8-27-2014

Graph Theoretic Methods for the Analysis of Data in Developing Systems

Kris H. Green
St. John Fisher College, kgreen@sjfc.edu

Bernard P. Ricca
St. John Fisher College, bricca@sjfc.edu

Publication Information

Green, Kris H. and Ricca, Bernard P. (2014). "Graph Theoretic Methods for the Analysis of Data in Developing Systems." *Quality and Quantity* 49.5, 2037-2060.

Abstract

A full examination of learning or developing systems requires data analysis approaches beyond the commonplace pre-/post-testing. Drawing on graph theory, three particular approaches to the analysis of data—based on adjacency matrices, affiliation networks, and edit distances—can provide additional insight into data; these methods are applied to student performance in a Calculus course. Data analysis methods based on adjacency matrices demonstrate that learning is not unidimensional, that learning progressions do not always progress monotonically toward desired understandings and also provide insight into the connection between instruction and student learning. The use of affiliation networks supports the concept development theory of Lev Vygotsky and also provides insight into how students' prior knowledge relates to topics being studied. Careful use of edit distances indicates a likely overestimate of effect sizes in many studies, and also provides evidence that concepts are often created in an ad hoc manner. All of these have implications for curriculum and instruction, and indicate some directions for further inquiry.

Keywords

fsc2015, Graph theory, concept development, edit distance, affiliation networks, learning progressions

Disciplines

Mathematics

Comments

This is a draft that was submitted for publication. The final publication is available at Springer via http://dx.doi.org/10.1007/s11135-014-0089-5. Published online on August 27, 2014.

Kris Green[1] and Bernard Ricca[1]

Graph Theoretic Methods for the Analysis of Data in Developing Systems

[1] Department of Mathematical and Computing Sciences, St. John Fisher College, 3690 East Avenue, Rochester, NY 14618, USA

Corresponding Author: Bernard Ricca, bricca@sjfc.edu, 585-899-3866 (office), 585-899-3872 (fax)

This is a draft that was submitted for publication. The final publication is available at Springer via http://dx.doi.org/10.1007/s11135-014-0089-5

Abstract

A full examination of learning or developing systems requires data analysis approaches beyond the commonplace pre-/post-testing. Drawing on graph theory, three particular approaches to the analysis of data – based on adjacency matrices, affiliation networks, and edit distances – can provide additional insight into data; these methods are applied to student performance in a Calculus course. Data analysis methods based on adjacency matrices demonstrate that learning is not unidimensional, and that learning progressions do not always progress monotonically toward desired understandings and also provide insight into the connection between instruction and student learning. The use of affiliation networks supports the concept development theory of Lev Vygotsky and also provides insight into how students' prior knowledge relates to topics being studied. Careful use of edit distances indicates a likely overestimate of effect sizes in many studies, and also provides evidence that concepts are often created in an *ad hoc* manner. All of these have implications for curriculum and instruction, and indicate some directions for further inquiry.

Keywords:

graph theory – concept development – edit distance – affiliation networks – learning progressions

1. Introduction

Developing a deeper understanding of student learning is one goal of educational research. St. Julien (1997) claimed that a general understanding of learning was “close to the holy grail” of education. Despite the effort that has gone into understanding learning, relatively little progress seems to have been made in understanding how students learn, how they develop new concepts and skills. While there are many reasons for this relative lack of progress, this paper will focus on three: the nature of development, the difficulties in collecting appropriate data, and the problems of data analysis.

The literature on development is diverse and voluminous, and choosing from among the possible approaches is difficult. Here, two important elements of the developmental literature will be chosen to guide the collection and analysis of data. The first element is to consider a *dynamic systems* approach (Thelen and Smith, 1994). Such an approach looks at observed behavior as consisting of the contextual coordination of underlying abilities. The second element is the idea of concept development as consisting of a progression from *heaps* to *complexes* to *concepts* (Vygotsky, 1986). While the details of this progression are not important here, the notion of concept development as a progression of different ways that various components can be coordinated is important.

Studying development, then, requires the collection of data that operates on at least two levels. One level is the overarching, and usually directly observable, behavior. In a school context, these behaviors are typically things that are measured on end-of-course tests and the like. Even when these tests are coupled with tests of prior knowledge to form a pre-/post-test set of data, these tests fail to capture the dynamics of the learning situation, showing only what can or cannot be done by students (Koopmans, 2014a). Further, a developmental approach requires that a second, and often less directly observable, level of data is required. In a classroom, these data consist of the elements and perceived contexts that drive student performance.

Clearly, a thorough understanding of learning will require a longitudinal approach to capture the development that takes place. There are several common approaches to the longitudinal study of systems. Singer and Willett (2003) presented three features of research design that are common to longitudinal analysis of data. These are:

- Three or more waves of data
- An outcome whose values change systematically over time
- A sensible metric for clocking time

(p. 9)

While designs that include the features highlighted by Singer and Willett, such as individual growth modeling, multilevel modeling, hierarchical linear modeling, etc., have been successful in a number of situations, because of the learning processes being investigated here, there are several reasons that these usual longitudinal approaches, which include pre-/post-test designs, do not work well in the learning situation considered here.

The data considered here are the performances of students during the process of learning. While the context – a first-year calculus course – is specific to this study, there are many situations in which data taken from students as they learn will be similar in structure to the data that are analyzed here. In particular, the salient features of this data structure are:

1. The task requires students to initially compare objects to one another, and to use the results of that comparison to choose a procedure for completing the task. (This type of task is rather general in school situations, as it may apply to choosing among techniques

of integration, choosing approaches to the solution of an algebra problem, organizing a five-paragraph essay, and others.)

2. The students come to the tasks with a great deal of prior knowledge. Much of this knowledge, at least on the surface, appears to be of use in the tasks. However, most of the prior knowledge consists of concepts and ideas that do *not* apply to the tasks and some of that prior knowledge actually blocks the successful completion of the task.
3. Students must both learn new concepts and techniques while simultaneously coordinating this new "knowledge" with their prior knowledge to address the task.
4. Although it appears that there are heuristics (Schonfeld, 1978) that can be used to address the tasks, such heuristics are in general not practical: Most of the heuristics are far too complicated to expect any student to remember them all, except in very limited contexts. Further the use of heuristics requires the application of only *codified knowledge* in the Vygotskian sense of independent ability. (This knowledge is the region "below" Vygotksy's (1980) *zone of proximal development* and by definition does not include what students are learning.)

In situations where these four features are important, such as the one studied here, there are several reasons that the standard pre-/post-test design, and indeed, most other common longitudinal approaches, are insufficient to understand the processes involved. First, because the learners are qualitatively different at the end of the study (compared to the beginning), the common approaches fail to adequately capture the causes of change (Koopmans, 2014b). Instead, these approaches describe whether a significant change has occurred, but not the reasons for these changes. Furthermore, there can be situations where qualitative changes in understanding have occurred – for example moving from *enactive* to *iconic* to *symbolic* representations or understandings (Bruner, 1966a, 1966b) – but which show no change in performance on particular tasks; in such cases the usual longitudinal design will likely show no growth, when substantial growth has occurred. (In the case of changing representations, Bruner claims that symbolic representation is necessary to use knowledge for further growth. In such cases, the change in representation may not become *evident* until a later time, even though learning had occurred.)

Second, learning often requires that specific examples or problems be categorized appropriately: For example, introductory physics courses often use equations as de facto categories, and encourage students to identify which equations are necessary to solve a problem, thereby requiring students to classify problems into categories/equations. Likewise, in some curricular approaches, students learning to interpret or write paragraphs must become proficient in the categories of *topic sentence*, *supporting sentence*, and *transition sentence*. (Note however, that in the first example there are often multiple ways to solve a problem, so a given problem may fall into multiple categories.) However, as Vygotsky (1986, and see below) showed, the manner in which learners classify objects, even when similar on the surface, may be quite varied in their features and affordances (Gibson, 1977), and so merely identifying the classification scheme of a learner is not sufficient.

Although classification schemes admit a metric (Diebel, Anderson and Anderson, 2005), a third way in which the common longitudinal approach can fail to adequately capture learning is that this metric is not easily interpretable or easily connected to the learning process. Piaget and Garcia (1991) note that the logic of learning is different than the logic of the expert, writing that "[w]ith respect to practical actions, we must distinguish their causal aspect (the outcome that is verifiable after the fact) from their anticipation which is inferential" (p. 4). Commenting on this, Doll (2008) said:

> Piaget is here making a distinction, important at the level of *practical operations*, between a logic of verification (the "truth" of a statement – the basis for the theory of assessment that educational research uses) – and what he calls operatory logic. Operatory logic can rightly be considered as the practical logic a child uses in developing his/her understanding….*this practical logic is one of development, not of verification*….To focus on the actual reasoning the learner uses – his/her intentions – is a departure from the validity *only* frame so dominant in current schooling. The validity of an action, statement, procedure is, of course, important; but to exclude a person's intentions/anticipations/inferences from the process of learning is to turn learning into a simplistic and mimetic act.
>
> In traditional acts of teaching what the child or learner *intends* to do is not considered, only what s/he did. For Piaget, such a non-developmental view misses the child's "constructions," his/her practical actions with all their "illogicalnesses."
>
> (p 28-9, italics in original)

The traditional approaches to studying development share the same problems as the traditional approaches to assessment in education that Doll critiques here: The logic embodied in the common longitudinal research approaches fails to follow the operatory logic of the learner, and instead typically follows only the logic of verification. That is, these approaches measure *only* against some final, stable, "expert" method of solving a problem rather than *also* considering the transitory coordination of components that are often used. This, as we will see below, often obscures what is taking place during development.

