JEM JOURNAL of EDUCATIONAL MEASUREMENT
NCME national council on measurement in education
Journal of Educational Measurement
Summer 2016, Vol. 53, No. 2, pp. 190–211

Using Networks to Visualize and Analyze Process Data for Educational Assessment

Mengxiao Zhu, Zhan Shu, and Alina A. von Davier
Educational Testing Service

New technology enables interactive and adaptive scenario-based tasks (SBTs) to be adopted in educational measurement. At the same time, it is a challenging problem to build appropriate psychometric models to analyze data collected from these tasks, due to the complexity of the data. This study focuses on process data collected from SBTs. We explore the potential of using concepts and methods from social network analysis to represent and analyze process data. Empirical data were collected from the assessment of Technology and Engineering Literacy, conducted as part of the National Assessment of Educational Progress. For the activity sequences in the process data, we created a transition network using weighted directed networks, with nodes representing actions and directed links connecting two actions only if the first action is followed by the second action in the sequence. This study shows how visualization of the transition networks represents process data and provides insights for item design. This study also explores how network measures are related to existing scoring rubrics and how detailed network measures can be used to make intergroup comparisons.

Advances in technology have expanded opportunities for educational measurement through changes to item design, item delivery, and data collection. Some examples include simulation-, scenario-, and game-based assessment and learning environments (DiCerbo & Behrens, 2012; Mislevy et al., 2014). These new computerized testing formats usually provide an interactive environment for students. To solve problems, students can choose among a set of available actions and take one or more steps to finish a task. All student actions are automatically recorded in system logs (Kerr, Chung, & Iseli, 2011), which can be used immediately for providing instant feedback to students in some cases, or later for diagnostic and scoring purposes in other cases (DiCerbo & Behrens, 2014).

Even in traditional assessment platforms, student actions can be collected with the help of external instrumentation, such as in the collection of keystroke (e.g., Almond, Deane, Quinlan, Wagner, & Sydorenko, 2012) and eye tracking data (e.g., Tai, Loehr, & Brigham, 2006). Data recorded by the system or through external devices reflect students' problem-solving processes and provide more information on *how* students try to solve a problem rather than merely *how well* students solve a problem. We call these data *process data*. Valuable information can be extracted from process data on students' problem-solving strategies or action patterns.

With the availability of rich response process data during problem solving comes the great challenge of building appropriate psychometric models to analyze these data. No matter whether they are system logs from scenario-based or game-based assessment tasks or data from keyboard or eye tracker records, the raw process data

 Copyright © 2016 by the National Council on Measurement in Education

17453984, 2016, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12107 by Arizona State University Acq & Analysis, Lib Continuations, Wiley Online Library on [04/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

are usually formatted as lines of coded and time-stamped strings. The vast amount of data on student's potential trial-and-error process makes it less than straightforward to detect patterns in problem solving. For scoring purpose, scoring rubrics and the evidence-centered design (ECD) framework (Mislevy et al., 2014) have been considered and used for scenario-based or game-based assessments. Several data analysis techniques and models have been explored to uncover problem-solving patterns. Here are a few examples. Inspired by methods in machine learning, researchers used methods such as cluster analysis (Bergner, Shu, & von Davier, 2014) and editing distance (Hao, Shu, Bergner, Zhu, & von Davier, 2014). Rooted in psychometric studies, researchers explored the potential of combining the concept of Markov models and item response theory (IRT) framework (Shu, Zhu, Hao, Bergner, & von Davier, 2014) in analyzing process data. Using methods merged from business process management, process mining techniques such as Petri net were also used to study behavioral patterns (Howard, Johnson, & Neitzel, 2010). More generally, researchers (DiCerbo, Liu, Rutstein, Choi, & Behrens, 2011) used digraphs to visualize and analyze sequential process data collected from assessment. As an effort to extend studies in the last direction, this article provides a generalizable method for representing and analyzing process data using networks that does not include complex rules as in Petri net. This article also introduces measures and analysis of global features and local patterns in networks that enable the application of network method to go beyond visualization.

With the observation that actions in process data are usually not isolated events but rather interconnected with each other, this study introduces methods and tools from network studies to visualize and analyze process data. In problem-solving processes, it is often observed that a student's future actions are impacted by his/her prior decisions. In process data, networks can represent sequences of actions, with actions as nodes and with directed links between actions representing the order of the actions. This study explores the application of network visualization and analysis as tools for understanding process data.

