Article

The Curriculum Prerequisite Network: Modeling the Curriculum as a Complex System

Preston R. Aldrich*

From the Department of Biological Sciences, Benedictine University, Lisle, Illinois

Abstract

This article advances the prerequisite network as a means to visualize the hidden structure in an academic curriculum. Networks have been used to represent a variety of complex systems ranging from social systems to biochemical pathways and protein interactions. Here, I treat the academic curriculum as a complex system with nodes representing courses and links between nodes the course prerequisites as readily obtained from a course catalogue. I show that the catalogue data can be rendered as a directed acyclic graph, which has certain desirable analytical features. Using metrics developed in mathematical graph theory, I characterize the overall structure of the undergraduate curriculum of Benedictine University along with that of its Biochemistry and Molecular Biology program. The latter program is shown to contain hidden community structure that crosses disciplinary boundaries. The overall curriculum is seen as partitioned into numerous isolated course groupings, the size of the groups varying considerably. Individual courses serve different roles in the organization, such as information sources, hubs, and bridges. The curriculum prerequisite network represents the intrinsic, hard-wired constraints on the flow of information in a curriculum, and is the organizational context within which learning occurs. I explore some applications for advising and curriculum reform. © 2015 by The International Union of Biochemistry and Molecular Biology, 43(3):168–180, 2015.

Keywords: *curriculum prerequisite network; directed acyclic graph; network; systems biology*

Introduction

There is widespread agreement that undergraduate science curricula in the United States need broad-scale reform [1–5]. A principal challenge is that many of the most active research fronts lie between the traditional boundaries of academic disciplines, and so the successful scientist must be able to harness resources in disjunct areas. This is especially true of biology which increasingly relies on multidisciplinary, integrative approaches to solve problems, as noted by AAAS Vision and Change in Undergraduate Biology Education: A Call to Action [5]. The old model is no longer tenable, wherein each discipline offers up its own siloed information, and it is left to the student to synthesize this information after being exposed to the separate information silos. Overall, the United States finds that it could do a better job at addressing this 21st century challenge. For example, it is clear that mathematics is not sufficiently integrated into our biology curricula compared to other nations [4].

The goal should be the production of integrated and coherent curricula, ones in which different topical areas reinforce and inform one another. Such efforts are underway in, for example, biomathematics [6] where many of the innovations consist of integrating math and biology content within courses. There is also a rising interest in the integration of courses across the curriculum [5], evidenced in part by the heightened popularity of forming learning communities [7, 8]. Still, this work generally is done without any comprehensive, global perspective on existing curricular organization, beyond what is readily available in a course catalogue.

All colleges and universities have a course catalogue. They are data-rich, and show course content and prerequisite mappings between courses. Unfortunately, course catalogues also are slow to yield information on relationships beyond one or two steps removed. Searchable electronic formats have improved their utility, yet the type of

Ⓢ *Additional Supporting Information may be found in the online version of this article.*

**Address for correspondence: Department of Biological Sciences, Benedictine University, Lisle, IL, USA. E-mail: paldrich@ben.edu.*

Received 29 August 2014; Accepted 16 January 2015

DOI 10.1002/bmb.20861

Published online 24 February 2015 in Wiley Online Library (wileyonlinelibrary.com)

information we end up with is still very much the same as it probably has been since the course catalogue was invented; we still read the catalogue as though it were a book.

As faculty and administrators, we need to leverage 21st century technology in order to interact with curriculum data more effectively. Not only must we keep up with current trends in our respective fields, such as how to interact with "big data" using the latest NCBI web resources, but we also must learn to interact with our "curricular big data" in new ways. Relevant here is course catalogue information along with enrolment and transcript information. We should use modern approaches that have proven useful for studying complex systems like the cell. Fortunately for biochemists and molecular biologists, this technology is already near at hand.

Networks offer a powerful paradigm for the visualization and analysis of complex systems [9, 10]. In biology, there are three disciplines that have a long and storied tradition in network representations, these being taxonomy and phylogenetics beginning with Linnaeus [11] and Darwin [12], food web ecology beginning with Elton [13], and biochemistry exemplified by the works of Krebs [14]. By 1840, the field of biochemistry was already grappling with a complicated morass of compounds and their interconversions [15], and by 1965, with the publication of Roche's first biochemical pathways wall charts, was entertaining the visual representation of an extremely complex system. Meanwhile, the field of mathematical graph theory was developing in a separate information silo [16, 17], and toward the end of the 20th century, physicists began applying tools from mathematical graph theory to the study of naturally occurring complex networks [18, 19]. These efforts revealed the ubiquity of general architectural features shared by social networks, power grids, brain neural networks, and networks of the cell. It is now commonplace to use networks to assemble whole genomes using de Bruijn graphs [20], to study whole transcriptomes [21], whole proteomes [22], and whole metabolomes [23], and represent drug-protein target interactions as networks [24], and the genic basis of illnesses using diseasome networks [25]. Reductionism gives way to holism.

Despite the importance placed on assessment in education, we lack a global, holistic assessment of the organization of our curricula. As we engage our national discussion on reshaping curricula for 21st century science, we lack a modern scientific view of the current shape of our curricula. Perhaps the closest we have is the process of "curriculum mapping" [26–28], which is a technical innovation from the field of curricular studies that has gained popularity [29–32]. Curriculum mapping is a template-based procedure in which faculty identify content as it is delivered during a term, and then share this information with other faculty. Curriculum reform might then follow based on this global information. The approach is especially popular in K-12 education. Unfortunately, the information is typically organized in tables and spreadsheets that are not inherently amenable to quantitative analysis or to visual assessment of global features.

So far, networks have not played a prominent role in curriculum assessment, though there have been attempts amidst a variety of other technical approaches for visualizing the knowledge domains of academic institutions. As early as 1984, student enrolment data in Drexel University's School of Library and Information Science were analyzed using multidimensional scaling to generate a two-dimensional map of the top forty two elective courses [33]. By 1990, concept maps were being used to improve a science curriculum [34]. By 2007, there began an upsurge in interest regarding curriculum visualization methods with most reports appearing in conference proceedings [35–40]. For example, the computer science program at Ball State University was visualized as a directed graph that was used during advising and recruitment events to explain the curriculum [36]. There is no published study to date that has explored as the primary object of study the network structure of course prerequisites for an entire college or university curriculum. The subject is interesting because it regards the fundamental constraints on information flow within an academic institution, and could serve as a "roadmap" to guide administrators and faculty as they build and revise a curriculum, and for students as they try to navigate the curriculum.

Research Questions

Here, I explore the Curriculum Prerequisite Network, or CPN, which is a network view of the knowledge system contained in a university course catalogue. Nodes (or vertices in graph theory) represent courses, and directed links (or arcs) between courses represent prerequisite requirements. The entire system is a directed graph or digraph.

I chose to work on the Benedictine University course catalogue because this is the curriculum with which I have the most familiarity, and this has aided my interpretation of the findings. Clearly, few would be intrinsically interested in the structure of any one institution unless it was their own. The goal here is, rather, to reveal a process for assessing curricular architecture in a holistic fashion using modern technologies that could be applied elsewhere and on a larger scale. Eventually, it will be interesting to compare the network structure of multiple institutions, but just as the field of comparative genomics had to await the sequencing of the first genome, Benedictine University serves as a reasonable place to begin this project. Research questions fall into four categories:

Method Feasibility

What issues are encountered when coding a CPN from a university course catalogue? Registrars are notoriously meticulous about maintaining consistent catalogue syntax, so it is a reasonable expectation that some of the coding might be automated. What other catalogue elements beyond the prerequisite link require attention? Answers to these questions will indicate whether this approach would be applicable to other institutions.