Fourth, it should be noted that when classification is involved, incorrect classification of an object is not merely the opposite of (or lack of) correct classification; the incorrect classification places an object in one of what may be many other categories. In most cases, categories are not ordered along a single dimension, and hence, data involving classifications are not unidimensional; in fact, it may be that "dimension" is not an appropriate concept for classification schemes.

Hence, a longitudinal scheme, even if carefully undertaken, may fail to capture important information about learning or development.

The present paper was motivated by a Calculus topic, *techniques of integration*; difficulties in the analyses of the data motivated the development of the data analysis techniques presented here; more details about the context of this paper will be found in Green and Ricca (under review). The particular data examined here are how students classified integration problems at various points during the study, and what underlying reasons students gave for their classifications. However, because the four features noted above apply to many learning situations, the methods developed here should be of interest to a broad audience.

2. Studying the Development of Concepts and Skills

Given the need to develop and use novel approaches to data analysis for this situation, a closer look at development and the shortcomings of commonplace approaches to developmental data analysis is appropriate here.

2.1. Development

Two important approaches to development are used here: consideration of development as a dynamic system, and Vygotsky's ideas about concept development.

2.1.1. A dynamical systems approach to techniques of integration.

Thelen and Smith's (1994) work on coordinating individual muscle motions into various means of locomotion indicated that such coordinations are both goal oriented and created on the fly, and further that significant practice of the individual motions occurs during the attempts at coordination and subsequent locomotion. Hence, the various ways in which an infant kicks her or his legs are the development of the muscle motions necessary for walking. Later, the various muscle contractions in the legs can be coordinated to allow for walking, running, jumping and other forms of locomotion. While the muscle motions remain the same, the coordinations are dependent upon the context (e.g., walking is good for flat ground on the earth, while hopping is better for movement on the moon when one is constrained by a space suit,) and the intent of the person (e.g., whether one is in a hurry or not). Regardless of the type of locomotion, the various leg muscles receive significant practice during locomotion.

Techniques of integration (e.g., *integration by parts*, *substitution*, etc.) and how to recognize which technique is appropriate to a given problem are common topics covered in first-year calculus courses (e.g., Chapter 7 of Stewart, 2007). The usual approach in calculus textbooks is to have students practice the various techniques of integration first and then have students learn to determine which technique should be used on a given problem, usually through the use of some sort of flow chart or other heuristic. The results of this approach to the topic are usually unsatisfying, both to students and to instructors (including both authors of this paper).

Hypothesizing that a process similar to locomotion could take place during the learning of calculus integration techniques – that rather than a fixed structure of techniques to try sequentially, each problem motivated a temporary coordination of skills (e.g., ability to use algebra to reduce a rational function or not, properties of exponents, recognition of derivatives of functions) into a technique (e.g., integration by parts) - led the authors to view the various integration techniques as components that had to be coordinated to make progress towards a solution. Coordinations of leg muscles into a walk, run, or jump based on an immediate context imply that calculus students need to learn to coordinate the various approaches to integration (and other mathematical knowledge) during the course of solving a particular problem. Hence, the original goal of this study was to investigate how an explicit focus on classifying integration problems would change student performance.

A similar coordinations approach perhaps can be used to examine other situations. The learning of multiplication in this view is similar to an approach presented by Kamii, Clark & Dominick, (1997) who recognized that the change from repeated addition to multiplication is one where the additions involved are coordinated sequentially (in repeated addition) or simultaneously (in multiplication). More generally, learning to solve quantitative problems involve the ability to choose among possible approaches/equations, and then to work through those in a particular order, thus coordinating various steps. Likewise, the writing of an essay involves the choice of and coordination of ideas along with the construction of sentences to express and link those ideas.

2.1.2. Concept formation and Vygotksy's method of *double stimulation*.

Vygotsky (1986) experimentally investigated concept formation through the method of *double stimulation*. In this approach, "[t]wo sets of stimuli are presented to the subject, one set as objects of his [sic] activity, the other as signs which can serve to organize that activity" (p. 103). For example, the objects could be blocks of various color, shape, height, and size, with a nonsense word written on the bottom of each block (hidden from the subject). The subject would be asked to classify the objects into appropriate groups. After a classification was performed, the experimenter would show the subject the bottom end of two of the blocks, revealing the word; the word named the group to which the object belonged. A process of reclassification and revelation would be repeated until a completely successful categorization occurred.

As a result of these experiments, Vygotsky (1986) outlined the development of concepts as falling into three basic phases (and multiple sub-phases): *heaps*, *complexes*, and *concepts*. Grouping by heaps results in "inherently unrelated objects linked by chance" (p. 110). Such groupings are highly unstable. When thinking in complexes, groupings are made "not only by…subjective impressions but also by *bonds actually existing between these objects*" (p. 112, italics in original). However, in a complex, these bonds are "*concrete and factual* rather than abstract and logical" (p. 113, italics in original)[1]. Because these bonds are concrete, "[*a*]*ny factually present* connection may lead to the inclusion of a given element into a complex" (p. 113, italics in original). Notice that this may mean that a resulting group may have no single concept describing it: Elements A and B may share, for example, a common color, while elements B and C may, on the other hand, share a common height which A does not share; this type of *associative* grouping is one of the sub-phases of the complexes phase. In the third phase, *concepts*, the groupings are made as a result of the application of an abstracted concept to the various elements.

One sub-phase of complexes, *pseudo-concepts*, can be outwardly identical with groupings according to concepts. Vygotsky claims that pseudo-concepts allow for

---

[1] Piaget distinguished between *operatory logic*, the practical and changing logic of development, and the *logic of verification*, which is the logic of the final edifice of knowledge. It will be important to keep in mind that what seems not-logical to the expert (using the logic of verification) may yield insight into the operatory logic of the learner. See Piaget and Garcia (1991) for more.

communication between adult and child (or, in our context, teacher and student) even before the full development of conceptual understanding. However, it must be remembered that the pseudo-conceptual processes of grouping, and the affordances (Gibson, 1977) of those groupings, are not the same as the processes and affordances of conceptual groupings.

The method of double stimulation does *not* require a fixed structure, and hence is compatible with the thesis that coordinations are created on the fly from existing components; both components and coordinations can change during the study.

Data collection in this study is modeled after the method of double stimulation. Participating students were given a number of integration problems to classify. After each classification, students were given some feedback regarding their work, although the feedback was not as simple as in the classifying objects task described above. (For more details, see Green and Ricca, under review.)

Vygotsky's approach to concept development fits well with Thelen and Smith's (1994) approach. Although Vygotsky is silent on the issue of whether or not heaps-complexes-concepts are formed from previously developed heaps-complexes-concepts, given that the various sub-phases have different organizing principles, it is reasonable to assume that a learner's performance does not develop from a previous performance, but rather from a new coordination of some underlying elements. This is the approach that will be taken here.

2.2. Developmental Data

Developing systems, which are one type of *complex system*, require data analyses that are different than is the case when a system is not complex. Several recent papers have drawn attention to further shortcomings in the usual analyses of such developing/complex systems, particularly in education. Opfer (2013) parallels Thelen and Smith (1994) by noting that teaching and learning are “context and condition dependent” (Opfer, 2013, p. 1) and called for more development of complexity related research methodology. In a similar vein, Koopmans (2014b) questioned the use of randomized controlled trials as a gold standard for education, and examined how a more dynamic systems approach to studies in education may produce a deeper understanding of what works in education. Gilstrap (2013) examines the issue of quantitative analyses directly, focusing on the need for more post hoc analyses when dealing with complex systems.

The specific move to considering the solution of an integration problem to be a temporary coordination of parts also limits the usefulness of commonplace data analysis methods. First, there may be, during any one student’s development of expertise, great variability in their attempts. Thelen and Smith (1994) stated “[a] dynamic approach elevates variability, both within and between individuals, into an essential element in the developmental process. Variability is a metric of stability and a harbiniger of change. Variability is also the essential ground for exploration and selection” (p. 341-2). As we will see, the variability encountered in this investigation is too large to determine statistical significance at a meaningful level by commonplace methods. Second, the data led the authors to recognize that - at least in this case – *being wrong* is not the opposite of *being right*. Hence, there is not a single axis along which to plot student results. This produces a very large number of possible student responses, and the number of paths of learning becomes larger still. These observations also led to a third problem with common methods: not all ways of being wrong, or of being right, are created equal; hence, assigning a “score” to student work is not trivial.

Taken together, these reasons demonstrate the need to analyze data from learning situations by methods other than the standard ones. This paper will explore several possible approaches to the examination of learning, and demonstrate how some approaches originating in graph theory can generate insight into developing systems when more commonplace methods cannot.

3. Approaches to Data Analysis

To fully benefit from data analysis, we need approaches that will enable us to study potential coordinations, to use both operatory logic and verification logic, and to not be restricted to a single dimension. Given the type of data collected, student categorization of problems, several approaches to the analysis of data yield insight, although no single one can be considered the overall "best" approach. Three useful approaches will be examined here:

1. An approach using the adjacencies of the data to examine the progression of student understanding, and the potential connections of those progressions to instruction. This approach will allow for "not right" to be along a different axis than "being wrong".
2. Using the affiliation networks (Wasserman and Faust, 1994) of the data to better understand the underlying approaches used by students. This will allow for a study of the stability of student ideas, which will give insight both into the formation of heaps, complexes, and concepts as well as the operatory logic of the students.
3. An analysis of the data using *edit distances* (Diebel, Anderson and Anderson, 2005) to better understand variability during learning. This will yield insight into how students coordinate components, and whether those coordinations are built on one another or constructed "on the fly".

While none of these approaches is strictly new, the authors have made some modifications to approaches, and examined how insightful each is to investigations such as this one.