The remainder of this article is organized as follows. In the next section, we formally define process data and introduce related network models, measures, and analysis tools using simple examples. We then describe a case study using data from a scenario-based task from the National Assessment of Educational Progress (NAEP) Technology and Engineering Literacy (TEL) assessment. In the final section, we discuss network models for process data, their advantages, and their limitations.

Visualizing and Analyzing Process Data Using Networks

The process data considered in this article are defined as a series of actions that individuals take in the problem-solving process used during educational learning or assessment. The major difference between process data and traditional item-response data, such as data from multiple-choice items, is that process data capture both a student's final solution to an item and, more importantly, the student's problem-solving process. Typically, process data can be represented using a sequence of actions. Each of these actions belongs to a finite set of available actions. Through this study, we analyze process data with the goal of exploring process data and finding patterns that

17453984, 2016, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12107 by Arizona State University Acq & Analysis, Lib Continuations, Wiley Online Library on [04/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

may exist in the problem-solving process, while taking the sequential features into consideration. We also explore how patterns in process data are related to the scores.

As a general representation, the process data collected from a single student during a scenario-based task are defined as a string of actions in order, $S = s_1, s_2, \ldots, s_n$, with each action s_i belonging to a predefined action set A, that is, $s_i \in A = \{a_1, a_2, \ldots, a_m\}$. Each action may be repeated multiple times in the action sequence, depending on the system constraints and the student's decisions. For instance, in some systems, some actions are disabled by design after they have been executed for a certain number of times. On the other hand, some students are more curious than others to try the same action multiple times. Here is a simple example of an action sequence: $S_1 = a_1\ a_2 a_1 a_3 a_3 a_4$. This action sequence has partial time information embedded, that is, it is recorded in order so that the actions in the front of the string are followed by the actions that occur after them. In this article, we do not discuss or explicitly make use of the information from the time stamps.

Network Representation and Visualization for Process Data

In process data, the action sequences are not an aggregation of independent activities; instead, they indicate the order of the activities taken by the students while solving problems. For instance, previous activities may potentially influence a student's future decisions. This article proposes to use networks, also called graphs (Bondy & Murty, 2008), to represent and visualize the sequential interdependence of actions in the process data. A single action sequence or multiple action sequences can be represented using networks indicating transitions of activities. In particular, weighted directed networks, also called valued directed graphs or weighted digraphs (Wasserman & Faust, 1994), are a natural choice. The reason is that they enable one to preserve the sequence of moving from one action to another, while accommodating the possibility for the same action transition pattern to appear multiple times in one sequence or an aggregated sequence set.

A weighted directed network $\mathcal{G}$ is defined by a set of nodes $\mathcal{V} = \{v_1, v_2, \ldots, v_g\}$, a set of links $\mathcal{L} = \{l_1, l_2, \ldots, l_t\}$, and a set of weights on the links $\mathcal{W} = \{w_1, w_2, \ldots, w_t\}$. This network can be described by a $g \times g$ adjacency matrix $\mathcal{M}$, in which rows and columns represent nodes and cells are the number of links between the node represented by the row and the node represented by the column.

To develop a weighted directed transition network from an action sequence, the basic idea is to use nodes in the network to represent actions and directed links to represent action transitions. To explicitly show the beginning and the end of a sequence, we include both the predefined action set A and two extra nodes, "Start" and "End," in the node set $\mathcal{V}$, that is, $\mathcal{V} = A \cup \{Start, End\}$. A directed link set $\mathcal{L}$ is created to represent transitions between actions. For instance, a link is created from the node representing action a_1 to the node representing action a_2 if a_1 is followed by a_2 in the sequence. A link from node a_1 to node a_2 is different from a link from node a_2 to node a_1. Table 1 shows the adjacency matrix for the simple example of an action sequence $S_1 = a_1\ a_2 a_1 a_3 a_3 a_4$. In this example, there are four actions, and thus, together with the two additional Start and End nodes, a 6×6 matrix is created. Each cell represents the frequency of transitions from the row action to the

17453984, 2016, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12107 by Arizona State University Acq & Analysis, Lib Continuations, Wiley Online Library on [04/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1
Adjacency Matrix of $\boldsymbol{S}_1(a_1a_2a_1a_3a_3a_4)$

	Start	a_1	a_2	a_3	a_4	End
Start	0	1	0	0	0	0
a_1	0	0	1	1	0	0
a_2	0	1	0	0	0	0
a_3	0	0	0	1	1	0
	0	0	0	0	0	1
End	0	0	0	0	0	0