DAG Topology

The second question concerns a specific aspect of the global topology, or shape, of a CPN; in its native state, is it a directed acyclic graph (DAG)? A DAG is a type of directed network that does not contain any cycles, i.e., one cannot leave a node and then later return to it [41]. If a native CPN is not a DAG, can a DAG architecture be produced without compromising the integrity of the catalogue information? This speaks to the amenability of CPNs to analysis by certain types of machine learning methods.

Global Topology

What is the topological structure of the CPN for an entire undergraduate curriculum? Of how many different parts is it comprised, how do they fit together, and how might this influence the flow of information? How are courses differentiated with respect to their position within a curriculum, and how might we represent this structure using metrics from mathematical graph theory? These questions address the type of information one is likely to get from the study of a CPN, of this or any other institution.

Program-Level Topology

Lastly, what sort of information is available when we drill down to the programmatic level and examine fine-scale linkages between collections of courses. For heuristic purposes, I focus on the Biochemistry and Molecular Biology program at Benedictine University. Is the level of integration desirable? Is the spacing between topics reasonable? Which are the most central courses? Which control the flow of information? Might a CPN aid in the navigation of the catalogue rules that organize the curriculum? These questions address factors that could influence the effectiveness of program-level revision efforts and the quality of student advising.

Methodology

From Catalogue to Network

The course catalogue of Benedictine University, Lisle, IL, was used for this study. Benedictine University is an independent, Catholic, comprehensive, 501I(3) institution of higher education, category 17 (DRU: Doctoral/Research Universities) of the Carnegie BASIC2010 classification. The main campus is in the Chicago metropolitan area, and serves a racially, ethnically, and religiously diverse student population of 7,434 students (56% undergraduates, 44% graduate). It offers 53 undergraduate majors, 12 masters programs, 2 PhD programs, and an EdD program. A primary strength of the institution is its well-established and highly regarded science programs.

Information from the undergraduate course catalogue for 2009 to 2010 was scraped from the university website using Python (see Supporting Information). After processing, the information was converted into network format using NetworkX (version 1.8), a Python package for the analysis of complex networks [42].

As an example, Fig. 1 shows a portion of a hypothetical course catalogue along with the CPN that would result. The dashed arcs are not coded from the catalogue shown, but are included in the network for heuristic purposes to illustrate mappings that would introduce cycles into the CPN (see Supporting Information). General rules and conventions follow.

Prerequisites

The prerequisite binding—course A must be taken before course B—establishes a parent-daughter, predecessor-successor relationship between the courses, one that is readily modeled as a directed graph. Although one could specify B $\rightarrow$ A, where B is referencing information obtained previously in course A, I chose the alternate method that better represents the flow of information from A into B, the arc A $\rightarrow$ B. If more than one prerequisite was specified as mandatory, then equal and full weights were applied to each arc. For example, if both course A and B were required before one could take course C, then arcs A $\rightarrow$ C and B $\rightarrow$ C were both given a weight of 1.0. However, if prerequisite rules presented an option such that, for example, either course A or B (but not both) was required before C, then the total weight of 1.0 was distributed equally across arcs (e.g., A $\rightarrow$ C, 0.5; B $\rightarrow$ C, 0.5).

Corequisites

A hard corequisite binding—coregistration in course X—establishes a symmetric relationship between courses, particularly if both courses reference the other as a corequisite. Here, it generally is assumed that a student will enrol for both courses A and B in the same term, as is typical of lecture/lab combinations in the sciences, or more broadly when a theory course is temporally bound to a practical applications course. We can represent this symmetric binding with a bidirectional arc between the courses in the CPN (A $\leftarrow\rightarrow$ B, Fig. 1), with equal weights in both directions (A $\rightarrow$ B, 1.0; B $\rightarrow$ A, 1.0). However, this bidirectional edge introduces a 2-member cycle, and the CPN no longer can be treated as a directed acyclic graph (DAG).

A soft corequisite binding—credit or coregistration in course X—establishes a slightly asymmetric relationship between courses, particularly if only one course of the two names the other as a corequisite. Although most students likely will take A and B together, the option exists to take course A first and B later. This allowance of temporal priority also suggests a certain amount of conceptual priority. One might ask, if a student were to take the courses in

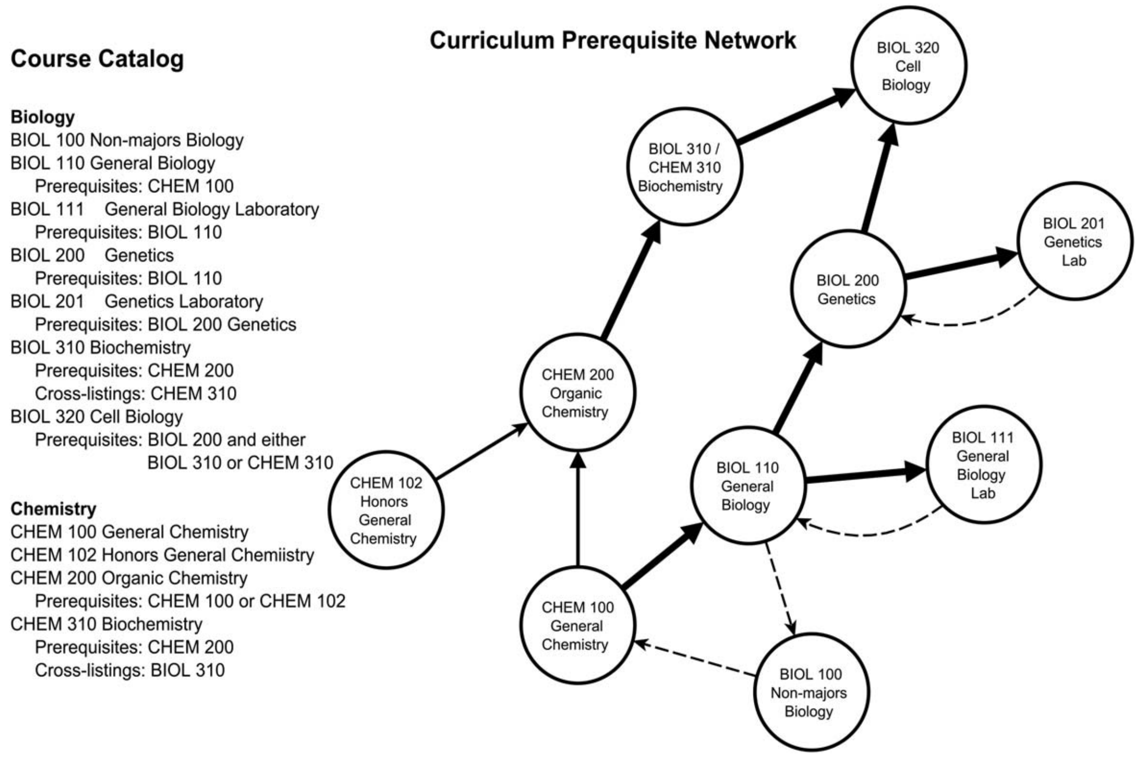

FIG 1 *Hypothetical course catalogue and the associated CPN. Rendered using yEd (http://www.yworks.com).*

different terms, which should come first? I expect most would agree that a lecture should precede a lab, not the other way around. By this argument, one might treat some corequisite courses as conceptually, if not temporally, prior in the curriculum.

In the Benedictine University catalogue, this was a reasonable interpretation since all corequisite pairings contained one course (A) that served as a "lecture" or "theory" course and the other (B) as a "laboratory" or "applied" course. Consequently, I elected to interpret all corequisite relationships as soft corequisite bindings, allowing their coding as prerequisites ($A \rightarrow B$), thus preserving the DAG structure; this was not critical to the analyses performed in the present study. The hypothetical example in Fig. 1 shows two instances of corequisite bindings between a lecture and a lab, wherein a cycle is introduced if a bidirectional arc is used, whereas DAG structure is preserved if only the lab carries the co(pre)requisite.

