectives on organizational classification.

REFERENCES

Albert, Stuart and David Whetten. 1985. "Organizational Identity," Pp. 263-295 in L. Cummings and B. Staw (eds.), *Research in Organizational Behavior*. Greenwich, CT: JAI Press.

Aldrich, Howard and Martin Ruef. 2006. *Organizations Evolving* (2nd Edition). London: Sage.

Betton, John and Gregory Dess. 1985. "The Application of Population Ecology Models to the Study of Organizations," *Academy of Management Review*, 10: 750-757.

Blei, David, A. Ng, and M. Jordan. 2003. "Latent Dirichlet Allocation," *Journal of Machine Learning*, 3: 993-1022.

Blei David and M. Jordan. 2005. "Variational Inference for Dirichlet Process Mixtures," *Journal of Bayesian Analysis*, 1:121-144.

Blei, David. 2011. "Introduction to Probabilistic Topic Models," Working Paper, Department of Computer Science, Princeton University.

Brint, Steven, Mark Riddle, and Robert Hanneman. 2006. "Reference Sets, Identities, and Aspirations in a Complex Organizational Field: The Case of American Four-Year Colleges and Universities," *Sociology of Education*, 79: 229-252.

Carnegie Commission on Higher Education. 1973. *A Classification of Institutions of Higher Education*. Berkeley, CA: Author.

Child, John. 1972. "Organization Structure and Strategies of Control: A Replication of the Aston Study," *Administrative Science Quarterly*, 17: 163-177.

Clark, Burton. 1972. "The Organizational Saga in Higher Education," *Administrative Science Quarterly*, 17: 178-184.

DiMaggio, Paul. 1986. "Structural Analysis of Organizational Fields: A Blockmodel Approach," Pp. 335-370 in B. Staw and L. Cummings (eds.), *Research in Organizational Behavior*. Greenwich, CN: JAI Press.

Frank, David John and Jay Gabler. 2006. *Reconstructing the University: Worldwide Shifts in Academia in the 20th Century*. Stanford, CA: Stanford University Press.

Haas, J., R. Hall, and N. Johnson. 1966. "Toward an Empirically Derived Taxonomy of Organizations," Pp. 157-180 in R.V. Bowers (ed.), *Studies on Behavior in Organizations*. Athens, GA: University of Georgia Press.

Hannan, Michael. 2010. "Partiality of Memberships in Categories and Audiences," *Annual Review of Sociology*, 36: 159-181.

Hannan, Michael and John Freeman. 1977. "The Population Ecology of Organizations," *American Journal of Sociology*, 82: 929-964.

Hannan, Michael and John Freeman. 1986. "Where Do Organizational Forms Come From?" *Sociological Forum*, 1: 50-72.

Hannan, Michael and Glenn Carroll. 1995. "An Introduction to Organizational Ecology," Pp. 17-31 in G. Carroll and M. Hannan (eds.), *Organizations in Industry: Strategy, Structure, and Selection*. New York: Oxford University Press.

Hannan, Michael, Laszlo Pólos, and Glenn Carroll. 2007. *Logics of Organization Theory: Audiences, Codes, and Ecologies*. Princeton, NJ: Princeton University Press.

Hodgkinson, Harold. 1971. *Institutions in Transition: A Profile of Change in Higher Education*. New York: McGraw-Hill.

Karabel, Jerome. 2005. *The Chosen: The Hidden History of Admission at Harvard, Yale and Princeton*. Boston, MA: Houghton Mifflin.

King, Brayden, Elisabeth S. Clemens, and Melissa Fry Konty. 2011. "Identity Realization and Organizational Forms: Differentiation and Consolidation of Identities among Arizona's Charter Schools," *Organization Science*, 22: 554-572.

Lagemann, Ellen. 1992. *The Politics of Knowledge: The Carnegie Corporation, Philanthropy, and Public Policy*. Chicago: University of Chicago Press.

Martin, Joanne, Martha Feldman, Mary Jo Hatch, and Sim Sitkin. 1983. "The Uniqueness Paradox in Organizational Stories," *Administrative Science Quarterly*, 28: 438-453.

McCormick, Alexander and Chun-Mei Zhao. 2005. "Rethinking and Reframing the Carnegie Classification," *Change*, 37: 50-57.

McKelvey, Bill. 1975. "Guidelines for the Empirical Classification of Organizations," *Administrative Science Quarterly*, 20: 509-525.