3.0.1. Participants

Fifteen students in a second semester Calculus course at a small Northeast college participated in this study. There were 10 females and 5 males in the study. All of the students were had declared a major which required the course (e.g., Chemistry, Mathematics) The course included 1 senior, 3 sophomores, and 11 first year students. Three of the students had previously taken the course and were repeating it to earn a higher grade. Most of the students worked in pairs during the investigation.

3.0.2. Data

The students were each presented with 20 cards in a pseudo-random order; each card had one integration problem printed on it. We will label these 20 problems a_1, a_2, a_3, a_4, b_1, b_2…,e_4, where the letter denotes to which group a problem belongs, and the subscript denotes a problem within the group; the student cards did not have this notation on them. Figure 1(a) shows a listing of the instructor's categorization, along with a brief description of the categories. These problems were chosen by the authors to consist nominally of 4 problems in each of five categories. Category A consisted of "basic" integration problems that students should have been able to solve using knowledge from their first semester Calculus course. Categories B, C, and D each consisted of integration problems requiring a single particular technique of integration. Category E consisted of harder integration problems that required the sequential use of two of the techniques used in categories B, C, and D.

Each of the five rounds of data collection consisted of (a) how students classified 20 integral problems into categories of perceived similarity based on how they would solve the problems, and (b) students' brief explanations of each category. In each round, the same problems were used, and students were allowed to use as many or few groups as they wished. After each round, the students received some feedback on their classification schemes, although the feedback varied somewhat from round to round (unlike the method of double stimulation). There was no formal instruction on the topic until after the fourth round; the fifth round of data was taken at the same time as the end-of-chapter test on the material. For more details about the problems, feedback and student justifications, see Green and Ricca (under review).

Although much care was given to the original choice of problems, during data analysis it became apparent that there were some alternative student classifications of integration problems for which the authors had some sympathy. These sympathetic classifications, although incorrect, increase the difficulty of some analyses, but often yield insight.

3.1. Representation of Data

Because the data consist of groupings of problems, it is appropriate to use *graph theory* to represent the student data. The basic elements of a *graph* are the *nodes* and any *edges* connecting those nodes to one another[2]. In this study, the graphs of data will be sets of 20 nodes, one node for each problem. In this graph, an edge exists between two nodes if those nodes were placed in the same group; two nodes with an edge connecting them are said to be *adjacent*. One way to represent these data is to draw the edges and nodes of the graphs. Figure 1(b) shows this representation of the data shown in Figure 1(a). In addition to the standard notion of edges and nodes, we will define two nodes to have a *gap* between them if they are not adjacent (i.e., not grouped together).

An alternate, and useful, representation of the data is through the use of an *adjacency matrix*.[3] In this representation, each row represents a node and each column also represents a node. The first row and first column each represent node a_1; the second row and second column

[2] This makes the graphs of graph theory qualitatively different from the usual graphs that appear in data analyses. The type of graph will usually be obvious from context; where it is not, the type of graph will be indicated.
[3] There is another matrix related to the adjacency matrix that is sometimes used in data analysis. The other matrix is is the *graph Laplacian* (von Luxburg, 2007), but the structure of our data does not make the graph Laplacian any more useful than the simpler adjacency matrix.

each represent node a_2, and so on. A "1" is placed in a cell to indicate that the row's node is grouped with the column's node. For example, a "1" placed in the 1^{st} cell of the 3^{rd} row indicates that nodes a_3 and a_1 are grouped together. (It is customary among mathematicians when referring to a cell in a matrix to note the row first, and then the column.) A "0" in a cell indicates that the two nodes have a gap between them (i.e., are not grouped together).

Clearly, the upper right triangle of the matrix is symmetrically related to the lower left triangle of the matrix. (E.g., if a_1 is connected to a_3, then a_3 is also connected to a_1. This results in a "1" in both the 3^{rd} cell of the 1^{st} row and the 1^{st} cell of the 3^{rd} row.) Further, it is obvious that each node will be grouped with itself, so that the elements on the diagonal from the upper left corner to the lower right corner of the matrix will all contain "1". For these two reasons, only the lower left portion of the adjacency matrix – that portion that lies strictly below the diagonal - will usually be displayed[4]. Figure 1(c) shows the adjacency matrix of the data in Figure 1(a). Although they are included in Figure 1(c), it is conventional not to include the row and column labels when there will be no confusion caused by their absence.

3.2. Links and Gaps

A quick study of the adjacency matrix of the Figure 1(c) shows that there are 30 links between the nodes, and 160 cells that are filled with zeroes. This adjacency matrix has the maximum number of correct links (30) as well as the maximum number of correct gaps (160). Naturally, student groupings will often differ from the instructor's grouping. Figure 2 shows a sample student grouping. This student's grouping has 13 correct links, and 146 correct gaps.

It is possible (although it did not happen in this study) for a student to look at the collection of problems, exclaim "I don't understand any of this!", and leave all the problems in a single heap. In this case, the adjacency matrix would consist of a "1" in every cell, as every

[4] Note that this means the diagonal is *not* displayed.

problem would be connected to every other problem. This would result in 30 correct links, the student having gotten all of the links that the instructor had, but also 160 incorrect links. This implies that *being correct* is not necessarily the opposite of *being incorrect*. Hence, the data cannot be reduced to a single axis, and it will be natural to then plot students' groupings on two axes: one axis (we chose this to be the horizontal axis) is the number of correct links, and the perpendicular axis is the number of correct gaps. We will refer to this fourth representation of data as a "links-gaps" graph. Figure 3 shows the instructor's groupings (point A) and the data from Figure 2 (point B) plotted on these links-gaps axes. It should be noted that the data on each axis are discrete values. Hence, this representation gives a maximum of 4991 = 161 x 31 possible locations for a student data[5].

The layout of the links-gaps graph axes indicates that a "perfect" score is represented by the point in the upper right-hand corner of the graph, and naively, one would think that "learning" would progress toward that point. While the 15 students each did end round 5 closer to the desired point than where each began, that progress was not monotonic. Figure 4 shows two sample progressions from round 1 through round 5. Not only did no student progress monotonically, of the 60 progressions that can be plotted on these graphs - 4 progressions for each of 15 students - only 9 of the progressions actually are directed toward the upper right.

These student progressions can yield additional information as well. A *pseudo-cluster analysis* can be performed, either on the individual round-to-round progressions, or on the pattern of student progressions taken across all rounds. These pseudo-clusters group together progressions which are close to each other; the criterion used is that the total distance of the clusters is to be minimized; the center of the pseudo-cluster is the average of the angles of the progressions included in that pseudo-cluster. We will refer to these analyses as pseudo-cluster

[5] Not all locations on the links-gaps plot are permissible. E.g., it is not possible to have exactly 29 correct links.

analyses, rather than cluster analyses (Aldenderfer and Blashfield, 1984) for two reasons. First, there is no obvious best distance function to use during the clustering. It is, for example, reasonable to use the differences between the slopes of progressions as a distance when examining round-to-round progressions. However, using the slope yields slightly different clusters compared to using the angle of the slope. Likewise, using the slope based on the change in links or gaps yields different clusters compared to using the change in links or gaps normalized by the maximum correct links or gaps, respectively. Second, given the relatively small sample size ($n = 15$), it is difficult to justify the use of clusters except in the most "obvious" cases. (Even the use of nonparametric statistics turned out to be insufficient for this purpose.) Nevertheless, there are some cases where a pseudo-cluster analysis of the data is helpful.

3.3. Affiliation Networks

The adjacency matrices, along with the student-identified categories, allow for the creation of a third representation of the data: an *affiliation network* (Wasserman and Faust, 1994). Used most frequently in the analysis of social networks, in an affiliation network, items that are grouped together are considered to be connected not to each other (as they are in the adjacency matrix) but to the category to which they belong. For example, the instructor's grouping shown in Figure 1(a) could be represented by the affiliation network shown in Figure 5. The affiliation networks created by students, naturally, are of more interest. Figure 6 shows the affiliation network created by the aggregation of student groupings after round 4. (The instructor's categories are drawn above the problems and other categories below the problems for clarity.) Because different students categorized problems into different groups, some problems are connected to multiple categories; the width of each edge is proportional to the number of students who made that link. Some student categories were considered to be the same even though their descriptions may have varied slightly. (E.g., Categories described as “trigonometric integrals”, “trig integrals”, and “integrals of sine and cosine” were all considered to be the same.)

The use of affiliation networks is preferable in this case to attempts to find the *community structure* of the graphs. A *community* is a subset of nodes of a graph whose members are relatively densely connected to one another, while being relatively not connected to other nodes of the graph. One of the measures of a graph’s community structure is its *modularity*, which compares the community structure to a random graph (Newman, 2006). Modularity can range from -0.5 to +1, with values greater than zero indicating a graph that has a community structure which exceeds the community structure expected by chance. A number of methods exist for detecting the communities in a graph, including *edge betweenness*, a.k.a., “Girvan-Newman”, (Newman and Girvan, 2004), *leading eigenvector* (Newman, 2006), and several others. Community structure, however, may not be easily interpretable, and different approaches to community detection will sometimes give different community structure. Application of seven common uni-modal community detection schemes for unimodal graphs, including *spinglass* (Eaton, 2012) and *Girvan-Newman* (Newman and Girvan, 2004) does find community structure in the data here. However, each of the methods yield different community structures, and the very small positive modularity values (less than 0.01) indicate that, while the communities are not random, none of the communities can be considered to be very well-defined. (This is roughly equivalent to saying that the community structures found were a statistically significant result for a not-too-small *p*-value, and with an extremely small effect size.) While there were occasionally correlations between the communities and the categories of the affiliation network, none of the resulting community structures fully reproduced the structure of the affiliation graphs, and each of the resulting community structures were significantly different than the affiliation graph structures. (E.g., the data in Figure 6 were found, by the Girvan-Newman algorithm to have one community of 13 nodes and seven communities of one node each; the large community consisted of the 13 problems that were connected to the "Substitution" category.) However, as none of the resulting community structures made a change in the resulting interpretations presented in Section 4, these methods were not used.