Cross-Listings

The cross-listing establishes an equivalency between courses A and B. One might choose to keep courses A and B separate in a CPN, if they were indeed distinct courses that served a similar but not identical role in a curriculum. A more literal reading of the catalogue would be to treat them as one and the same course. I chose the latter option and merged cross-listed course nodes, creating a composite node (A/B) inheriting all of the bindings of the original nodes.

Hard-wired Versus Soft-Wired Relationships

The course catalogue also contained several diffuse (soft-wired) bindings between courses, such as prerequisite rules that specify Junior or Senior standing before enrolment in a course. Although such bindings could be modeled in a CPN, I chose for this exercise to focus exclusively on bindings that were explicitly stated for specific courses (hard-wired).

Analysis of Network Topology

The CPN was initially constructed as a directed graph. Python/NetworkX [42] was used to assess whether it also was a directed acyclic graph (DAG). Cycles were removed when detected, resulting in a CPN with a DAG architecture. The DAG-CPN (hereafter, CPN) was then visualized using Pajek [43], Gephi [44], and yEd [45], and analyzed using standard graph metrics with NetworkX and Gephi (see also Supporting Information for further details).

Node Degree

The degree of a node (k) in a directed graph is the sum of the in-coming arcs (k_{in}, or in-degree) and the out-going arcs (k_{out}, or out-degree). Weighted degree [46] was evaluated as:

$$k_i = \sum_{j=1}^{N} a_{ij} w_{ij} \tag{1}$$

where k_i is the weighted degree of node i, and $a_{ij} = 1$ if there exists a connection between nodes i and j, and $a_{ij} = 0$ otherwise. The strength of the connection is adjusted by the weight w_{ij}, which is 1.0 for all links when evaluating an unweighted node degree. A node with a very high degree is termed a "hub" [9], and may be important in the channelling of information. A course with a high in-degree has numerous prerequisites and is inherently integrative. A course with a high out-degree is called as a prerequisite by many other courses, which means it is likely a critical information source in the curriculum.

Shortest Path

The shortest path between two nodes in a graph is the fewest steps one can take in traveling from one to the other node, constrained by the rules that one must follow arc directionality and form a contiguous series of links. In a CPN, where arcs represent prerequisite relationships, one link or step typically denotes the passage of at least one semester. The more steps there are separating two courses, the more opportunity there is for a student to forget material learned in the earlier course, unless there are opportunities later in the curriculum for reinforcement [47, 48]. Thus, as one moves across the CPN through time, one can think in terms of an extinction time for concepts. Course content and prerequisite bindings could be engineered to mitigate this decay.

Betweenness Centrality

Betweenness centrality (b_i) is another popular index, measuring the extent to which a node lies on the shortest paths between all the other nodes in a graph [49]. As such, it speaks to the broader-scale traversability of a network and the role played by individual nodes in that level of connectivity. A course with high betweenness tends to act as a bridge or conduit between large but otherwise isolated regions of a curriculum.

Connected Components and Other Knowledge Communities

A connected component is a set of nodes, and their links, such that all nodes connected to this group are included in the set. Connected components analysis of a CPN assesses the extent to which the curriculum is subdivided into independent, disconnected groupings of courses that do not call one another as prerequisites. A single connected component in a CPN represents an independent knowledge community or domain. By graph theory definition, a single node counts as a single connected component, and so an individual course that is unlinked to any other courses would also count as a single, independent knowledge community.

There are other types of knowledge communities. Some encompass multiple connected components of a CPN while others are embedded within a single, larger connected component, and some require algorithmic detection.

Administrative knowledge communities correspond to the agreed-upon collections of courses associated with individual departments and colleges. Since a department might offer courses with no prerequisite requirements, the administrative knowledge community can contain several independent knowledge communities (connected components).

A latent knowledge community is a set of courses that are more connected to one another than they are to the rest of the curriculum. By definition, a latent knowledge community is embedded within a larger independent knowledge community (connected component), and its borders need not coincide with administrative, departmental boundaries. Community detection in complex networks is a very active area of research since it is widely believed that such communities represent modules that perform separate but integrated tasks in the overall network [50]. The brain has many such modules where, for instance, neurons in the occipital lobe link mostly to other neurons within that lobe, but also integrate with other regions of the brain allowing the broader coordination and assimilation of vision [51–53]. Here, I used the Gephi [44] implementation of the Louvain method [54] to study the BMB program, searching for the optimal partition under the formation of two communities.

Other Metrics

There are many other metrics that have been developed for assessment of network topology through the fields of mathematical graph theory and network science [9, 41]. Many of these have transparent pedagogical interpretations when applied to CPNs, but it is beyond the scope of the present study to review this vast literature and form all of these connections. The metrics I employ here are intended as representing the types of analyses that one could perform.

Biochemistry and Molecular Biology Program

The BMB program at Benedictine University follows the guidelines of the American Society of Biochemistry and Molecular Biology (ASBMB). This intensive and research-oriented program prepares students for interdisciplinary research, education, and vocation in the natural sciences. There are 27 courses required by the BMB major. I formed a new CPN from these nodes and their links (an induced subgraph); all other nodes and links were discarded except for two math courses outside the major (college algebra and trigonometry). Their inclusion here facilitates the analysis of math integration in the program (see also Supporting Information).

Results of CPN Assembly

Method Feasibility

The catalogue information was readily transformed into a CPN using the methods described herein. The syntax of the catalogue was highly regular, and most of the information was readily coded as pairwise bindings between courses. There exist some aspects of the catalogue that were not captured by this CPN such as the soft-wired rules, but these could be addressed if it was deemed important. The approach should be scalable to a large research university without much adjustment. On a fine scale, it also was easy to represent an individual program as a CPN subgraph (BMB program). This task would be amenable to the coding of prerequisite information by hand if a department or program were interested only in the organization of their own unit, and not in the institution-wide picture.

DAG Topology

Directed acyclic graphs (or DAGs) are a type of directed network that lacks cycles, i.e., one cannot leave a node and then later return to it, and these are critical to binary decision diagrams, certain Bayesian and neural networks, and other forms of machine learning [55–57]. Analysis of the CPN using NetworkX indicated that the catalogue, interpreted literally, contained coding for 109 cycles, and so was not a DAG. All of these cycles were due to corequisite bindings between a lecture or theory course (A) and a laboratory or applications course (B). In order to rectify this, the lab prerequisite was removed from each lecture course, while the labs retained the lecture prerequisite, yielding the arc: A → B. Analysis confirmed that the modified CPN had a DAG architecture. All subsequent analyses were performed on this DAG-CPN (hereafter, CPN). It could be argued there was some sacrifice of reality in re-coding the corequisite relations as prerequisites, but if the judgment was made that losing the DAG structure was warranted in order to maintain a literal coding, there are still many interesting things that could be done with a CPN as a simple directed graph.

Curriculum Assessment

The CPN allows assessment of curricular organization (topology) and the information flow that this architecture engenders. Global information entails knowing precisely how any given course connects to another course in the curriculum, and how all courses inter-connect on average over the whole institution. On a finer scale, one also can extract subsets of nodes and their links (subgraphs) to assess programs, departments, or colleges within the institution. Thirdly, it is not only faculty and administrators that need to assess curricula; students must assess a curriculum as a consumer of the product, and a program-level CPN could aid them in this activity. See Supporting Information for additional content and for summary tables of CPN metrics.