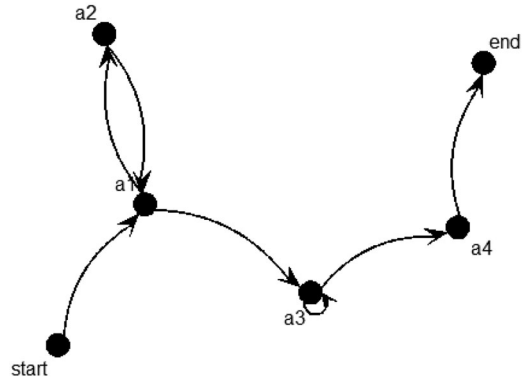

Figure 1. Network visualization of $S_1(a_1a_2a_1a_3a_3a_4)$.

column action. In this example, each action transition happens only once, that is, link weights w_is are all 1s, but in other cases, if an action transition happens more than once, link weights w_is can be represented using the numbers in the corresponding cell, indicating how many times that transition was observed.

To present aggregated patterns from a subgroup of students, transition networks can also be built to include multiple action sequences. The most straightforward way is to use the weights of the links to indicate how frequently action transition patterns were observed over the subgroup. An alternative process is to build the individual transition networks first, and then add the network matrices together to get the overall transition network. We will give examples of subgroup networks in the example section of this article.

The benefit of using transition networks to represent process data is twofold. First, process data can be graphically presented by plotting the transition networks, which provides visual access to hidden patterns in process data. Figure 1 shows the visualization of the action sequence $S_1 = a_1\ a_2a_1a_3a_3a_4$ using a Fruchterman and Reingold layout (Fruchterman & Reingold, 1991). This layout, which is a variation of the force-directed placement method, places nodes with stronger connections (in terms of the number and the strength of connections among nodes) closer together in the two-dimensional space. Compared to character strings, network visualizations are

17453984, 2016, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12107 by Arizona State University Acq & Analysis, Lib Continuations, Wiley Online Library on [04/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

relatively easier for human eyes to process. By observing the patterns in the transition network, useful information about the item design and common pitfalls can be identified. For instance, if a certain link is extremely highly weighted but is not part of the correct answer, it could signal that there is a potential design problem with the task or that the task is exposing a common problem-solving misconception.

The second benefit of representing process data using networks is that many existing tools and methods in social network analysis (SNA) can be directly applied. Originating in sociology and later adopted by researchers from biology, physics, computer science, and many other fields (Newman, 2003), SNA provides many useful measures and methods (e.g., Wasserman & Faust, 1994) for analyzing interconnected data. One of the examples of SNA is the study of the phenomenon of the so-called *small world* (Watts & Strogatz, 1998). By collecting and analyzing data on human connections, scientists found surprisingly that humans in a social system are connected through a very small number of "hops." Similar highly locally clustered and globally connected structures are also observed in many natural and human-constructed systems. To analyze and characterize patterns in transition networks generated from process data, several global network measures, such as density and centralization, and measures of local patterns, such as reciprocity and triad census (Davis & Leinhardt, 1972; Wasserman & Faust, 1994), can be very useful. The next section introduces these measures in detail.

Network Measures for Transition Networks

Transition networks generated from process data are directed and weighted to preserve the sequential information and to accommodate repeated action transitions. Thus, we focus on corresponding measures for directed weighted networks. The most basic measure is the degree of a node. For weighted directed networks, each node has both in-degree d_I, which measures the number of incoming links multiplied by their weights, and out-degree d_O, which measures the number of outgoing links multiplied by their weights. Due to the construction of the transition networks connecting two adjacent actions, it can be easily seen that $d_I = d_O$ for all action nodes, with the exceptions of the two added indicator nodes of Start and End. The in-degree and out-degree for the Start and End nodes are always fixed, with $d_I = 0$ and $d_O = 1$ for the Start node, and $d_I = 1$ and $d_O = 0$ for the End node for transition networks generated from one transition sequence. Without losing much information we are interested in, we consider only measures using in-degree and omit the equivalent measures using out-degree. In the current analysis, we also exclude self-loops for two reasons. First, self-loops in the current setting represent immediate repetitions of the same actions, while this study emphasizes transitions between different actions. Second, self-loops may introduce unnecessary mathematical complications in network measures and are usually not considered in the literature (Wasserman & Faust, 1994).