Not surprisingly, the use of community detection methods for bipartite graphs was more successful than any of the uni-modal approaches, for they use more salient information. However, the two methods (Melamed, 2014; Barber, 2007) used found either the affiliation graph, or a very similar graph, (one with a single problem categorized differently than the most common categorization for that problem on affiliation graph) for each wave of data. These differences also made no change in the results presented in Section 4. Hence, the use of affiliation networks is better for these data than existing community detection schemes, except perhaps for huge networks.

3.4. Edit Distance

Sorting and grouping tasks have often been analyzed using the *edit distance* (Diebel, Anderson, and Anderson, 2005), and it reasonable to apply edit distances to the data in this study. The edit distance is a measure of the similarity or difference between two (graph theory) graphs, and is found be counting the minimum number[6] of links that would have to be changed (edited) to transform one graph into the other. As an example, consider the following student grouping (from round 4) of problems into categories:

"Simple" – $\{a_2, a_3, a_4, c_1, e_2\}$

[6] Finding this *minimum* number can be computationally difficult, however. See Diebel, 2005.

"Integration by Parts" – {c_2, c_3, c_4}

"Substitution" – {a_1, b_1, b_2, b_3, b_4, e_3}

"Partial fractions" – {d_1, d_2, d_3, d_4}

"Mixed" - {e_1, e_4}

This example is an edit distance of 4 from the instructor's grouping, and is at the point (19, 144) on a links-gaps graph. The edit distance is 4 because only the following 4 switches must be made to transform the student grouping into the instructor grouping:

Switch 1 – c_1 into "Integration by Parts"

Switch 2 – e_2 into "Mixed"

Switch 3 – a_1 into "Simple"

Switch 4 – e_3 into "Mixed"

This relatively small edit distance accompanies a grouping that appears very similar to the instructor's, which is a desired property of a measurement.

Edit distances can be calculated not only between the students' groupings and the instructor's groupings, but also between two different student groupings. In the former case, some indication of the progress of students' progress toward the correct solution can be found, while in the latter case, some information about the impact of the various feedbacks provided to the students and some insight into the variability of student work can be gained.

Despite the usefulness of edit distance, a three-dimensional space defined by axes for edit distance, links, and gaps, would still not be sufficient to distinguish all the differences between groupings. There are, for example, situations where two different groupings have the same number of correct links, correct gaps, and edit distance. For example, the following two groupings both have 9 correct links, 143 correct gaps, and an edit distance of 9 (from the instructor's grouping):

Grouping I: {$\mathbf{a_1}$, $\mathbf{a_2}$, $\mathbf{a_3}$, d_1, e_1}, {$\mathbf{b_1}$, $\mathbf{b_2}$, $\mathbf{b_3}$, e_4}, {$\mathbf{c_1}$, $\mathbf{c_2}$, $\mathbf{c_3}$}, {a_4, b_4, d_4}, {c_4, d_3, e_2}, {d_2, e_3}

Grouping II: {$\mathbf{a_1}$, $\mathbf{a_2}$, $\mathbf{a_3}$, $\mathbf{a_4}$, d_1}, {$\mathbf{b_1}$, $\mathbf{b_2}$, c_3, e_4}, {$\mathbf{c_1}$, $\mathbf{c_2}$, e_1}, {$\mathbf{d_3}$, $\mathbf{d_2}$, e_3}, {b_3, c_4, e_2}, {b_4, d_4}

The edit distance from one grouping to the other is 9, which indicates that the groupings are different in some sense. In fact, the two groupings are qualitatively different: In each group, the bold items are correctly linked with one or more other bold items. The first grouping has three groups each with three correctly matched items. The second grouping has one group with four correctly matched items and three groups each with two correctly matched items.

4. Applications to Analysis of Data

The application of these graph theoretic techniques to analyze data facilitates a number of interesting insights, including:

- The difference between surface features and salient features, and the stability of each of these.
- The connections between the feedback given and the development of student conceptual understanding.
- Support for Vygotsky's (1986) heaps-complexes-concepts approach to concept development.
- Support for Thelen and Smith's (1994) notion of on-the-fly creation of concepts.

Although each of these insights can contribute to an understanding of developing systems, none of these insights are gained through the usual longitudinal (including pre-/post-test) schemes. Instead, these insights are gained only because the stability of classifications, the differences between "being wrong" and "not being right", and the use of edit distances, are included in the analyses.

4.1. Surface and Salient Features

Figure 7 shows the affiliation network created by the aggregation of student groupings after round 1, round 4, and round 5. (The instructor's categories are drawn above the problems and other categories below the problems for clarity.) Because different students categorized problems into different groups, some problems are connected to multiple categories; the width of each edge is proportional to the number of students who made that link. Some student categories were considered to be the same, even though their descriptions may have varied slightly. (E.g., Categories described as "trigonometric integrals", "trig integrals", and "integrals of sine and cosine" were all considered to be the same.) Further, some students, because of their course history, knew of the various approaches to integration, and sometimes would use those categories, although not always correctly. As might be expected, more students used the instructor's categories as the rounds progressed. (Recall, however, that no explicit instruction other than the feedback was given between round 1 and round 4.) In other words, as students progressed, they used more *salient* features of the integration problem, and fewer *surface* features, to group the problems. (This is gratifying to the instructor, as it indicates that at least some of the desired learning took place.)

A closer examination of these graphs shows additional interesting features. Figure 8 shows the same data as Figure 7(c), except that the only links that are shown are those used by more than one student. This has the effect of, at least informally, removing outliers from the data, and shows that students overwhelmingly moved toward use of the desired categories (even if they were not always interpreting those categories correctly). A comparison of these figures shows that some groupings appear to be relatively stable: For example, problem b_1 was placed into a trigonometry category even though there was no feedback supporting such a classification. A closer look at the individual student data shows that the students who originally classified b_1 as a trigonometry integral almost all had that same classification in round 5. These and other persistent incorrect classifications can yield some insight into precisely what students are coordinating; see Green and Ricca (under review).

The use of affiliation networks, therefore, allows some insight into whether students are interacting with surface or salient features of a problem. While it is generally presumed that instruction from the teacher should help focus students on particular parts of a problem (Reynolds, 2005), the nature of development may impede that process. It should be noted that only an examination of the affiliation network would show this, as simple testing of whether or not a student could correctly compute an integral would not necessarily identify this. Were an instructor or researcher to request something like “explain your work” or asks “why did you choose this approach?”, then implicitly a part of the affiliation network is being investigated. More insight into the difficulty of focusing students through instruction can be gained by looking at links-gaps plots.

4.2 Connections Between Instruction and Development

A look at link-gap plots for each student’s progressions through the rounds is also enlightening. (See Figure 4.) Two interesting features can be seen in these plots. As noted above, relatively few of the progressions (only 1 of 8 in Figure 4, and only 9 of 60 in the complete data set) are directed toward the upper right-hand corner. Although none of the progressions shown in Figure 4 were directed opposite this desired progression (i.e., toward the lower left-hand corner of the graph), student progress was nonmonotonic. Further analyses of these progressions are interesting.

A pseudo-cluster analysis (see above) was used to look for underlying structure in the data. The slopes and magnitudes of the student progressions from round to round were scaled to make comparisons of progressions with different starting points comparable. Pseudo-clusters were based solely on the direction of student progression from round to round.

Consider the four corners of the links-gaps graphs: upper right (the instructor's solution), upper left (each problem in its own group), lower left, and lower right (all problems in a single group). As noted above, of the 60 student progressions from round to round only 9 were directed toward the instructor's grouping (i.e., being a vector pointing into the first quadrant). A number of the progressions, however, could be considered as representing an "improvement" in student understanding as they were directed with a slope angle that pointed along an arc that we will refer to as the *arc of improvement*, the arc from the upper left corner through the upper right corner to the lower right corner; there were 43 (out of 60) such clusters. Figure 9 shows a representation of the student progressions from round 1 to round 2: Each student's progression is represented by an arrow, and the slope and magnitude of the arrow represent the slope and magnitude of the change on the links-gaps plots. In this diagram, a cluster whose center angle is directed toward the right can be visually identified; this cluster corresponds to the pseudo-cluster found through calculations.

A pseudo-cluster analysis of the data from round to round revealed two statistically significant pseudo-clusters containing more than 1/3 of the students: a pseudo-cluster of 10 students moving almost directly to the right from round 1 to round 2 and a similar pseudo-cluster of 10 students progressing from round 3 to 4. In both cases, the angle of the pseudo-cluster center was slightly negative (-6.2° and -8.7°, respectively). In other words, the two pseudo-clusters showed improvement in identification of correct links, but with a slight diminishment in the identification of correct gaps. Further, the analyses revealed only two significant pseudo-clusters with a pseudo-cluster center angle in the 1st quadrant; neither of these was very tightly clustered, and neither was very large: they consisted of 4 and 3 students, respectively.

A pseudo-cluster analysis of the progression from round 1 to round 5, however, was substantially different. Four distinct pseudo-clusters were found:

- Cluster 1: 5 members, cluster center angle = 42°, magnitude = 0.22
- Cluster 2: 6 members, cluster center angle = -24°, magnitude = 0.21
- Cluster 3: 2 members, cluster center angle = 98°, magnitude = 0.21
- Cluster 4: 2 members, cluster center angle = -180°, magnitude = 0.07

Overall, 13 of the 15 students improved their groupings, as indicated by the cluster center angles pointing along the arc of improvement; 5 of the 15 progressed toward the instructor solution, while the remainder of the 13 students did better in identifying links or gaps, but not both.