Global Topology

Interpreting the Visual Shape of a CPN

The full undergraduate CPN of Benedictine University is shown in Fig. 2. It contains 1,097 nodes (courses) and 770 arcs (prerequisite bindings) representing the 2009 to 2010 catalogue. It is important to recognize that the main source of information in a CPN, and in most networks, is the presence/absence of a link between two nodes. Unlike a principal component analysis, for example, where the point position in multidimensional space is itself critical, there are multiple possible and equally correct projections of any one network onto a two-dimensional or three-dimensional space. Here, I have used the Kamada-Kawai force-directed method [58]. Node positions and the lengths of links do carry visual information, but their values are not rigidly fixed (see Supporting Information).

Course Connectedness (Node Degree)

The connectivity of the CPN was highly heterogeneous. The average node connected to the weighted equivalent of one other node (weighted $k = 1.10$), with half of these connections incoming and half outgoing. However, this "average" node was poorly represented in the actual CPN (Fig. 2). Nearly a third of the nodes ($n = 328$, 29.9%) resided in the largest connected component where the mean degree was (weighted $k = 2.44$) over twice that in the CPN overall (Supporting Information Table 1). Over half of the nodes ($n = 559$, 51.0%) had a degree of zero; these were courses that carried no prerequisite requirement. Many, though not all, of these courses represented elective credit that students might take outside their major, and such courses cannot be burdened with prerequisites if they are to serve this role in the curriculum.

Figure 3 shows variation in out-degree centrality for courses in the largest connected component of the curriculum (see also Supporting Information Table 2). A course with a high out-degree is called as a prerequisite by many other courses, which means it is likely a critical information source in the curriculum. Courses outside the natural sciences were well-represented (60%) in the top 10 list for out-degree centrality. Several of these courses (e.g., PLSC 102 American Government) resided in the smaller connected components where they served as a prerequisite to many courses within their own department. Within the largest connected component, the General Biology course (BIOL 108) occupied an influential position in the curriculum as a primary information source or hub since so many other courses called it as a direct prerequisite ($k = 29.0$).

Reinforcement and Extinction Times (Path Lengths)

The average shortest path connecting nodes in the largest connected component involved 2.48 steps, which fully binds three to four courses in a sequence. For a semester system, this translates to a typical tier of courses linked together in a 2-year block. This also represents the time scale over which reinforcement should be managed if courses occurring later in the curriculum are to cohere with the earlier prerequisites.

Not surprisingly, the longest of all the shortest paths within the largest component consisted of only six steps (diameter of six, linking seven nodes), which translates to seven sequential semesters or 3.5 academic years. This makes sense given the standard graduation target is 4

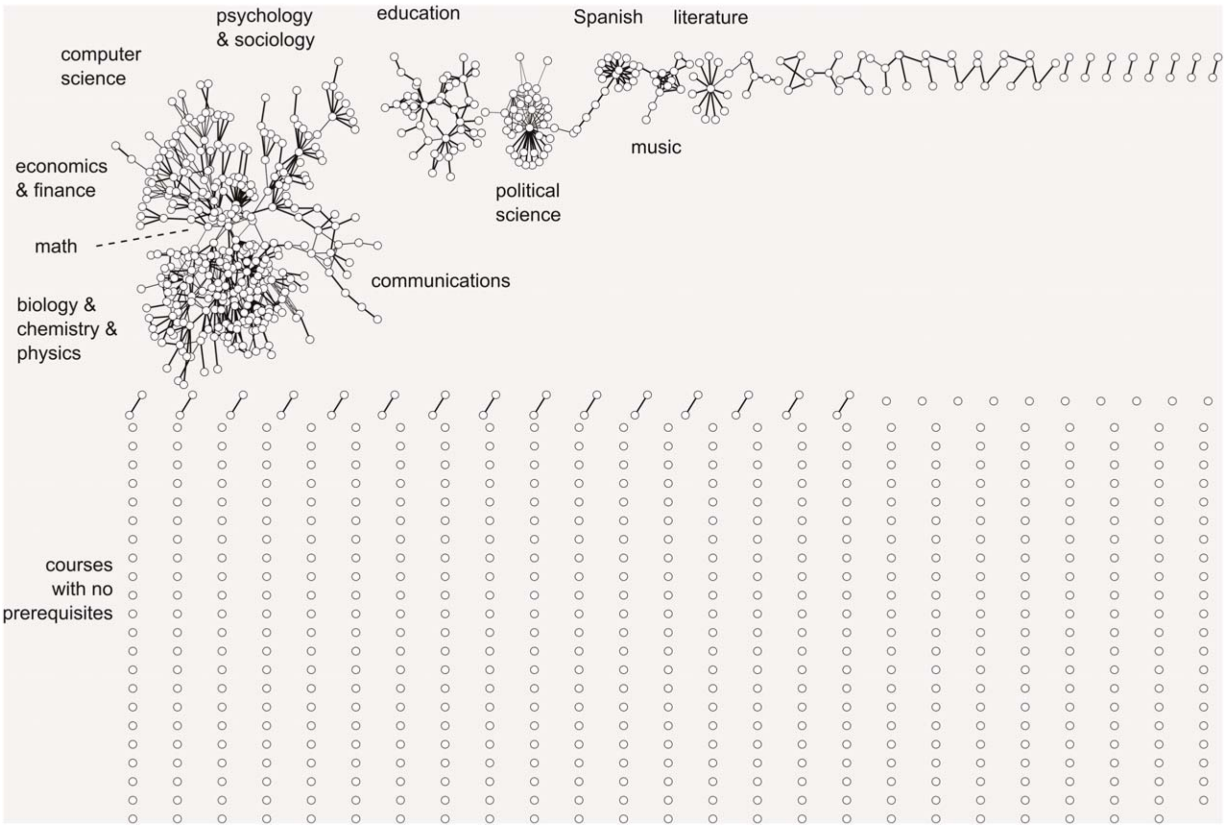

FIG 2 *Entire 2009-10 undergraduate curriculum of Benedictine University as a Curriculum Prerequisite Network (CPN; rendered using Pajek [41]).*

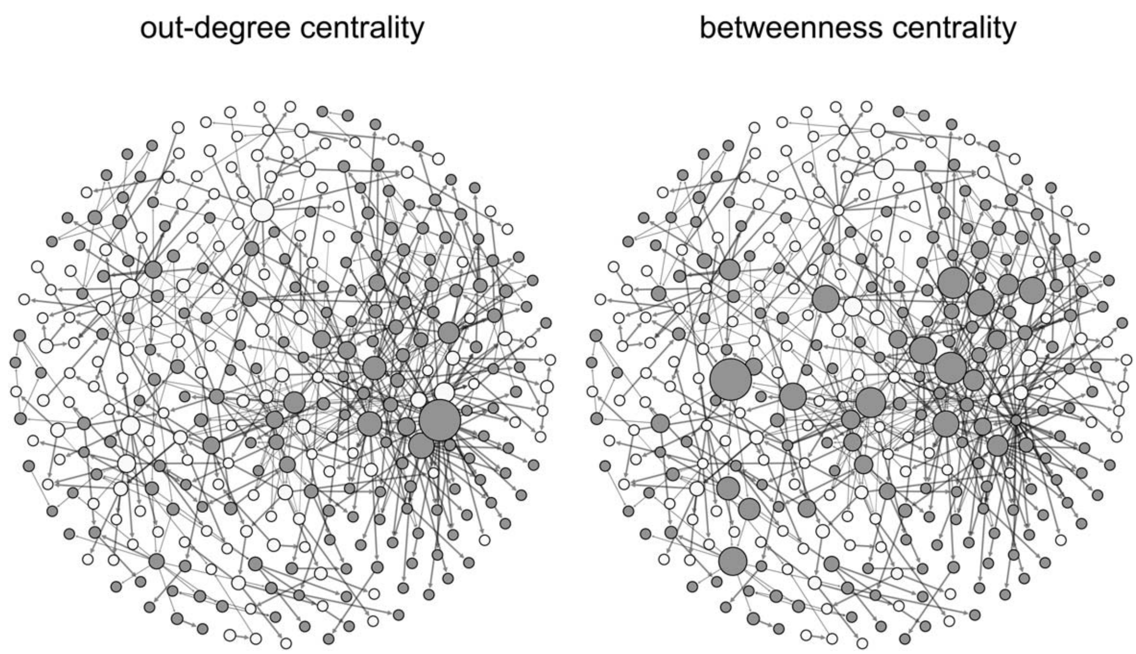

FIG 3 *Largest connected component of the Benedictine University undergraduate CPN (dark nodes, College of Science courses; node size represents out-degree centrality (left) and betweenness centrality (right; using Gephi).*

