# VISUALIZING RELATIONAL NETWORK EFFECTS AMONG NORTH AMERICAN COLLEGES AND UNIVERSITIES

MIKAILA MARIEL LEMONIK ARTHUR

This analysis considers self-defined peer relationships within a network of 1,292 North American colleges and universities. The analysis explores elements of network structure, including degree distribution, average shortest path, and modularity. These factors can be useful for better understanding the decisions that colleges and universities make about their self-defined peers as well as for better understanding the clustering of colleges and universities by type and/or geography. Additionally, this analysis considers the way that network structure and proximity might affect curricular offerings by examining the presence of academic programs in women's studies among network members.

## OBTAINING DATA

The data included in this analysis comes from an ongoing project examining curricular change in higher education in the United States (specifically, the emergence of tertiary-level academic programs in women's studies, Asian American studies, and queer/LGBT studies). In particular, network data is important to this study because it allows for the possibility of testing whether diffusion, mimetic effects, and/or institutional isomorphism shape the curricular change decisions of particular colleges and universities. Prior research in this area has tended to focus on diffusion across categories of institutions—looking, for instance, at whether colleges and universities in the same locality or who meet certain structural criteria will have similar curricular offerings ("positional peers"). Therefore, this particular study was design to instead explore the effects of higher education institutions' own self-defined peers (here referred to as "relational peers"). Accrediting agencies in the United States require each college or university to define a list of peer institution that the college or university will use for comparative analysis. Each college and university has substantial latitude to define its peers in the manner that it chooses: for instance, peers may be defined on the basis of membership in the same state university system, similarity in student demographics, geographical co-locality, or aspiration (the colleges or universities a particular institution wishes to become like).

To develop the network data, first a random sample of 60 colleges and universities was drawn. To be included in the sampling frame for this sample, a college or university must have received at least one dollar of federal research funding in the period 1997 through 2003 and it must have awarded four-year undergraduate degrees. Each of the 60 colleges and universities was contacted directly and asked for a list of its own self-defined peers; the resulting list contained 525 colleges and universities. Each of these colleges and universities was then contacted and asked for a list of its own self-defined peers. Some colleges and universities provide their peer lists to the Department

of Education's Integrated Post-Secondary Data Analysis System (IPEDS) as part of their annual reporting process; these lists were obtained where possible for those colleges and universities which did not respond to direct contacts. Between direct contacts and the IPEDS system, peer lists were not obtained for 12 colleges and universities. Therefore, **nodes** in this dataset represent individual colleges and universities who qualified for inclusion either by being part of the initial random sample of 60 colleges and universities or by being included on one or more lists of peer institutions. **Edges** represent self-defined peer relationships; they are directional, as peer relationships are not necessarily reciprocated. Instances in which a college or university indicated a peer ambiguously (for instance, listing "University of California" instead of a specific campus in the UC System) are excluded from the analysis, as are institutions offering only graduate or specialized study (such as divinity schools).

The resulting dataset contains 1,292 colleges and universities in North America. Data on institutional and curricular characteristics was obtained from available data, including the IPEDS system, the Carnegie Corporation, and CollegeSource (an online database of college and university course catalogs). The data utilized in this analysis contains a subset of the variables in the complete dataset; the only outcome variable included is the presence of a women's studies program (major, minor, concentration, or certificate) in 2005, while other variables (all as of 2005) include:

- Geographic data: state, zip code, and Census 9-region classification code;
- Demographic data: total enrollment, percentage of enrollment consisting of undergraduate students, percentage of faculty female, and percentage of students female;
- Institutional classifications: public, private non-profit, or private for profit classification; Carnegie classification (using the 2000 classification system);
- Admission data: admission rate and yield rate; and
- Financial data: tuition and net assets.