Given the existence and characteristics of these pseudo-clusters – no uniform progress in any progression, but relatively large overall progress – some interpretation is necessary. Because of the relatively small sample size, these interpretations may lack generalization, but coupled with all of the data analyses in this study, these interpretations should be seen as valid. The first general interpretation is that the feedback given tends to have an overall effect on students: in two of the three progressions where feedback was given to the students, approximately 2/3 of the students moved in the direction of more correct links. Hence, it does appear that students respond favorably to some types of feedback. However, as none of the progression clusters involved more than ¾ of the students, it appears that there may be an upper limit on the efficacy of some classroom interventions; this provides evidence of the need to differentiate instruction even at the level of a second-semester Calculus course[7].

The directions of the two largest pseudo-clusters indicates an increase in the number of correct classifications, but relatively little change in the number of incorrect classifications. This first gives further support to the relative independence of correct and incorrect classifications, and hence, to the idea that learning (at least in this situation) is not unidimensional. The move to the right on the links-gaps graph indicates relatively stable (although not necessarily correct) student categories. The slight downward trend indicates that this stable group has more mis-grouped problems than the group from which the new member was taken. This implies that the feedback given in these two cases (which, both times, involved the results of performing the integration, and the second of which included the identification of the method that was used) does little to break up student misconceptions[8].

The third pseudo-cluster, found in the progression from round 4 to round 5, (when there was no direct instruction, as the course had moved on to other topics,) was striking. Without classroom feedback, students tended to improve largely by removing incorrectly linked problems from larger groups. Further, this cluster was quite tightly grouped around its center angle. It is possible that more time is necessary for removing links because of the slower pace of student reflection or metacognition, or because the process of breaking a connection is harder, but interpretation of this result will require additional work.

The substantial difference between the results of any given links-gaps progression from round to round and from the start to the end provides further evidence that learning requires a

---

[7] Perhaps this is the educational analog of a quotation attributed to Abraham Lincoln: "You can teach some of the students all of the time, and all of the students some of the time, but you can't teach all of the students all of the time."

[8] The authors acknowledge that there is not universal agreement in the literature on whether to use *misconception*, *naïve conception*, *prior conception*, etc. We have chosen arbitrarily to use the term *misconception*.

number of different processes. Taken as a whole, the students in the class appeared to learn from beginning to end; however, a pre-/post-test approach would only capture that there is a difference between beginning and ending student knowledge about integration, and would fail to capture the potentially important pseudo-cluster information and the connection between the type of feedback given and the change in student understanding. Learning does not appear to occur along a single axes, nor is all learning prompted by the same external pressures; learning requires the coordination of many components.

4.3. Heaps, Complexes, and Concepts

In addition to these insights into learning, a closer analysis of the groupings is interesting when compared to Vygotsky's approach to concept development. First, most students showed some obvious (to the authors) coherence in their groupings. Hence, it is unlikely that many students had groupings that could be considered heaps. While it can be argued that a category labeled "other" or "miscellaneous' or "I don't know" can be considered heaps, there were only 10 of these among the 362 groups[9] created over the 5 rounds. Four of the 10 occurred in the first round; 1 of those 4 groups (and 1 of the remaining 6 "other" groups) contained a single element, and therefore cannot really be considered a heap.

We would expect groupings into heaps would be very unstable, as they would have been essentially random groupings to begin with. Of the 37 nodes included in the 8 possible heaps, only 8 of those nodes were connected in the subsequent round, and none of those nodes were included in a subsequent group that was labeled "other", etc. This change in groupings is relatively large compared to the other groupings, and this is an indication that student at least, at some level, some students used heaps of problems as a grouping. While the argument could be

[9] There were actually 374 groups created, but one student did not follow directions, and failed to give category names in two rounds, resulting in 12 unlabeled groups. Examining the created groups shows some coherence to the authors, however, because there were no student labels, it was decided to ignore those categories in this situation.

made that students grouped the problems together merely to complete the task, such a grouping would still fulfill Vygotsky's definition of heap: the reasoning was essentially random relative to the task. Additional data, in the form of interviews with students or closer observation of the process of grouping would be needed to make a definitive judgment.

Determining the difference between complexes and concepts is more difficult. Stability of grouping is, unfortunately, not a good indicator of the difference between all complexes (especially pseudo-concepts) and concepts: Some misconceptions are widely known (Dole and Sinatra, 1998) to be hard to change, and one student in this study consistently used the same surface level features throughout the course to group problems. Although the authors do not claim that all misconceptions are complexes rather than concepts, for the purpose of grouping problems, it is reasonable to treat misconceptions as complexes: Complexes are *concrete and factual*, which is true of the groupings that consider only surface level features, rather than *abstract and logical* (Vygotsky, 1986). Hence, those students who used surface level features in their groups created complexes.

A more useful distinction between complexes and concepts, however, comes from the affordances (Gibson, 1977) of the groupings. In particular, concepts are more flexible (Bruner, 1966a, 1966b) and could allow students to actually perform (or at least attempt with a legitimate method) the integration. A further examination of student misconceptions reveals more detail. Figure 10 shows the same data as Figure 7(c) but this time with three changes. First, only those edges that were seen on the responses of more than two students are shown. Second, all the correct remaining edges were removed. Third, nodes that were then not connected to a group were removed from the figure. The result demonstrates the existence of some groupings that are attractive to multiple students, are reasonable (on the surface), but are incorrect (because they reflect an incorrect understanding of integration). We will refer to these groupings as *sympathetic*. One such sympathetic grouping in Figure 10 is the inclusion of problems b_2, b_3, and c_2 in the group of regular integrals. It is reasonable (although wrong) for students to place those edges into the “regular” integral group because each of those integrals is among those that are included in the tables of integrals on the inside cover of the text that was used for the course. However, despite their inclusion in that reference, none of these three integrations can be performed by simple means. This sympathetic group, then, indicates the use of a (Vygotskian) complex rather than a concept: the complex does not have as one of its affordances guidance on how to perform the problem. First, students with these sympathetic groupings would be unable to perform the steps required to integrate problem c_2 (because of their failed classification) which is what occurred in this course. (Students with conceptual groupings would be able to attempt the integration, but may fail to complete the integration for other reasons.) Second, and perhaps more importantly, is that the concrete nature of a complex is not useful in other contexts. (Most of the students in this study, because they were not math majors, are unlikely to ever encounter those other contexts, but that an issue of curriculum, and not of learning or development.)

A second type of sympathetic grouping can also be found in the data, particularly with regard to the problems of the E-group. Most of the E-group problems require the sequential use of multiple approaches of B, C, or D. In many cases, students classified E group problems according to which technique would be used first in the sequence. While it is tempting to follow similar analyses of these sympathetic groups, it is unclear that such analyses are appropriate: The data have too much variation for the relatively small sample size, additional data would be needed to know if a student’s use of a grouping indicated a concrete approach or not.

Sympathetic groupings, although important for analyzing student learning, are not typically available from community detection schemes. (Only one of the attempts to use community detection schemes in the current study found a sympathetic group.) Hence, failure to work with the affiliation groups as was done in this study would most likely cause a decrease in the ability to understand student learning. (Again, it is possible that for very large networked data sets the use of a community detection scheme may be more efficient, but it is likely that the edges from the "correct" groups and the idiosyncratic edges would have to be removed first.)

4.4. Construction "On the Fly"

Table 1 shows a summary of the edit distances over the five rounds of data collection. From those data, it appears that, overall, the students in the study are "learning" because they are getting closer to the "correct" answer. However, there are two difficulties with this assessment.

The first problem with assuming that students have "progressed" in their understanding is the difficulty in interpreting the significance of the changes in edit distances. While it is possible

	Round 1	Round 2	Round 3	Round 4	Round 5
Mean edit distance	10.40	10.00	9.47	8.67	8.20
Standard Deviation	1.24	1.89	1.68	1.91	1.70

Table 1. Summary of edit distances, measured from the student groupings to the instructor grouping, across the five rounds.

to compute statistical significance for the edit distances (and doing so indicates significance at a $p = 0.05$ level from round 1 to round 5) a closer analysis indicates some difficulty in using a naïve interpretation of significance. Table 2 shows the edit distances between the rounds. Notice that the inter-round edit distances are much larger than the standard deviation of the edit distances from each round to the instructor's groupings. Hence, the variation used in calculating statistical significance from the beginning to the end is not representative of the variation that occurs during learning; it is likely to be a substantial underestimate. Given the extreme variability exhibited by students here, the authors suspect that the general reliability of educational measures is overstated[10]. Therefore, any statistical significance (or effect size) calculated is suspect.

	Round 1 to Round 2	Round 2 to Round 3	Round 3 to Round 4	Round 4 to Round 5
Mean edit distance	7.73	9.67	10.00	9.40
Standard Deviation	3.20	1.88	1.60	1.12

Table 2. Summary of edit distances, measured from the student grouping in one round to the student grouping in the subsequent round.

Second, and perhaps more important, is that there is no clear indication, from either theory or data, about the relative weights to be given to links and gaps. Although we have treated links and gaps equally, Carlsson (2009) noted that when there is no natural measurement of distance, then the use of that distance in analysis is arbitrary, and may not be appropriate for the data. Hence, the examination of links-gaps plots, and the corresponding edit distances, indicates that the interpretations of the learning progressions are to be treated with skepticism. (This problem does not exist in a uni-dimensional approach to data analysis, although the variations between rounds would still make the interpretation of statistical significance suspect in those cases.)