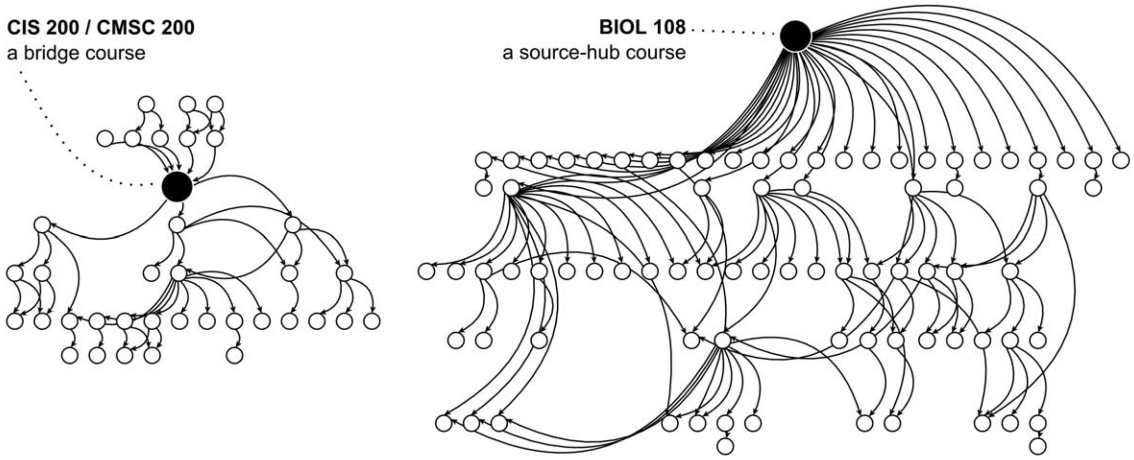

FIG 4 A bridge and a source-hub (rendered using yEd).

years for a baccalaureate degree. Traversability metrics are summarized in Supporting Information Table 1.

Information Bridging (Betweenness Centralities)

Several courses carried a disproportionate load in bridging information through the largest connected component of the curriculum (Fig. 3). All courses in the top 10 betweenness centrality list (Supporting Information Table 3) were from science, technology, engineering, and math (STEM) disciplines; 60% were either mathematics or chemistry.

The composite course on Computer Programming (CIS 200/CMSC 200) had the highest betweenness value. It bridged several early courses with several later ones, as opposed to the BIOL 108 General Biology course which served mainly as an information source (Fig. 4). Computer programming is not an entry-level course, but is sufficiently critical at a point so that many computer courses eventually call it as a prerequisite. As such, the course serves as an information conduit or bridge. That it is cross-listed (CIS/CMSC 200) contributes to its effectiveness at binding different segments of the curriculum. In total, 92 of the nodes (8.4%) were composite courses resulting from cross-listings.

General Chemistry II (CHEM 123) had the next highest betweenness centrality, and was the only course appearing in both top 10 lists (Supporting Information Tables 2 and 3). There was a significant correlation between weighted node degree and betweenness (Spearman rank order correlation $\rho = 0.66$, $p < 0.001$, Fig. 5). Whereas CIS/CMSC 200 had the highest betweenness but rather low degree and BIOL 108 had the highest degree but low betweenness, CHEM 123 had fairly high degree and betweenness, serving as both an information source and a bridge.

Knowledge Communities (Connected Components)

The full CPN contained 599 connected components, or independent knowledge communities (weakly connected components, see Supporting Information). Nearly a third of the courses ($n = 328$, 29.9%) were bound together in the largest

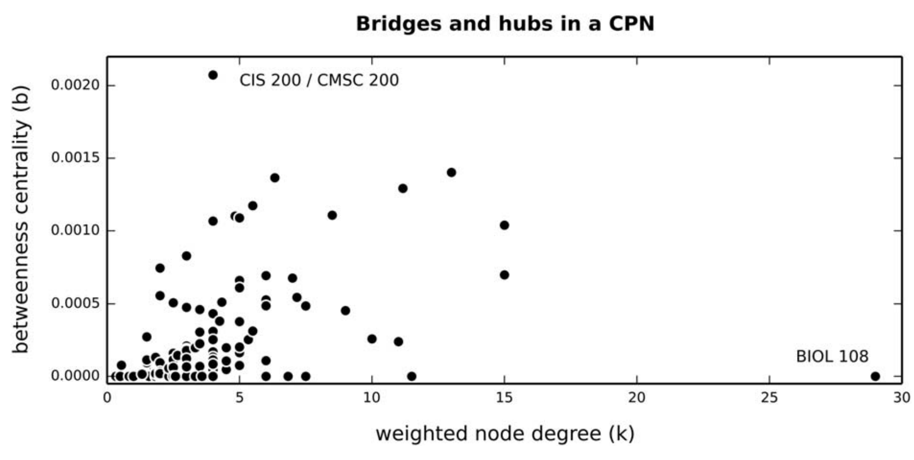

FIG 5 Weighted node degree versus betweenness centrality.

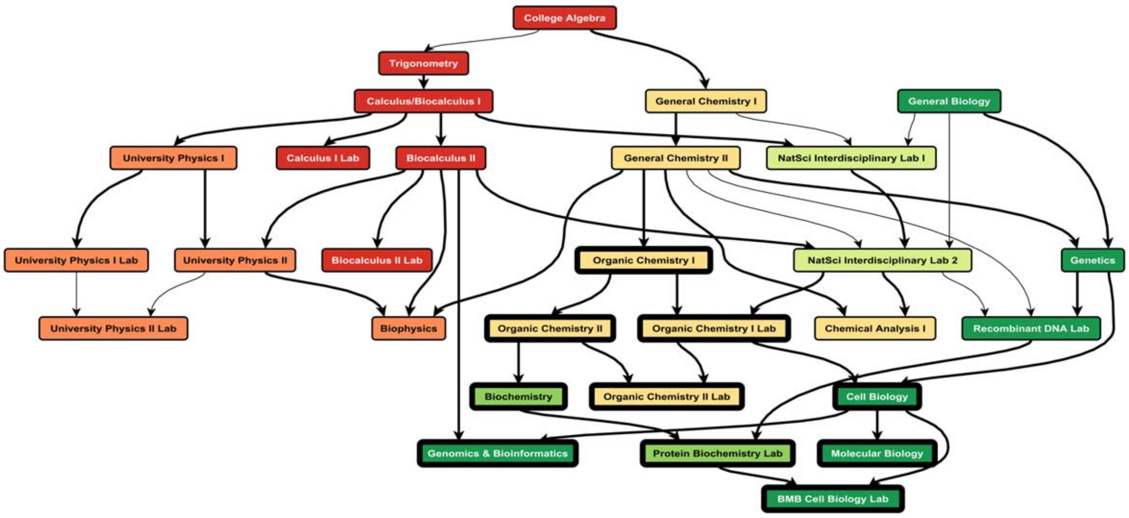

FIG 6 *Induced subgraph of the Benedictine University CPN showing the Biochemistry and Molecular Biology (BMB) major. Red, mathematics; orange, physics; yellow, chemistry; light green, natural science; medium green, biochemistry; dark green, biology. Arcs denote preprequisite relationships. Arc weights indicate whether the prerequisite must be taken (heavy line) or if other courses might stand in place of that course as an alternative prerequisite (light line; node outlines (heavy vs. light) represent the two communities that were identified using the Louvain method in Gephi; (rendered using yEd).*

connected component which contained over two thirds of the total prerequisite bindings, or arcs ($n = 530$, 68.8%). This largest knowledge community consisted predominantly of STEM courses offered by the College of Science (190 COS, 57.9%; 138 other, 42.1%). As Benedictine University has a tradition of excellence in the sciences, it is not surprising that science courses would occupy a central position in the curriculum; natural science students represent 36% of all declared undergraduate majors.