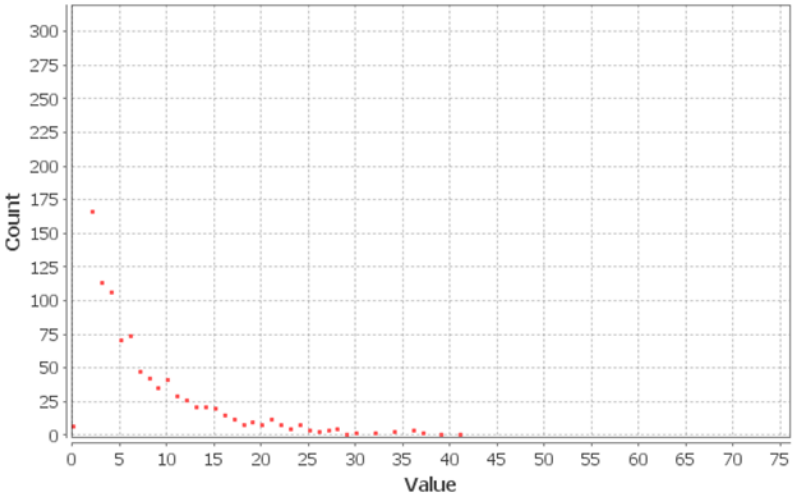
The resulting dataset was analyzed using Gephi network visualization software and SPSS statistical software; the Gephi and SPSS files are available for download from the author's website at http://www.ric.edu/faculty/marthur/research.html.

## ANALYSIS

### *Network Structure*

The average degree in the network is 6.173[1]. However, a closer examination of in-degree and out-degree is warranted. The distribution of in-degree and out-degree are markedly different. As shown at left, the in-degree distribution roughly follows the pattern of an inverse exponential function, with many nodes having low in-degree; while only 7 have in-degree of 0, hundreds

FIGURE 1. IN-DEGREE DISTRIBUTION

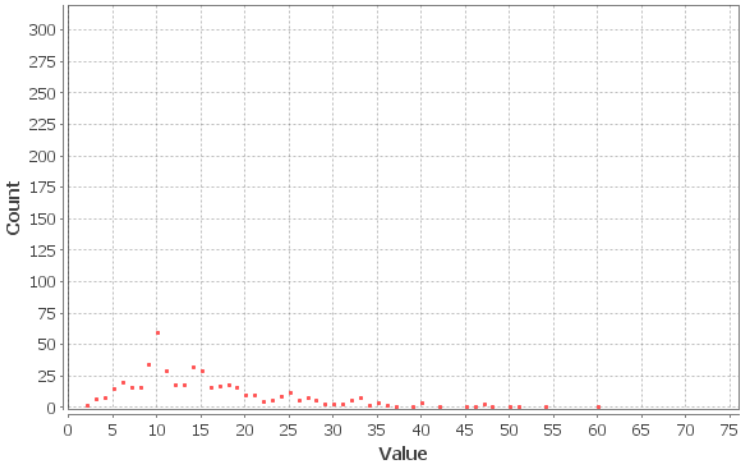



[1] To compute in-degree and out-degree, use the Gephi procedure *statistics* → *average degree*. In-degree and out-degree for individual nodes can then be obtained in the Data Laboratory.

have in-degree between 1 and 5, and the median in-degree is 4. Just a few having high in-degree. Eight institutions have in-degree of over 35; seven of these are flagship public universities[2] and one is a selective liberal arts college[3].

In-degree, in this analysis, is a representation of how many other colleges and universities have selected a given college or university as a peer. There are, of course, a variety of strategies undergirding peer selection, such as the selection of competitors for student enrollments, members of the same athletic consortium, those structurally or demographically similar in terms of IPEDS or other metrics, and "aspirational peers" (those colleges or universities that a given campus would like to emulate in its quest for greater status and prestige). This variation makes it more difficult to draw a clear conclusion about what it means to be a college or university with high in-degree, but we can assume that such institutions are generally regarded as successful and worth competing with or emulating.

FIGURE 2. OUT-DEGREE DISTRIBUTION




The out-degree distribution, on the other hand, is much flatter, with some clustering around 10. This may be due to the fact that while there are no specific and universal requirements for how many peers a college or university must select, there are costs and limitations to selecting too few or too many. For example, the selection of too few peers makes it difficult for Institutional Research offices to draw substantial or significant comparisons to practices elsewhere, and thus only 14 of the colleges and universities which have indicated peers have fewer than five such peers. On the other hand, with too many peers, data collection and analysis can become much more challenging; however, four institutions have fifty or more self-defined peers[4]. Additionally, 723 of the institutions in the dataset were incorporated in the last phase of data collection and thus did not indicate their own peers, the median and modal out-degree is 0.