As noted in the above analysis of edit distances related to Tables 1 and 2, there is clearly a great deal of variability during learning. While not unexpected – Barsalou (1983) and Thelen and Smith (1994) both highlight variability during learning and the performance of tasks - the magnitude of the variability given by edit distances indicates that commonplace studies of learning may be radically underestimating variability and hence, overstating the significance of pre-/post-test results *as indicators of learning and student understanding*. For example, a simple hypothesis test of student edit distances in this study from the first to the fifth round of data

[10] This problem of reliability is sometimes referred to as the so-called "rubber ruler" problem of statistics: Intervals which are reported as equal in the data are not actually equal.

indicates a significance difference at the $p < 0.01$ level, but the large edit distances from round to round tells a story that undercuts such a claim of significance.

What are we to make of this great variability in the face of (statistically) significant progress? One explanation is that the students are creating their groupings *not* from their grouping in a previous round, but at each stage, they are creating their groupings on the fly – as claimed by Thelen and Smith (1994) – by coordinating some underlying (and unmeasured) components. This, if true, would have powerful implications for teaching: The usual paradigm of teaching is to attempt to build new student knowledge from old student knowledge by adding to an existing behavior or concept. However, if students are re-creating their knowledge at each stage of learning, then the paradigm of teaching must change. The widespread dissatisfaction with the outcome of schooling lends some additional credence to this interpretation, but clearly additional work is needed before such a claim can be fully warranted. However, it is important to note that it is only through the use of edit distances that these insights are gained.

Two last interesting results from the link-gap plot are quite striking. First, approximately six weeks occurred between the completion of the unit on integration (round 4) and the final exam (round 5). During that time, nothing in the course relied on or further developed the material involved in this study. Hence, while students likely reviewed this material for the final, there was no further explicit instruction. Despite this, the scores on the integration section of the final exam were statistically significantly ($p < 0.05$) higher than the scores on similar problems on the unit test. This indicates that, even in the absence of explicit instruction, students were still able to improve their on-the-fly coordination of the elements. In analogy with locomotion, this indicates that perhaps some of the underlying components may have been strengthened by the intervening material: More practice at walking, for example, leads to better jumping because the underlying muscles used in both are strengthened by performance of either.

Second, the scores on the integration section of the final exam were well predicted by students' performance on the five rounds of this study, although not in the manner that one might expect: Of the 11 pieces of data that could potentially contribute to the regression – correct links and correct gaps for each of the five rounds, and the score on the unit test – only two of those pieces were necessary in a linear regression to get a r^2 value of 0.79 and a significance of $p < 0.01$: the number of correct links on round 3, and number of correct gaps on round 1[11]. Note that the performance on integration problems on the unit test was not a significant predictor of performance on the final exam. The full implication of this regression is unclear, although it certainly implies that the assumption of a linear progression of learning through the course is not justified: a linear progression of learning would result in larger contributions to the regression from later rounds.

5. Conclusion

The use of graph-theoretic techniques for data analysis can provide insight into data on developing systems. The use of adjacency matrices can show that seemingly simple data may exist in a multi-dimensional space (rather than the usual assumed single dimension), while the use of affiliation networks can demonstrate changes in student approaches to problem solving, and their use of surface or salient features of a problem. Further, link-gap plots derived from the adjacency matrices provide insight into connections between the learning context (e.g., curriculum or instruction) and changes in student understanding; these also show the possibility of pushing students into undesired metastable states.

[11] Other regressions, such as using a multiplicative combination of independent variables, were also attempted; none of them were any better than this regression.

Edit distances, while not overly useful in a straightforward manner, can highlight the nature of variability during learning, and also call into question the results from some of the usual approaches to significance in learning systems. In addition, an analysis of edit distances supports Thelen and Smith's (1994) notion that actions are coordinated on the fly according to context, which in turn has further implications for curriculum and instruction.

All three of the approaches here were used to refine the common longitudinal approaches to data collection and analysis, including the use of pre-/post-tests, in order to gain more insight into the dynamics of learning. The fruitfulness of the approach indicates that the use of longitudinal analyses is not inappropriate, but rather than the appropriate data sets – ones which rely less on metrics – are necessary to gain more insight into learning and development. In particular, a switch to viewing learning as requiring active categorizing by students rather than application of existing categories presented to students, can provide rich data and deeper analyses. Although we are hesitant to theorize about learning outside of our own discipline, it does appear that the use of the methods presented here could provide insight into any learning situation, provided the data collected can be viewed as groupings by students. This does not seem to be a terribly restrictive requirement however: Any time that a choice is made in a situation, it is equivalent to placing that situation into the chosen category, even if that category is an action. Even a relatively small set of data – this study included 5 waves of data from each of 15 students – seems capable of being fruitful. In this way, the use of graph theoretic methods of analysis indicates potential avenues for further inquiry, particularly when coupled with the use of microgenetic methods (or at least multiple waves of data) to further understand the components that are coordinated in student approaches, both for integration techniques in Calculus, and probably in other areas as well.

6. References

Aldenderfer, M. & Blashfield, R.: Cluster Analysis. Sage Publications, Thousand Oaks, CA (1984)

Barsalou, L.: Ad hoc categories. Mem. & Cognition. 11 (3), 211-227 (1983)

Bruner, J. On cognitive growth: I. in J. Bruner, R. Oliver, & P. Greenfield, eds., Studies in cognitive growth. John Wiley & Sons, Inc., New York, NY, pp. 1-29 (1966)

Bruner, J. On cognitive growth: II. in J. Bruner, R. Oliver, & P. Greenfield, eds., Studies in cognitive growth. John Wiley & Sons, Inc., New York, NY, pp. 30-67 (1966)

Carlsson, G.: Topology and data. Bull. (New Series) of the Am. Math. Soc. 46 (2), 255-308 (2009)

Diebel, K., Anderson, Ruth, & Anderson, Richard: Using edit distance to analyze card sorts. Expert Systems, 22 (3), 129-138 (2005)

Dole, J. & Sinatra, G.: Reconceptualizing change in the cognitive construction of knowledge. Educational Psychologist. 33 (2/3), 109-128 (1998)

Doll, W. Response to Proulx: “Maturana is not a constrctivist”…nor is Piaget. Complicity: An International Journal of Complexity and Education 5 (1), 27-31 (2008)

Eaton, E. & Mansbach, R.: A Spin-Glass Model for Semi-Supervised Community Detection. Proc. of the Twenty-Sixth AAAI Conference on Artificial Intelligence. 900-906 (2012)

Gibson, J. The theory of affordances. In R. Shaw and J. Bransford, eds., Perceiving, acting, and knowing: Toward an ecological psychology. pp. 67-82 (1977)

Gilstrap, D.: Quantitative research methods in chaos and complexity: From probability to post hoc regression analyses. Complicity: An International Journal of Complexity and Education, 10 (1/2), 57-70 (2013)

Green, K. & Ricca, B.: Understanding students' difficulties with integration through a card sort task. Submitted to Investigations in Math. Learning. (under review)

Kamii, C., Clark, F., & Dominick, A.: Teaching to facilitate "Progressive Schematization" or reflective abstraction? The Constructivist, 12 (1), 9-14 (1997)

Koopmans, M.: Nonlinear change and the black box problem in educational research. Nonlinear Dynamics, Psychology, and Life Sciences, 18 (1), 5-22.

Koopmans, M.: Change, self-organization and the search for causality in educational research and practice. Complicity: An International Journal of Complexity and Education, 11 (1), 20-39 (2014b)

Newman, M., & Girvan, M.: Finding and evaluating community structure in networks. Phys. Rev. E. 69 (2), 026113-026127 (2004)

Newman, M.: Modularity and community structure in networks. Proc. of the Nat. Academy of Sci. of the USA, 103 (23), 8577-8582 (2006)

Opfer, V. D.: Methodological implications of a complexity theory-based approach to understanding teacher learning. Paper presented at the Annual Meeting of the Am. Educational Research Assoc. San Francisco, CA (2013, April)

Piaget, J. & Garcia, R.: Toward a logic of meanings. Lawrence Erlbaum and Associates, Rahwah, NJ (1991)

Reynolds, S.: Learning is a verb. Holcomb Hathaway , Phoenix, AZ (2005)

St. Julien, J.: brain research (and practices). Message posted to http://xmca.ucsd.edu/yarns/2054 (1997, March 2). Accessed 26 February 2014

Schoenfeld, A.: Presenting a strategy for indefinite integration. Am. Math. Monthly, 85(8), 673-678 (1978)

Singer, J. & Willett, J.: Applied longitudinal data analysis: Modeling change and event occurrence. Oxford University Press, New York, NY (2003)

Stewart, J.: Calculus: Early transcendentals. Cengage Learning, Independence, KY (2007)

Thelen, E., & Smith, L.: A dynamic systems approach to the development of cognition and action. : MIT Press, Cambridge, MA (1994)

von Luxburg, U. A tutorial on spectral clustering. Statistics and Computing, 17 (4), 395-416 (2007)

Vygotsky, L.: Mind in society. : Harvard University Press, Boston, MA (1980)

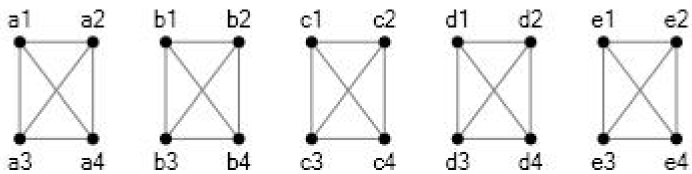
Vygotsky, L.: Thought and language (Revised edition). MIT Press, Cambridge, MA (1986)

Wasserman, S. & Faust, K.: Social network analysis: Methods and applications. : Cambridge University Press, New York, NY (1994)