There were a few independent knowledge communities of much smaller size though still comprised of more than one course. Each group represented the offerings from a single department, such as psychology or political science. Several of the components were radially symmetric (Fig. 2) with several courses calling a single, key course as a prerequisite, and these had some of the highest node degrees in the study.

Over half of the courses ($n = 559$, 51.0%) were each an independent knowledge community, unbound to other courses by prerequisites. I address these further in the curriculum reform section.

Other Analyses

There are numerous other ways in which one might assess global topology of a CPN, but space limitations do not permit a full review here. The goals and aspirations of an institution or program should drive the questions that are posed, and the analyses used to address these questions would follow. Clearly, administrators are likely to find global analyses useful since it is their job to maintain a broad oversight of the curriculum. However, this is not to say that individual faculty and departments should be uninterested in global analyses since it is through this perspective that one determines how an individual course or program integrates with the rest of the institution.

Program-level Topology

The induced subgraph of the Biochemistry and Molecular Biology (BMB) program (inclusive of two un-called math courses) contained 29 courses and 46 arcs (Fig. 6). A fine-scale assessment that parallels the global assessment follows:

Course Connectedness (Node Degree)

Two courses shared the highest weighted all-degree of 6.0: CHEM 123 General Chemistry II and MATH 221 Biocalculus II. Both of these courses were information sources in the program, each carrying a single prerequisite and each called as a prerequisite by five other course equivalents (weighted values allowing for alternative prerequisites). Not far behind was the NTSC 152 Natural Science Interdisciplinary Lab II with an all-degree of 5.5; this lab required 3.0 course equivalents in prerequisites, and 2.5 course equivalents called it as a prerequisite (see later section on information bridging).

Reinforcement and Extinction Times (Path Lengths)

Long paths of several courses called in succession as prerequisites represent ample opportunities for information loss. CPNs can be useful for managing retention by rendering these pathways explicit and obvious, and providing a framework on which to map reinforcement events. One might assess path lengths within a program, within a discipline within the program, between disciplines, or one could target specific course pairs.

In the BMB program (Supporting Information Table 4), the longest path in the program was six steps (3 years).

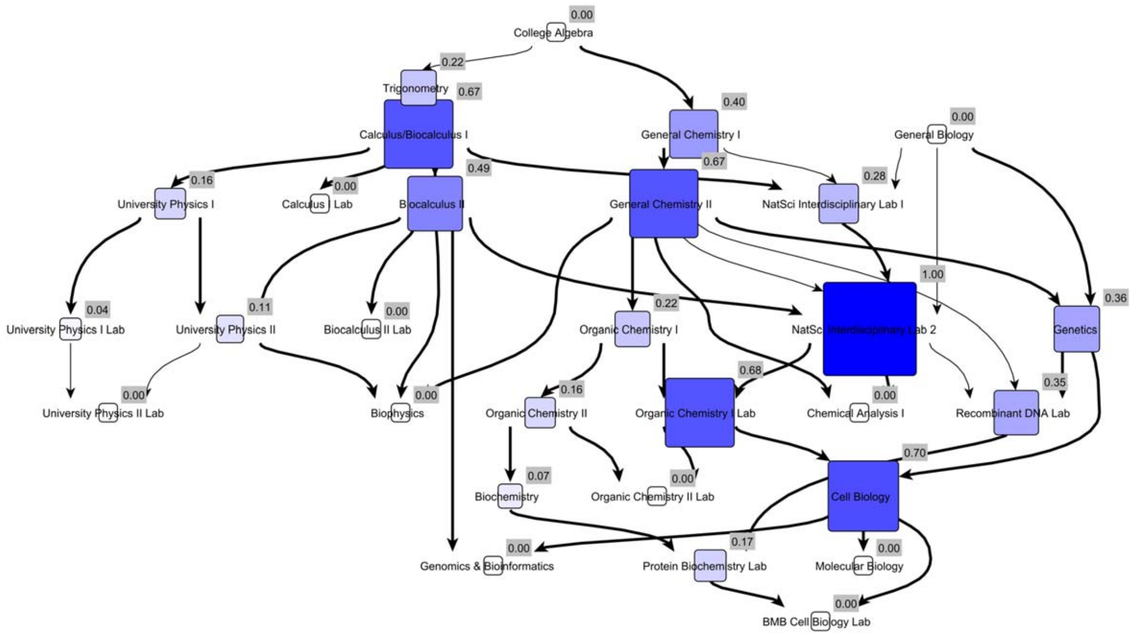

FIG 7 *Betweenness centrality of courses in the Biochemistry and Molecular Biology (BMB) major at Benedictine University. Numbers, node sizes, and node color intensities indicate the betweenness centrality (ranging between 0 and 1.0) for each course. Layout is the same as Figure 6 (rendered using yEd). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]*

This represents a path beginning with college algebra and ending in the capstone experience, BMB Cell Biology Lab.

The maximum number of steps between courses within a discipline, or disciplinary depth, offers another vantage point on extinction times. In the BMB major, math and chemistry both require a maximum of four steps (2 years), biology three steps, physics two, and biochemistry one step (Supporting Information Table 4). In addition to managing retention, disciplinary depth shows how far a program explores a particular area of study.

Alternatively, one might be concerned with cross-department integration. For the BMB major, I calculated the mean shortest path distances from the nearest math course for the different disciplines within the program (Supporting Information Table 5). Biochemistry was farthest from math (4.0 steps), biology (2.8 steps) and chemistry (2.4 steps) intermediate, while natural sciences (1.0 step) and physics (1.4 steps) were most integrated with math courses.

Information Bridging (Betweenness Centralities)

The betweenness centralities of courses in the BMB program are shown in Fig. 7. The course with the highest value is the NTSC 152 Natural Science Interdisciplinary Lab II which integrates General Chemistry and General Biology. Other key bridge courses are Cell Biology, Organic Chemistry II Lab, General Chemistry II, and Biocalculus I. These courses and labs channel a substantial amount of information between different segments of the BMB curriculum.

Knowledge Communities (Latent Communities)

All of the BMB courses resided in a single connected component, suggesting overall programmatic integration. It can be interesting to ask, though, whether some groups of courses are more integrated with one another than they are with other parts of the curriculum. Such latent knowledge communities are "naturally occurring" modules of courses that, by virtue of the pattern of prerequisite linkages, are sharing information more heavily compared to exchanges with other courses. The Louvain method, when set to detect two communities, identified a core of 10 courses (heavy node outlines, Fig. 6) consisting of organic and biochemistry, molecular and cell biology, and genomics and bioinformatics, along with associated labs. All other courses were partitioned into a larger community of nineteen courses.

Advising and Student Assessment of the Curriculum

Students must adhere to the many rules established in the course catalogue, and advisors help them navigate the curriculum. A program-level CPN that reveals the broad-scale picture of the educational program would aid both parties.