Global measures. To capture global features of the transition networks, we consider two commonly used network measures, weighted network density, and centralization. It is worth noting that in networks these terms are defined differently from terms with similar names in general statistics. Briefly, the former captures the

17453984, 2016, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12107 by Arizona State University Acq & Analysis, Lib Continuations, Wiley Online Library on [04/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

overall chance of observing links in the network, and the latter captures the extent to which nodes differ from each other in terms of their connections in the network.

Weighted density. For unweighted networks, where links either exist or do not exist between two nodes, the density of a network is defined as the number of links divided by the total number of possible links. For a directed network with g nodes and t links, there can be a maximum of $g(g-1)$ links. (In directed networks, a link from node v_1 to node v_2 is different from a link from node v_2 to node v_1.) The corresponding density is $\frac{t}{g(g-1)}$, which is between 0 and 1.

For a weighted directed network, there are $g(g-1)$ possible links, and all links should be weighted by their values w_i; thus, the weighted density (Wasserman & Faust, 1994):

$$D_w = \frac{\sum_{i=1}^{t} w_i}{g\,(g-1)}\,.$$

Since the link weights can be larger than 1, the values of the weighted density can be larger than 1. Theoretically, density for a weighted and directed network captures the average strength of the links. In the analysis of process data, given the fixed number of possible actions, for one action sequence the weighted density captures the average frequency of transitions between actions. If compared across multiple action sequences, the weighted density also indicates the relative length of the sequence.

Centralization. Centralization measures the extent to which the nodes in a network are different from each other in terms of their importance. The importance of nodes can be measured in several different ways, the most basic being *degree centrality* (Freeman, 1979; Wasserman & Faust, 1994), which is calculated by counting the number of connections to or from the focal node, that is, the degree of the node. As discussed above, in the transition networks generated from process data, for all action nodes in-degrees are always equal to their out-degrees. Here, we calculate the degree centrality using in-degree only, that is, $C_D\,(v_i) = d_I\,(v_i)$. For process data, the degree centrality is a node-level measure and captures how popular a certain action is among all available actions. Degree centralities for nodes can be used to capture the most and least frequent actions observed in action sequences.

Centralization is thus a global measure that captures the variability of the node-level indices. We follow the definition used by Freeman (1979). For a network $\mathcal{G}$, the degree centralization is given by

$$C_D = \frac{\sum_{i=1}^{g}\left[\max_{v\in\mathcal{V}} C_D\,(v) - C_D\,(v_i)\right]}{\max\sum_{i=1}^{g}\left[\max_{v\in\mathcal{V}} C_D\,(v) - C_D\,(v_i)\right]}\,,$$

where $\max_{v\in\mathcal{V}} C_D(v)$ is the largest value of degree centrality $C_D(v_i)$ for any node in the network; the numerator is the sum of the difference between each node's degree centrality and the largest value; and the denominator $\max\sum_{i=1}^{g}[\max_{v\in\mathcal{V}} C_D(v) - C_D(v_i)]$ is a normalizing factor, calculated as the maximum possible sum of differences over all possible networks with the same

17453984, 2016, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12107 by Arizona State University Acq & Analysis, Lib Continuations, Wiley Online Library on [04/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

number of g nodes. The maximum possible sum of differences in this case can be achieved when the network is structured as a star, with one node connected to all other nodes and the rest not connected to each other. The corresponding theoretical maximum is $(g-1)^2$ for directed networks. The minimum value of the centralization score is 0, which indicates that all nodes have the same degree (Freeman, 1979). The higher the centralization score, the more unequal are the degrees. In the analysis of process data, the measure of centralization captures the dispersion of the attention or attempts given to different actions. A lower centralization score indicates that the student took all actions about the same number of times, while a higher centralization score indicates that the student favored some actions more than others.

Dyadic and triadic local patterns. Besides global features on all actions, we are also interested in local transition patterns among actions. In this article, we focus on dyadic (two nodes) and triadic (three nodes) local patterns, because they are the basic building blocks of the whole structure and capture a lot of local dynamics.