_____.1982. *Organizational Systematics: Taxonomy, Evolution, Classification*. Berkeley, CA: University of California Press.

McPherson, Miller. 1983. "The Ecology of Affiliation," *American Sociological Review*, 48: 519-532.

Mohr, John and Francesca Guerra-Pearson. 2010. "The Duality of Niche and Form: The Differentiation of Institutional Space in New York City, 1888-1917," *Research in the Sociology of Organizations*, 31: 321-368.

Oakely, Francis. 1997. "The Elusive Academic Profession: Complexity and Change," *Daedalus*, 126: 43-66.

Parsons, Talcott and Gerald M. Platt. 1973. *The American University*. Cambridge, MA: Harvard University Press.

Pentland, Brian and Henry Rueter. 1994. "Organizational Routines as Grammars of Action," *Administrative Science Quarterly*, 39: 484-510.

Phillips, Damon. 2002. "A Genealogical Approach to Organizational Life Chances: The Parent-Progeny Transfer among Silicon Valley Law Firms, 1946-1996," *Administrative Science Quarterly*, 47: 474-506.

Pinder, Craig and Larry Moore. 1979. "The Resurrection of Taxonomy to Aid the Development of Middle Range Theories of Organizational Behavior," *Administrative Science Quarterly*, 24: 99-118.

Pugh, D., D. Hickson, and C.R. Hinings. 1969. "An Empirical Taxonomy of Structures of Work Organizations," *Administrative Science Quarterly*, 14: 115-126.

Renzulli, Linda. 2005. "Organizational Environments and the Emergence of Charter Schools in the U.S.," *Sociology of Education*, 78: 1-26.

Ruef, Martin. 2000. "The Emergence of Organizational Forms: A Community Ecology Approach," *American Journal of Sociology*, 106: 658-714.

Ruef, Martin and Kelly Patterson. 2009. "Credit and Classification: The Impact of Industry Boundaries in 19th Century America," *Administrative Science Quarterly*, 54: 486-520.

Scott, W. Richard and Gerald F. Davis. 2007. *Organizations and Organizing: Rational, Natural, and Open Systems Perspectives*. Englewood, NJ: Prentice Hall

Thelin, John. 2004. *A History of American Higher Education*. Baltimore, MD: Johns Hopkins Press.

Whetten, David. 2006. "Albert and Whetten Revisited: Strengthening the Concept of Organizational Identity," *Journal of Management Inquiry*, 15: 219-234.

Zuckerman, Ezra. 1999. "The Categorical Imperative: Securities Analysts and the Illegitimacy Discount," *American Journal of Sociology*, 104: 1398-1438.

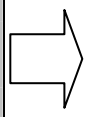
Table 1. *Inductive Approaches to Defining Organizational Forms*

Focus with respect to role of perception	Focus with respect to organizational boundaries: Internal	Focus with respect to organizational boundaries: External
Objective	**'Blueprints'** **Typical Method:** Surveys of Internal Structures and Routines **Examples:** Haas et al. 1966; Pugh et al. 1969; Brint et al. 2006 *	**'Resource Niches'** **Typical Method:** Analysis of Conditions or Relationships Supporting Organization **Examples:** McPherson 1983; DiMaggio 1986; Renzulli 2005 *
Subjective	**'Organizational Identities'** **Typical Method**: Interpretation of Mission Statements and Self-Depictions **Examples:** Martin et al. 1983; Albert and Whetten 1985; King et al. 2011 *	**'Cultural Codes'** **Typical Method:** Analysis of Public Discourse or External Classifications **Examples:** Zuckerman 1999; Ruef 2000; Hannan 2010 *

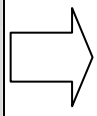
Source: Table adapted from Aldrich and Ruef (2006: 115).
Note: Asterices (*) identify analyses that are oriented toward the field of education.

Table 2. *Reported Characteristics for Sample College in IPEDS Database*

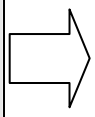
Name:	**Shelton State Community College** (2007)
Location:	**Tuscaloosa, AL**
Carnegie Classification:	Associates College
Institutional Control:	Public
Degrees Awarded:	Certificates (up to two years), Associate's Degrees
Educational Offerings:	Academic, Occupational, Recreational, Adult Basic
Special Learning Opportunities:	Distance Learning, ROTC, Weekend / Evening Classes
Transfer Credits:	Dual Credits, AP Credits
Other Institutional Characteristics:	Has Library, HBCU
Gender Composition:	54.3% Female
Racial Composition:	64.7% White, 29.1% Black, 1.1% Hispanic, 5.2% Other
Age Composition:	47.5% Under 22; 36.3% Between 22-29; 16.2% Over 29
International Students:	0.1%
In-State Students:	99.5%
Mission Statement:	Shelton State Community College is a public open-admission comprehensive community college whose primary mission is to provide accessible postsecondary education, training, and community educational opportunities.