The average shortest path is 4.78[5]. Interpretation of the average shortest path in the context of the diffusion of organizational innovations must consider the fact that communication between organizations is considerably more constrained than communication between individuals. For example, were this a network of individuals considering whether to adopt a new technology, simple person-to-person communication might be enough to enable diffusion of the innovation. In

[2] In order from lowest to highest in-degree, University of Minnesota Twin Cities, University of North Carolina Chapel Hill, University of Texas Austin, Ohio State University, University of Wisconsin Madison, University of Michigan Ann Arbor, and, at in-degree 41, University of Illinois Urbana-Champaign
[3] Carleton College, MN
[4] In order from lowest to highest out-degree, Kentucky Wesleyan College, University of the Pacific, Boston University, and, with out-degree 60, the University of Missouri St. Louis
[5] Average shortest path can be obtained via the Gephi procedure *statistics → average path length.*

contrast, diffusion of an innovation across organizations requires active information-gathering as well as decision-making involving a complex network of responsible parties.

### *Network Effects*

The network can be partitioned into nine modularity classes[6]. These modularity classes do in fact roughly correspond to real divisions within the institutional field of higher education. As Table 1 below demonstrates, the proportion of colleges and universities in each modularity class which offer women's studies programs varies greatly, with such programs being non-existent in the two community college modularity classes and rare among Christian and for-profit colleges while being almost universal among selective public universities and selective liberal arts colleges. This pattern is statistically significant at $p<0.001$ with a fairly strong association (Cramer's $V=0.501$). A graph of the network colored by modularity class and labeled by college or university name can be found in Appendix A.

TABLE 1. MODULARITY CLASSES AND WOMEN'S STUDIES PROGRAMS

| Modularity Class Number & Organizational Type | Percentage of Colleges & University in Modularity Class | Percentage of Those in Modularity Class with a Women's Studies Program |
|---|---|---|
| **0) Moderate Prestige & Smaller Private Colleges** | 10.76% | 48.2% |
| **1) Selective Liberal Arts Colleges** | 6.04% | 92.3% |
| **2) Selective Public Universities** | 13.47% | 87.9% |
| **3) North Carolina Community Colleges** | 2.01% | 0% |
| **4) Less Selective Mid-Atlantic Colleges** | 3.41% | 31.8% |
| **5) Selective Private Universities** | 14.4% | 74.2% |
| **6) Less Selective Public Universities** | 27.79% | 54.6% |
| **7) Christian and For-Profit Colleges** | 21.21% | 22.3% |
| **8) Missouri Community Colleges** | 0.93% | 0% |
| **Overall Network** | 100% | 54.3% |

Another way to visualize the distribution of women's studies programs across the network is to color the nodes by the presence or absence of women's studies. A network graph showing this distribution is available in Appendix B. Looking at the network graph provides an intuitive understanding of the pattern of adoption of women's studies. The graph shows that there is a central "spine" of colleges and universities in the network among which women's studies programs are almost universal, while at the peripheral edges—which include less prestigious and less selective institutions—such offerings are less likely. The primary exception to this pattern is a swath of colleges and universities not offering women's studies that appear at the center of the graph in the cluster of selective private universities. On closer inspection, this set of colleges and

[6] Modularity classes can be obtained via the Gephi procedure *statistics* → *modularity*. The modularity classes of individual nodes can then be obtained in the Data Laboratory, and in the Graph view, the graph can be partitioned and colored by modularity.