Dyadic patterns capture the structures of two nodes and the links between these nodes. Holland and Leinhardt (1970) proposed using a dyad census system, and counted the three possible types of dyads in a directed network, known as mutual, asymmetric, and null dyads. The definitions of these three types of dyads are straightforward. Mutual dyads have two-way links between two nodes; asymmetric dyads have one-way links between two nodes; and null dyads have no links in between. Any dyads in a directed network fall into one of these three categories. In analyzing process data, we may be more interested in the nonnull dyads than in null dyads, that is, the transitions among the actions. So, we propose to use the aggregated measure *reciprocity* (Wasserman & Faust, 1994) to capture the dyadic patterns in the transition networks. In this article, we adopt the definition of reciprocity as the number of mutual dyads divided by the total number of nonnull dyads. For transition networks developed from process data, reciprocity captures the tendency of students to revisit immediately previous actions.

Extending the ideas from the dyad census, *triadic patterns* are captured by a triad census (Davis & Leinhardt, 1972), which provides detailed statistics on all three nodes and link combinations. As shown in Figure 2, there are 16 possible isomorphic patterns for triads. Each pattern is named using the convention developed by Holland and Leinhardt (1970) and Davis and Leinhardt (1972). The first character indicates the number of mutual dyads in the triad; the second character indicates the number of asymmetric dyads in the triad; the third character indicates the number of null dyads in the triad; and a potential extra letter in the end indicates the shape of the triad for the cases when the first three characters are the same. The letter D means down; the letter U means up; the letter C means cyclic; and the letter T means transitive. Taking the pattern 120D in Figure 2 as an example, the name indicates one mutual dyad, two asymmetric dyads, and zero null dyads in the triad. At the same time, the two asymmetric dyads are both pointing downward. For a directed network with g nodes, there are $C_g^3 = \frac{g(g-1)(g-2)}{3!}$ triads, and each one is isomorphic to one of the triadic patterns in the triad census.

In transition networks constructed from process data, these triadic patterns capture transitions beyond immediately adjacent actions. For example, if we observe a large number of 030C structures in a student's transition network, it indicates that this

17453984, 2016, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12107 by Arizona State University Acq & Analysis, Lib Continuations, Wiley Online Library on [04/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

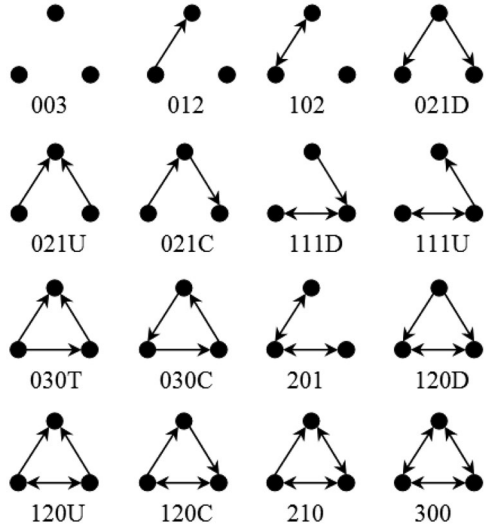

Figure 2. Triad census: triadic isomorphism patterns.

student prefers to solve the problem following some procedures and always comes back to a certain action after conducting the other two actions. On the other hand, the pattern of 021C indicates that a student does not revisit previous steps but goes straight ahead. The triad patterns provide useful measures of action transition patterns for each student and can also be used to compare action patterns between students or groups of students.

The NAEP TEL Wells Task: A Case Study

We illustrate the use of transition networks to visualize and analyze process data collected through one of the scenario-based tasks in the TEL assessment (http://nces.ed.gov/nationsreportcard/tel/) as part of the NAEP project for the National Center for Education Statistics. In general, TEL tasks are problem-solving tasks based on interactive scenarios that reflect realistic situations. The goal of these tasks is to "measure students' capacity to use, understand, and evaluate technology, as well as their ability to understand technological principles and strategies" (Institute of Education Sciences, 2013, p. 1). We use the task called Wells, which was released to the public after the pilot. In this task, students are given basic information on how a hand water pump works, and then they are asked to fix a pump which is not working properly. Interested readers can try it out at the TEL website (http://nces.ed.gov/nationsreportcard/tel/wells_item.aspx).