Curriculum Reform

Curriculum reform generally is done without a clear view of the shape of the existing curriculum. The CPN provides one solution to this problem by rendering important aspects of the overall curriculum visible and subject to quantitative, systems-level analysis. These analyses should be driven by the goals and aspirations of the institution and its faculty, programs, and administrators. Each of the metrics outlined in the assessment section above could be used as a lens through which to identify problems and guide reform efforts. But rather than re-stating these points here, I distinguish between curriculum integration and curriculum coherence as general goals to aspire to in the use of CPNs for reform.

Curricular Integration

A curriculum is integrated if its courses are linked "properly." There is no one correct topology, but there are clearly many incorrect topologies. For example, departments that form independent information silos represent an incorrect topology. This silo model may have been adequate for a past age, but the 21st century requires more integration according to the AAAS [5] and many other sources.

As an example, the BMB program at Benedictine University exhibits cross-department integration in its CPN. Still, there are departmental partitions evident in the topology. When courses were differentially labeled with discipline-specific colors, it became evident there was strong integration across topics within a discipline, but less integration between disciplines. In particular, math and physics appear on the left, chemistry, biochemistry, and biology on the right of the projection shown in Fig. 6. Some of this is to be expected and even desired; whether this poses a problem is for the program to decide. Should biology be topologically closer to math, like physics? Biocalculus is already an important information source for the program, but it is directly called as a prerequisite by only one biology department course (Genomics & Bioinformatics), the other four courses residing outside the department (physics, math, natural science). Biocalculus II is an implicit prerequisite for any courses calling the second natural sciences lab as a prerequisite, and perhaps that is sufficient. Then it is more a matter of making sure that the later courses cohere in their content with the earlier courses.

A strategy for controlling curricular integration is through the management of bridge courses, which can carry a considerable load of the information traffic between disciplinary areas. One might experiment with increasing the betweenness centrality of an existing bridge course by selectively adding a prerequisite call from an otherwise isolated part of the curriculum. While this might seem arbitrary, and likely would not be practical in many cases, it could prompt the development of new and innovative ways of delivering knowledge and optimizing the learning experience. An important source of bridging can be the cross-listed course, which serves to unite separate departments through the use of a shared course that is fully integrated into each of the departmental curricula.

Managing knowledge communities through rewiring a CPN is another method of curricular integration. The presence of latent knowledge communities shows how curricular wiring can produce modules of courses whose topology is self-reinforcing. One might capitalize on this structure to improve content delivery, or chose a different architecture and re-wire the curriculum to include other courses within a programmatic core for example, such as the one identified in the BMB program (Fig. 6).

Curricular Coherence

A curriculum is coherent if it makes sense. A curriculum might be integrative yet ineffective at delivering a learning experience. Here, one might consider the coherence of information as it flows forward through a curriculum, or the coherence of information as it flows laterally across concurrent courses within a term.

Forward-Flow Coherence

Forward-flow coherence occurs when courses actually use the information established by prerequisite courses. It is possible that a course could serve as a prerequisite in name only if it does not sufficiently use and build upon prior knowledge. Opportunities for reinforcement legitimized by prerequisite bindings should be used to strengthen retention [47, 48], which does not necessarily happen if prior topics are treated as already covered, rather than incorporating new topics into this already established knowledge framework [59].

Although course titles or descriptions in a catalogue do not offer enough detail to track fine-grained information flows, one could develop a survey instrument and query the faculty teaching relevant courses, and then overlay this fine-grained information onto a CPN. A good place to begin would be bridge courses. Actively loading bridge courses with content from early and late in the curriculum is one way of furthering both knowledge integration and coherence; not to do so is a missed opportunity.

Lateral-Flow Coherence

Lateral-flow coherence occurs when courses in the same semester reinforce one another in meaning. In a "learning community" [7, 8], often a group or cohort is engaged in temporally overlapping learning experiences. These might be seemingly unrelated courses from different departments whose content is made to cohere in a fashion that students co-enrolled in both courses are prompted to see connections where they might not have had the courses been treated as independent silos. Alternatively, interdepartmental coherence might be sought within a classroom through team-teaching.

There are many opportunities to foster coherence in a curriculum without rewiring the prerequisite relationships. Over half of the nodes in the Benedictine curriculum ($n = 559$, 51.0%) had a degree of zero; these were courses that carried no prerequisite requirement. Many of these courses represented elective credit that students might take outside their major. With this many independent knowledge communities, there are numerous degrees of freedom in the system. Given the importance of academic freedom in the delivery of classroom content, this is not automatically a negative quality but perhaps even an asset. And, based on my own experiences at the institution, the faculty do work to cohere some of these courses with others offered on campus. But it does suggest that there are many opportunities to introduce coherence in this collection of unbound courses without introducing prerequisite requirements, should that be seen as desirable.

Summary

There is a need for curricular reform that includes greater integration and coherence. Curriculum prerequisite networks provide a means of assessing curricula, and they provide a frame for managing curricular revisions. Networks offer a powerful paradigm with which to model complex systems and visualize broad-scale structure, and there is an abundance of relevant theory and metrics from the field of mathematical graph theory that informs this practice. With the first full-institution CPN of Benedictine University, it is evident that many of these metrics have a clear pedagogical application and interpretation. The overall approach will be scalable to other and larger institutions. Through adjusting the links between courses, one can optimize the performance of a curriculum, or at the very least make faculty and programs aware of the roles their parts are playing in the greater whole. Efforts at revision should be mindful of how prerequisite bindings influence overall integration. Moreover, coherence should be a goal so that curricula make sense to the students that navigate them. Such efforts to transform the way we interact with curricula should serve to empower faculty and administrators with much-needed information on the practices and infrastructure of their own organizations. And on the most basic level, as faculty we should seek to understand the academic framework within which we deliver content, since this framework represents the institutional presuppositions, sensu Wittgenstein [60], that must integrate and cohere with our own courses.

Acknowledgments

The author thanks Benedictine University for the use of its catalogue data, and Ed Ferroni, BMB program director, for discussions about the program, and two anonymous reviewers for comments.

References

[1] National Research Council. (1999) Transforming Undergraduate Education in Science, Mathematics, Engineering, and Technology, National Academies Press, Washington, DC.

[2] National Research Council. (2003) Improving Undergraduate Instruction in Science, Technology, Engineering, and Mathematics: Report of a Workshop, National Academies Press, Washington, DC.

[3] National Research Council. (2009) A New Biology for the 21st Century: Ensuring the United States Leads the Coming Biology Revolution, National Academies Press, Washington, DC.

[4] Steen, L. A. (2005) Math and Bio 2010: Linking Undergraduate Disciplines, Mathematical Association of America, Washington, DC.

[5] AAAS. (2010) Vision and Change: A Call to Action, American Association for the Advancement of Science, Washington, DC.

[6] Ledder, G., Carpenter, J. P., and Comar, T. D. Eds. (2013) Undergraduate Mathematics for the Life Sciences: Models, Processes, and Directions, Mathematical Association of America, Washington, DC.

[7] Lave, J. (1988) Cognition in Practice: Mind, Mathematics and Culture in Everyday Life, Cambridge University Press, Cambridge.

[8] Lave, J. and Wenger, E. (1991) Situated Learning: Legitimate Peripheral Participation, Cambridge University Press, Cambridge.

[9] Newman, M. E. J. (2010) Networks: An Introduction, Oxford University Press, Oxford.

[10] Barabási, A. -L. (2012) The network takeover. Nat. Phys. 8, 14–16.

[11] Linnaeus, C. (1758) Systema Naturae, Laurentius Salvius, Stockholm.

[12] Darwin, C. (1859) On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life, John Murray, London.

[13] Elton, C. S. (1927) Animal Ecology, Sidgwick and Jackson, London.

[14] Krebs, H. A. (1937) The intermediate metabolism of carbohydrates. Lancet 230, 736–738.