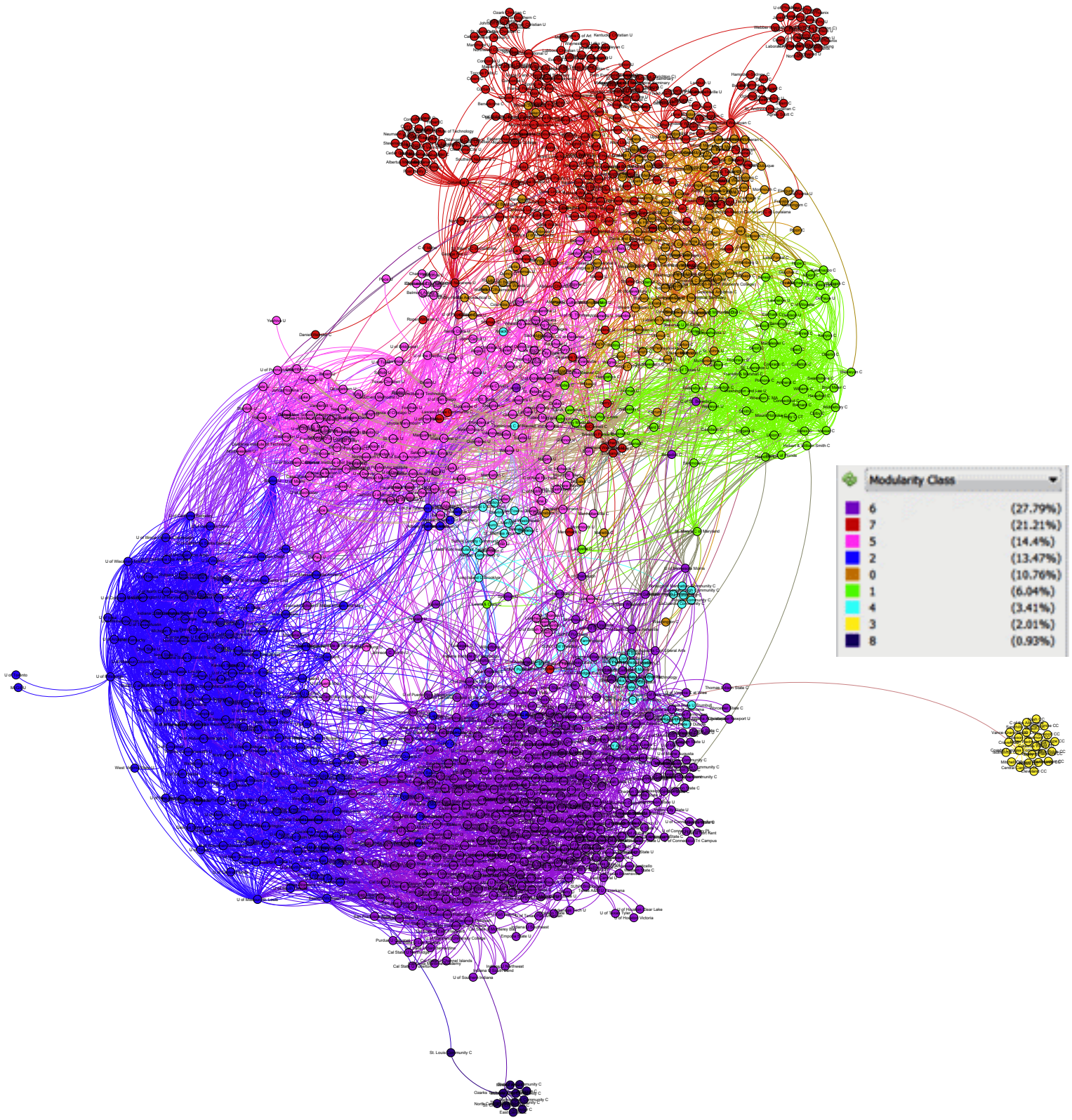
universities consists primarily of those falling into one of two categories: Catholic universities and technical or engineering institutions. Presumably, the fact that this set of colleges and universities does not offer women's studies programs accounts for the lower proportion of women's studies offerings in modularity class 5.

## INTERPRETATIONS

The data utilized in this analysis does have limitations. Most importantly, the colleges and universities included in the dataset do not capture the complete network of colleges and universities in the United States. This limitation occurs due to the fact that data was collected for only two hops out from the initial seed colleges and universities. The dynamics present in this limited network may not be the same as those which would be present if the entire population of colleges and universities were included. In particular, specialized institutions, tribal colleges, and community colleges are underrepresented in this analysis. Additionally, this analysis is limited because of the narrow scope of the data included. The larger-scale project will look at several additional outcome variables, will consider geographical network effects by geocoding the data, and will explore change over time by examining whether the presence of an academic program in 2005 is predicted by the percentage of peer colleges and universities that offered that program in 1995.

Despite these limitations, this analysis suggests several analytical insights. Considering in-degree and out-degree for self-defined peer networks enables a better understanding of the considerations colleges and universities have as they develop their lists of peers. It also enables the analyst to determine which colleges and universities—and which *types* of colleges and universities—are most frequently seen as appropriate for comparison and/or emulation. Furthermore, examining the modularity class distribution as determined analytically from the data provides further support to the oft-cited notion that there are "multiple worlds" of higher education, rather than all colleges and universities belonging to a single institutional field. Finally, this analysis provides support for the idea that network structure and self-defined peer relationships affect the adoption of organizational innovations by considering the way that the presence of women's studies programs is clustered within the network. Women's studies programs are more likely to be present among the more central and high-prestige segments of the network, while they remain quite rare among the peripheral nodes. Further analysis will explore the degree to which diffusion is responsible for this pattern and whether the pattern differs when less common innovations are considered.

# Appendix A: Network Colored by Modularity Class



# Appendix B: Network Colored by Presence of Women's Studies Program