Figure Captions

$\{a_1, a_2, a_3, a_4\}$ – Basic integrals
$\{b_1, b_2, b_3, b_4\}$ – Integrals requiring the use of substitution
$\{c_1, c_2, c_3, c_4\}$ – Integrals requiring the use of integration by parts
$\{d_1, d_2, d_3, d_4\}$ – Integrals requiring the use of partial fractions
$\{e_1, e_2, e_3, e_4\}$ – Integrals requiring the use of multiple methods

(a)



(b)

	a_1	a_2	a_3	a_4	b_1	b_2	b_3	b_4	c_1	c_2	c_3	c_4	d_1	d_2	d_3	d_4	e_1	e_2	e_3	e_4
a_1																				
a_2	1																			
a_3	1	1																		
a_4	1	1	1																	
b_1	0	0	0	0																
b_2	0	0	0	0	1															
b_3	0	0	0	0	1	1														
b_4	0	0	0	0	1	1	1													
c_1	0	0	0	0	0	0	0	0												
c_2	0	0	0	0	0	0	0	0	1											
c_3	0	0	0	0	0	0	0	0	1	1										
c_4	0	0	0	0	0	0	0	0	1	1	1									
d_1	0	0	0	0	0	0	0	0	0	0	0	0								
d_2	0	0	0	0	0	0	0	0	0	0	0	0	1							
d_3	0	0	0	0	0	0	0	0	0	0	0	0	1	1						
d_4	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1					
e_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
e_2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1			
e_3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1		
e_4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	

(c)

Figure 1 (a) List of integral problems and the categories to which they belong. (b) A network representation of the same data. (The category names are not shown.) (c) The adjacency matrix of the same data. (The category names are not shown; shading is only to assist in reading).The minimum possible number of correct links is 0, as is the minimum number of correct gaps

0																		
0	1																	
0	1	1																
0	0	0	0															
0	0	0	0	1														
0	0	0	0	1	1													
1	0	0	0	0	0	0												
0	0	0	0	0	0	0	0											
0	0	0	0	0	0	0	0	1										
0	0	0	0	0	0	0	0	0	0									
0	0	0	0	0	0	0	0	0	0	1								
0	0	0	0	0	0	0	0	0	0	0	0							
0	0	0	0	0	0	0	0	0	0	0	0	1						
1	0	0	0	0	0	0	1	0	0	0	0	0	0					
1	0	0	0	0	0	0	1	0	0	0	0	0	0	1				
0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0			
0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1		
0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	1	
0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0

Figure 2 Sample student adjacency matrix. (Rows and columns are the same as in Figure 1(c); the upper left "0" is actually row 2 (i.e., problem a_2), column 1 (i.e., problem a_1).)

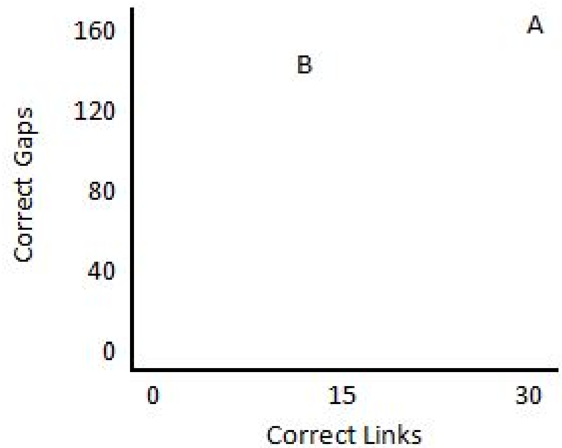
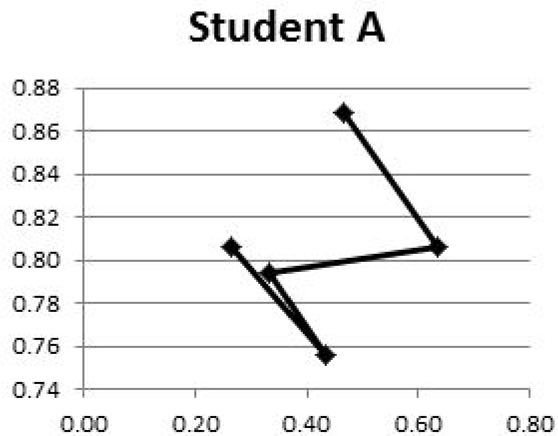
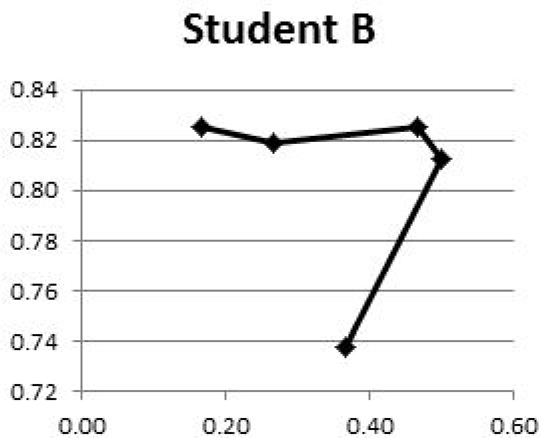


Figure 3 A sample link-gap plot



(a)



(b)

Figure 4 Link-gap plots for two students, shown on axes normalized by the maximum number of links and gaps. (a) This student's round 1 grouping was near (0.25, 0.81), and progressed through round 5, ending near (0.5, 0.87). (b) This student's round 1 grouping was near (0.16, 0.82) and progressed through round 5, ending near (0.38, 0.74). Each plot shows only part of the possible region, and each plot has different scales

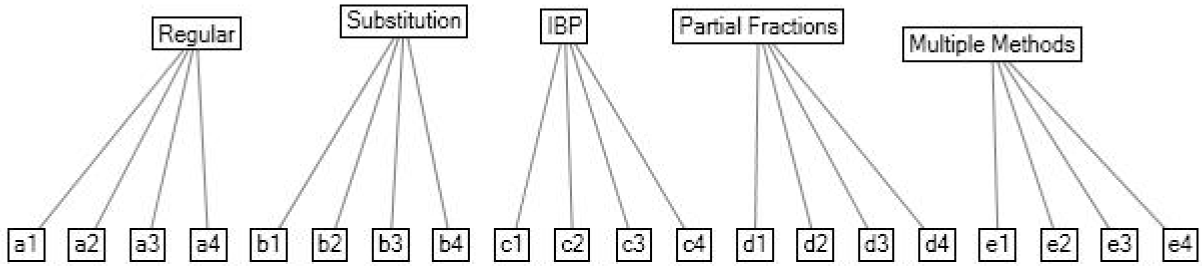


Figure 5 Affiliation network display of the data from Figure 1

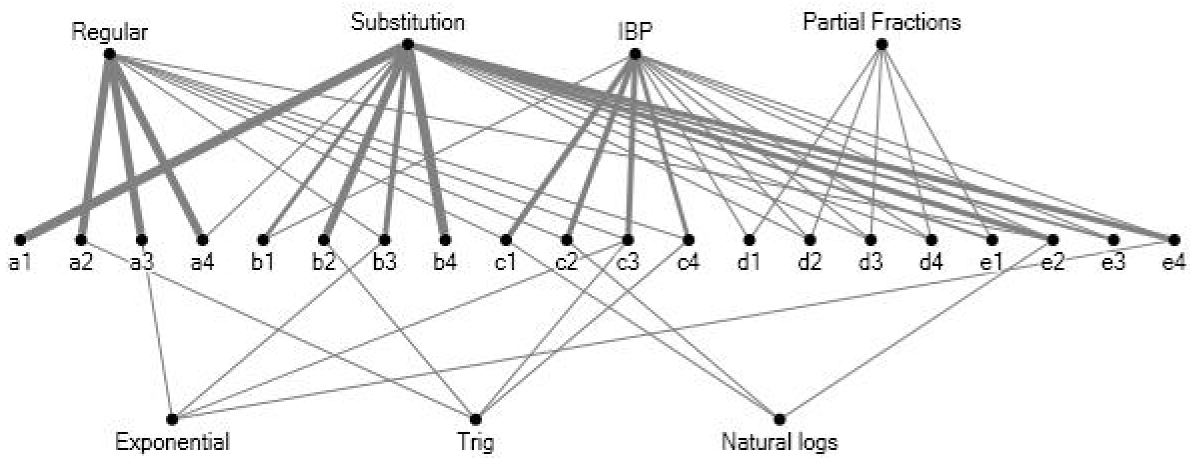
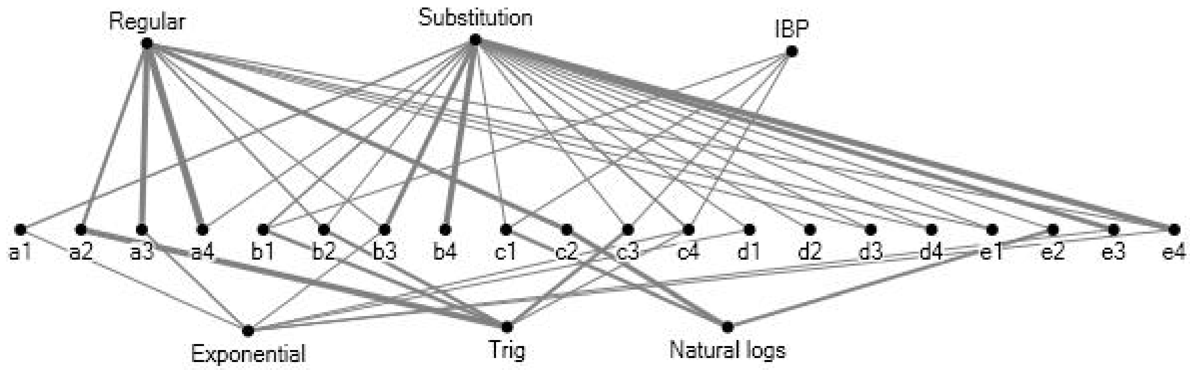
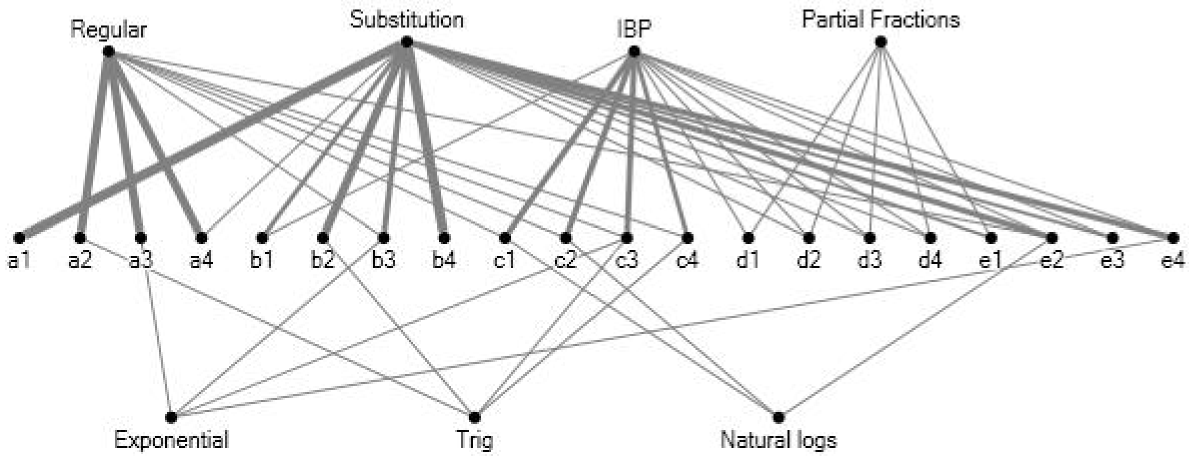