[15] Grzybowski, B., Bishop, K., Kowalczyk, B., and Wilmer, C. The 'wired' universe of organic chemistry. Nat. Chem. 1, 31–36.

[16] Euler, L. (1736) Solutio problematis ad geometriarn situs pertinentis Comment. Acad. Sci. Imp. Petropolitanae 8, 128–140.

[17] Erdős, P. and Rényi, A. (1959) On random graphs. Publ. Math. 6, 290–297.

[18] Watts, D. J. and Strogatz, S. H. (1998) Collective dynamics of 'small-world' networks. Nature 393, 440–442.

[19] Barabási, A. -L. and Albert, R. (1999) Emergence of scaling in random networks. Science 286, 509–512.

[20] Compeau, P. E. C., Pevzner, P. A., and Tesler, G. (2011) How to apply de Bruijn graphs to genome assembly. Nat. Biotechnol. 29, 987–991.

[21] Davidson, E. and Levin, M. (2005) Gene regulatory networks. Proc. Natl. Acad. Sci. USA 102, 4935.

[22] Bonetta, L. (2010) Protein-protein interactions: Interactome under construction. Nature 468, 851–854.

[23] Palsson, B. Ø. (2006) Systems Biology: Properties of Reconstructed Networks, Cambridge University Press, Cambridge.

[24] Yildirim, M. A., Goh, K. -I., Cusick, M. E., Barabási, A. -L., and Vidal, M. Drug-target network. Nat. Biotechnol. 25, 1119–1126.

[25] Goh, K. -I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., and Barabási, A. -L. (2007) The human disease network. Proc. Natl. Acad. Sci. USA 104, 8685–8690.

[26] Jacobs, H. H. (1997) Mapping the Big Picture: Integrating Curriculum and Assessment K-12, Association for Supervision and Curriculum Development, Alexandria, VA.

[27] Jacobs, H. H. (2005) Getting Results with Curriculum Mapping, Association for Supervision and Curriculum Development, Alexandria, VA.

[28] Jacobs, H. H. (2009) The Curriculum Mapping Planner: Templates, Tools, and Resources for Effective Professional Development, Association for Supervision and Curriculum Development, Alexandria, VA.

[29] Udelhofen, S. (2005) Keys to Curriculum Mapping: Strategies and Tools to Make It Work, Sage Publications, Inc., Thousand Oaks, CA.

[30] Plaza, C. M., Draugalis, J. R., Slack, M. K., Skrepnek, G. H., and Sauer, K. A. (2007) Curriculum mapping in program assessment and evaluation. Am. J. Pharm. Educ. 71, 1–8.

[31] Kelley, K. A., McAuley, J. W., Wallace, L. J., and Frank, S. G. (2008) Curricular mapping: Process and product. Am. J. Pharm. Educ. 72, 1–7.

[32] Uchiyama, K. P. and Radin, J. L. (2009) Curriculum mapping in higher education: A vehicle for collaboration. Innov. High. Educ. 33, 271–280.

[33] White, H. D. and Calhoun, K. (1984) Mapping a curriculum by computer J. Am. Soc. Inf. Sci. 35, 82–89.

[34] Starr, M. L. and Krajcik, J. S. (1990) Concept maps as a heuristic for science curriculum development: Toward improvement in process and product. J. Res. Sci. Teach. 27, 987–1000.

[35] Sommaruga, L. and Catenazzi, N. (2007) in Proceedings of the Twelfth International Conference on 3D Web Technology, Association for Computing Machinery, New York, NY, pp. 177–180.

[36] Gestwicki, P. (2008) In Frontiers in Education Conference, 2008. FIE 2008, 38th Annual, Institute of Electrical and Electronics Engineers (IEEE), pp. T3E–13–T3E–14.

[37] Kabicher, S. and Motschnig-Pitrik, R. (2009) In Ninth IEEE International Conference on Advanced Learning Technologies, 2009 (ICALT 2009), Institute of Electrical and Electronics Engineers (IEEE), pp. 742–743.

[38] Mitchell, G. G. and Howlin, C. P. (2009) Presented at the 2009 China-Ireland International Conference on Information and Communications Technologies (CIICT), Maynooth, Ireland. pp. 256–259.

[39] Zucker, R. (2009) ViCurriAS: A curriculum visualization tool for faculty, advisors, and students. J. Comput. Sci. Coll. 25, 138–145.

[40] Siirtola, H., Raiha, K. -J., and Surakka, V. (2013) Interactive Curriculum Visualization, Information Visualisation (IV), 2013 17th International Conference, Institute of Electrical and Electronics Engineers (IEEE), London, pp. 108–117.

[41] Gross, J. L. and Yellen, J. (2004) Handbook of Graph Theory: Discrete Mathematics and Its Applications, CRC Press, London, 2004.

[42] Hagberg, A., Swart, P., and Chult, D. S. (2008) Exploring Network Structure, Dynamics, and Function Using NetworkX, Vol. LA-UR-08–05495, Los Alamos National Laboratory Los Alamos, NM.

[43] Batagelj, V. and Mrvar, A. (1998) Pajek - Program for large network analysis. Connections 21, 47–57.

[44] Bastian, M., Heymann, S., and Jacomy, M. (2009) In Third International AAAI Conference on Weblogs and Social Media, Association for the Advancement of Artificial Intelligence (AAAI).

[45] yWorks. (2014) yEd Graph Editor [Online]. Available: http://www.yworks.com/en/products/yfiles/yed/, (Last accessed April 14, 2014).

[46] Barrat, A., Barthélemy, M., Pastor-Satorras, R., and Vespignani, A. (2004) The architecture of complex weighted networks. Proc. Natl. Acad. Sci. USA 101, 3747–3752.

[47] Karpicke, J. D. and Blunt, J. R. (2011) Retrieval practice produces more learning than elaborative studying with concept mapping. Science 331, 772–775.

[48] Inda, M. C., Muravieva, E. V., and Alberini, C. M. (2011) Memory retrieval and the passage of time: From reconsolidation and strengthening to extinction. J. Neurosci. Off. J. Soc. Neurosci. 31, 1635–1643.

[49] Brandes, U. (2001) A faster algorithm for betweenness centrality. J. Math. Sociol. 25, 163–177.

[50] Fortunato, S. (2010) Community detection in graphs. Phys. Rep. 486, 75–174.

[51] Chen, Z. J., He, Y., Rosa-Neto, P., Germann, J., and Evans, A. C. (1991) Revealing modular architecture of human brain structural networks by using cortical thickness from MRI. Cereb. Cortex NYN 18, 2374–2381.

[52] Bullmore, E. and Sporns, O. (2012) The economy of brain network organization. Nat. Rev. Neurosci. 13, 336–349.

[53] Sporns, O. (2010) Networks of the brain. The MIT Press, Cambridge, MA.

[54] Blondel, V., Guillaume, J., Lambiotte, R., and Lefebvre, E. Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. 2008, P10008.

[55] Akers, S. (1978) Binary decision diagrams. Comput. IEEE Trans. On. 27, 509–516.

[56] Pearl, J. (1988) Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, Burlington, MA.

[57] Haykin, S. O. (2008) Neural networks and learning machines. Prentice Hall/Pearson, New York.

[58] Kamada, T. and Kawai, S. (1989) An algorithm for drawing general undirected graphs. Inf. Process. Lett. 31, 7–15.

[59] Committee on Developments in the Science of Learning. (2000) How People Learn: Brain, Mind, Experience, and School, 2nd ed., National Academies Press, Washington, DC.

[60] Wittgenstein, L., in G. E. M. Anscombe and G. H. von Wright, Eds. (1969) On Certainty. D. Paul and G. E. M. Anscombe (transl.). J. & J. Harper Editions, New York, NY.