Internal Structures and Routines



Demographic Niche



Organizational Identity

Table 3. *LDA Categories Inferred from the Internal Structure and Routines of U.S. Postsecondary Institutions* (Data Source: IPEDS Module on Institutional Characteristics)

Category	Attributes	Institutional Examples
Associate Colleges	Public, Associate's Degree, Transfer Credits, Distance Learning	Eastern Arizona College, Sacramento City College
Community Colleges I	Public, Adult Basic & Recreational Instruction, No Bachelor, Transfer Credits, Distance Learning	Glendale Community Asnuntuck Community, Community College of Aurora
Community Colleges II	Public, Adult Basic & Recreational Instruction, No Bachelor, Transfer Credits, Distance Learning	Rich Mountain Community, Los Angeles Pierce College, Housatonic Community
Comprehensives I (Career-Focused)	Bachelor or Master's Degree, Teacher Certification, Study Abroad, Transfers	Saint Joseph College, University of Miami
Comprehensives II (Liberal Arts)	Bachelor or Master's Degree, Teacher Certification, Study Abroad, Transfers	Husson College, Concordia University, St. Paul
Cosmetology and Med Tech Schools	Private For-Profit, No Bachelor or Advanced Degrees, No Transfer Credits	Arkansas Beauty College, First Institute
Cosmetology Schools II	Private For-Profit, No Bachelor or Advanced Degrees, No Transfer Credits	New Tyler Barber College, Elegance International
District and System Offices	Not applicable	City Colleges of Chicago, U-Hawaii System Office
Liberal Arts Schools	Bachelor's Degree, Teacher Certification, Transfer Credits, Study Abroad	Knox College, Lycoming College
Professional Schools	Private NP, Master's / Professional Degree, No SLO, No Transfer Credits	Southwestern Law School, Fielding Graduate University
Religious Liberal Arts Schools	Private NP-Religious, Bachelor or Master's Degree, Teacher Certification, Transfer Credits	Campbellsville University, College of the Holy Cross, Gordon College
Research Universities	Bachelor, Master, and Doctoral Degrees, Teacher Certification, AP Credit	University of Idaho Drake University
Technical and Art Institutes I	Bachelor but no Advanced Degrees, Transfer Credits, Distance Learning	ITT Technical Institute Devry University
Technical and Art Institutes II	Private For-Profit, Bachelor or Associate's Degree, AP Credits	Indiana Business College, New England Inst. of Art
Trade Schools I	Private For-Profit, Two-Year Certificates, No SLO, No Transfer Credits	Refrigeration School, Taylor Business Institute
Trade Schools II	No Bachelor or Advanced Degrees, No SLO, No Transfer Credits	Bridgerland Applied Tech, Everest College-Reseda
Trade Schools III	Private For-Profit, No Bachelor or Advanced Degrees, No SLO, No Transfer Credits	South Coast College, Stenotype Institute
Vocational Schools	No Bachelor or Advanced Degrees, No SLO, No Transfer Credits	Lincoln Technical Institute, Marinello School of Beauty

Note: "SLO" refers to special learning opportunities, including distance learning, ROTC, study abroad, teacher certification, and weekend / evening classes.

Table 4. *LDA Categories Inferred from the Demographic Niches of U.S. Postsecondary Institutions* (Data Source: IPEDS Module on Fall Enrollments)