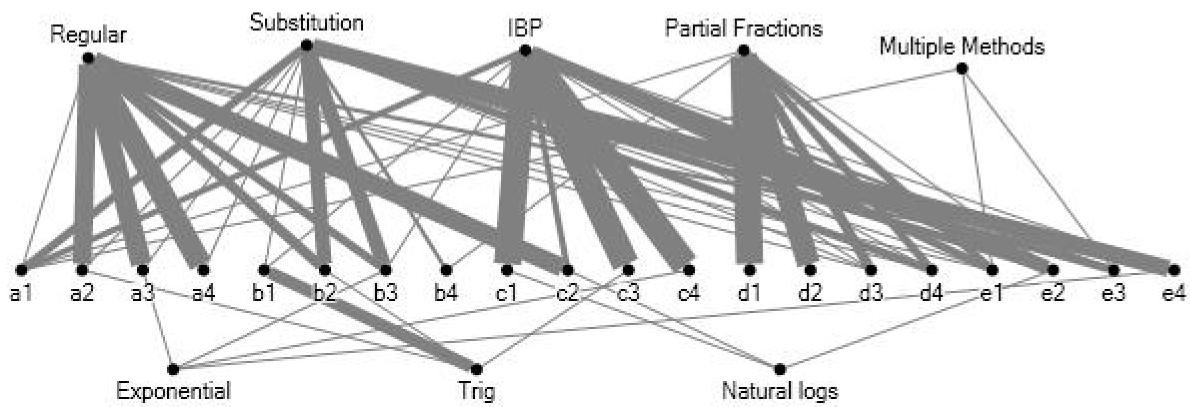
Figure 6 Affiliation network created from the aggregate student groupings of round 4. (The edge thickness between a problem and its category is proportional to the number of students who placed that problem into the associated group. Idiosyncratic groups - those that were only identified by one student - are omitted.)



(a)



(b)



(c)

Figure 7 (a) Student groupings from round 1. (b) Student groupings from round 4. (c) Student groupings from round 5. (The edge thickness between a problem and its category is proportional to the number of students who placed that problem into the associated group.)

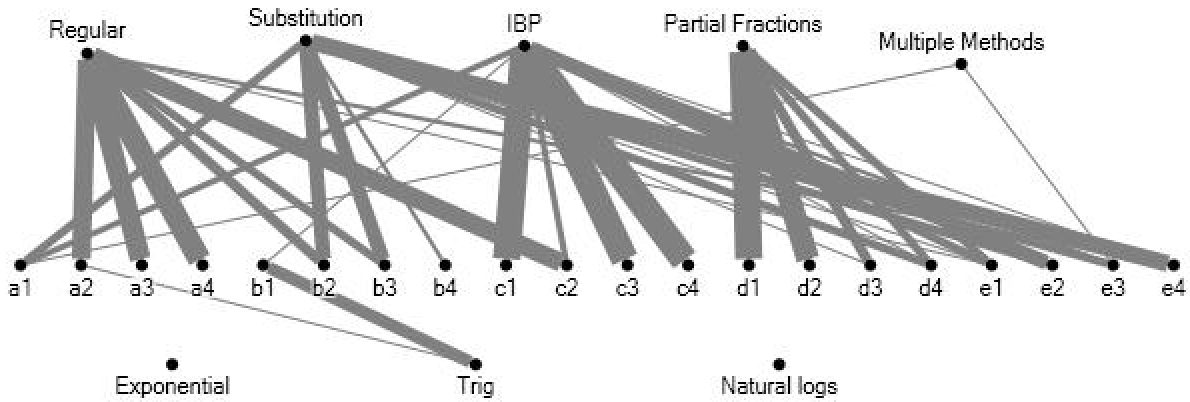


Figure 8 Round 5 data, showing only those links used by more than one student. Note that the use of surface features to group problems has almost disappeared

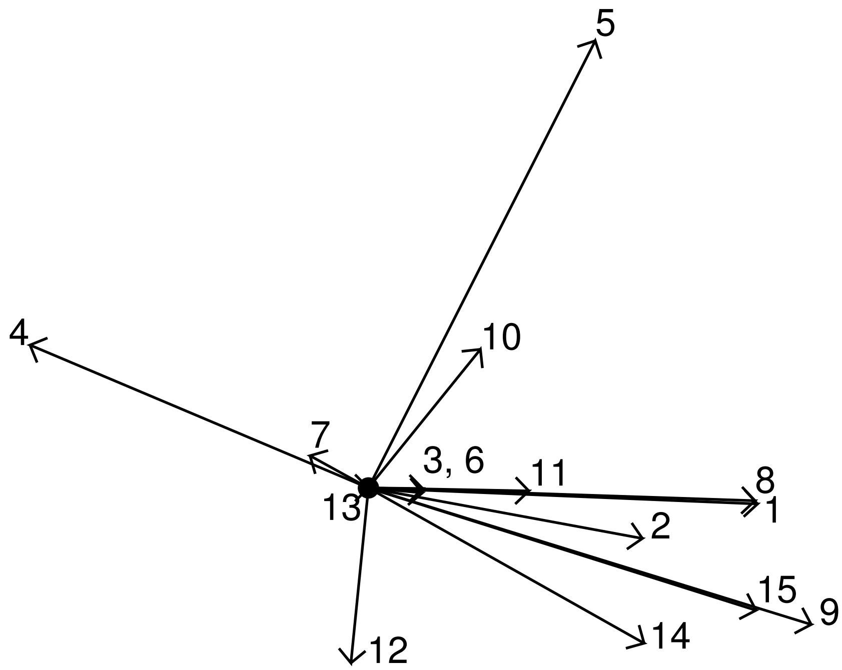


Figure 9 A graphical representation of the 15 students' progressions from round 1 to round 2. The direction and magnitude of each arrow is found from the student change on a links-gaps plot; the directions have been corrected to reflect the effect of the starting position on the links-gaps graph. Labels indicate the student number. Note the clustering of progressions toward the right and slightly down

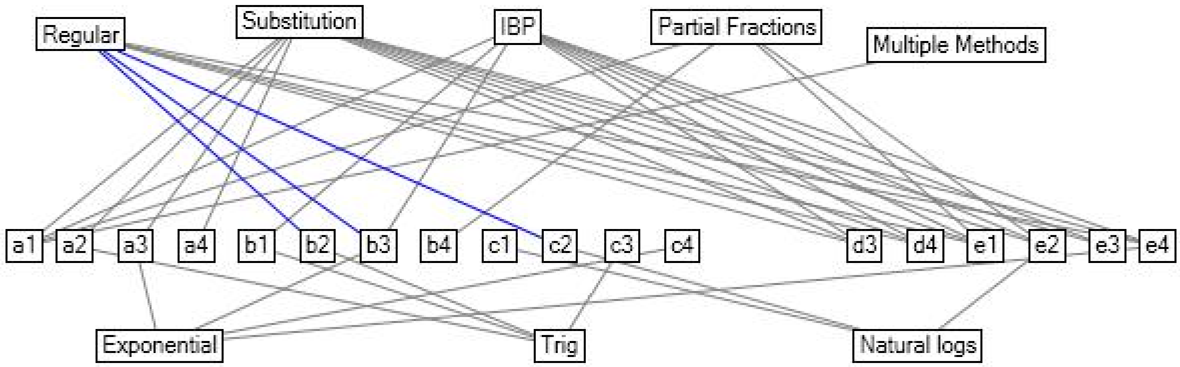


Figure 10 Round 5 groupings, with all some edges and nodes removed (see text). For clarity, all edges are the same width, regardless of how many students used that particular grouping