Category	Attributes	Institutional Examples
Career Colleges I	Medium Size, More Women, Older Students	Stautzenberger College, Career Technical College
Career Colleges II	Small Size, More Women, Oldest Students, Many Asian Students	Brown Mackie College, Indiana Business College
Career Colleges III	Medium Size, More Women, Older, Many Black and Hispanic Students	Concorde Career College, St. Louis College of Health Careers
Community Colleges I	Very Large Size, Many International, Part-Time Students	Riverside Community College, Butler Community College
Community Colleges II	Large Size, Part-Time Students, Few Asian Students	Appalachian Technical College, Edison State Community
Community Colleges III	Very Large Size, Part-Time, International Students	Cumberland County College, Neumann College
Community-Oriented Colleges	Large Size, Part-Time, International Students	North Florida Community, Buena Vista University
Continuing Ed Colleges I	Very Small Size, Oldest Students, Part-Time Students	Southeastern Business College, South Texas Barber College
Continuing Ed Colleges II	Small Size, Oldest Students	Antioch University, California Career College
Liberal Arts Colleges	Large Size, Young Students, Few In-State Students	Occidental College, Colorado College
Men's Military Schools	Medium Size, More Men, Young Students	Massachusetts Maritime Academy
Men's Vocational I	Small Size, More Men, Many Hispanic and Asian Students	Golf Academy of the Carolinas, Tennessee Technology Center
Men's Vocational II	Medium Size, More Men, Older Students	American Film Institute, Pennco Tech
Universities I	Very Large Size, Many International Students	University of South Alabama, CSU-Sacramento
Universities II	Very Large Size, Many International, Some Part-Time Students	Indiana State University, SUNY at Albany
Women's Vocational I	Very Small Size, Mostly Women, Many Hispanic and Asian Students	Professional Choice Hair, Dayton School of Hair
Women's Vocational II	Very Small Size, Mostly Women, Many Hispanic and Asian Students	Associated Technical College, Artistic Beauty College
Women's Vocational III	Very Small Size, Mostly Women, Many Minority Students	Toni & Guy Hairdressing, California Hair Design

Table 5. *LDA Categories Inferred from Mission Statements of U.S. Postsecondary Institutions* (Data Source: IPEDS Module on Institutional Characteristics) *

Category	Identity Claims	Institutional Examples
Art and Music Schools	School, Design, Music, Department, Art, College, Accrediting, Association	Hussian School of Art, Conservatory of Recording Arts
Career Colleges	Career, Skills, Training, Technical, Employment, Provide, Business, Job	College America, Medina County Career Center
Christian Colleges	Christian, God, Church, Seminary, Theological, Ministry, Jesus, Faith	Boise Bible College, Lutheran Theological Seminary
Community Colleges I	College, Community, Programs, Technical, Services, Associate, Transfer	Estrella Mountain Community, Wilkes Community
Community Colleges II	College, Community, Services, Quality, Needs, Support, Accessible, System	Coffeyville Community, South Piedmont Community College
Cosmetology Schools	Cosmetology, Beauty, State, Field, School, Pass, Training, Hair, Industry	Fayetteville Beauty College, Award Beauty School
Globally-Oriented Colleges	Global, Community, World, Values, Knowledge, Develop, Society, Diversity	Salem International University, Lafayette College
Liberal Arts Schools I	College, Arts, Liberal, Professional, Learning, Student, Personal, Diverse	Wheaton College, James Madison University
Liberal Arts Schools II	College, Human, Arts, Women, Justice, Commitment, Liberal, Intellectual	Siena College, Albright College
Massage Schools	Massage, Therapy, Providing, Quality, Dedicated, Highest, Institute, Graduates	New York Institute of Massage, E. Grady School of Esthetics and Massage Therapy
Medical Schools	Healthcare, Medical, Program, Research, Professional, Clinical, Practice	Jefferson College of Health, Academy of Oriental Medicine
Medical Tech Schools	Center, Medical, Engineering, Computer, State, Science, Student	Cleveland Institute of Dental-Medical Assistants, McLeod Regional Med. Center School
Nursing Schools	Nursing, Healthcare, Practice, Promote, Needs, Demonstrate, Competent, Skills	Episcopal School of Nursing, Medcenter 1 College of Nursing
Research-Oriented Universities	University, Research, State, Graduate, Undergraduate, Public, Programs	University of Mississippi, Eastern Kentucky University
Student-Oriented Universities	Community, Learning, University, Excellence, Values, Student, Diversity	Berkeley City College, Cameron University
Talmudical Seminaries	Understanding, Jewish, Seek, Moral, Ethical, Help, World, Means, Build	Telshe Yeshiva, Yeshiva Toras Chaim Talmudical Seminary
Technical Schools	Provide, Quality, Technology, Care, Employees, Services, Health, Focused	Chubb Institute, High-Tech Institute
Trade Schools	School, Training, Law, Industry, Career, Skills, Hands, World, Program, Classes	Tulsa Welding School, New England Culinary Institute

* Excludes seven residual categories, with low mission statement proportions or highly heterogeneous vocabulary.