In order to fix the pump, the students are provided with a repair manual, which lists five potential problems with the pump and related fixes. For each of these five problems, two actions are available, Check and Repair. When the corresponding button is clicked, the former action checks whether or not the pump has a related problem, and the latter action repairs the pump to fix that problem. In the data collected

17453984, 2016, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12107 by Arizona State University Acq & Analysis, Lib Continuations, Wiley Online Library on [04/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

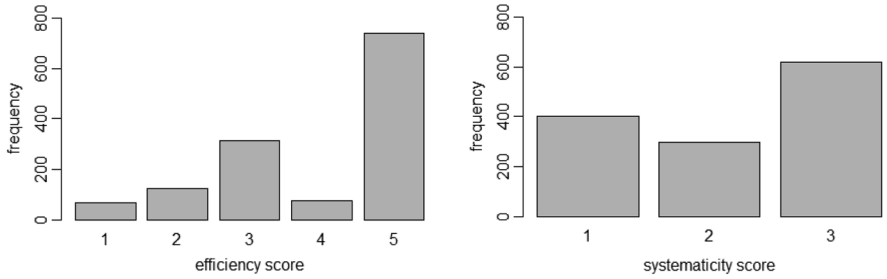

Figure 3. Bar plots of the efficiency and systematicity scores.

on students' problem-solving processes, these actions are coded as C1 through C5 for checking and R1 through R5 for repairing. Each action can be performed in any order but can be performed only once, as the button will be disabled after it has been clicked once. Besides checking and repairing actions, the student can also check whether the pump is working properly using the Check Pump action, which is coded as P in the data. The Check Pump action can be conducted at any time, and there is no limit on how many times this action can be conducted. In total, we have 11 actions available to the students.

The task will end only when the pump is fixed. So, in terms of fixing the pump, all students must fix the pump successfully, which makes it uninteresting to look at whether or not the pump is fixed. Instead, student performance is measured on two dimensions in the scoring rubric related to students' problem-solving processes. The first dimension, efficiency, captures how fast a student fixes the pump. As only problems 4 and 5 exist (this information is made available through several items prior to the current task), actions including C4, C5, R4, R5, and P are considered necessary actions to fix the pump, and actions such as C1, C2, C3 and R1, R2, R3 are considered unnecessary actions. Consequently, students are rated higher if they only take the necessary actions, and lower if they also take unnecessary actions. The more unnecessary actions they take, the lower the efficiency score. For efficiency, students are rated on a scale of 1 to 5. The second dimension, systematicity, captures students' problem-solving procedures. According to the scoring rubric, it is considered a systematic way to fix the pump if a student first checks for a certain problem, then repairs this problem, and then tests whether the pump is working. If a student conducts the repair action before checking the corresponding problem, this set of actions is considered unsystematic. Students are rated on a scale of 1 to 4 on systematicity, with a higher score indicating following the problem-solving routine and a lower score indicating not following that routine. Detailed scoring rubrics are available in the appendix. The data set used in this study contains the problem-solving process data, efficiency, and systematicity scores, and several items of anonymized student information, including gender, on 1,318 students from the eighth grade who participated in the pilot study for the NAEP TEL assessments. The distributions of efficiency and systematicity scores are shown in Figure 3. About half of the students got the highest score, 5, on efficiency and the rest distributed on the other four

17453984, 2016, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12107 by Arizona State University Acq & Analysis, Lib Continuations, Wiley Online Library on [04/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

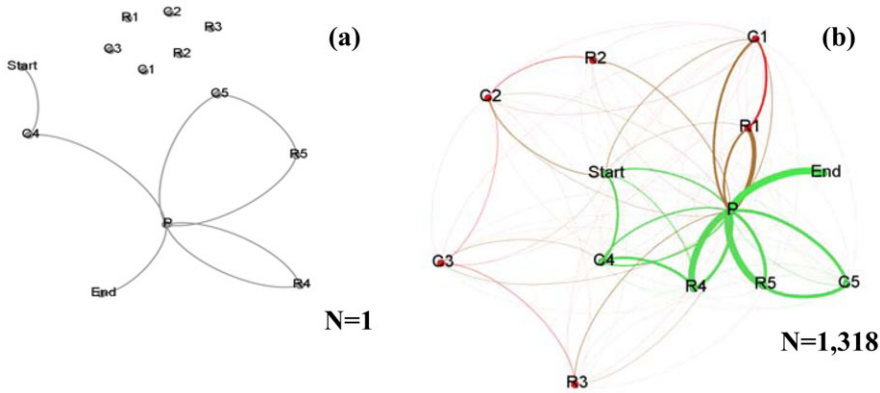

Figure 4. Visualization of transition networks from the Wells task: (a) one student with action sequence C4,P,P,R4,P,C5,R5,P and (b) transition network from a sample of 1,318 students.

categories. In comparison, systematicity scores are more equally distributed across the three categories.

Network Visualizations and Insights on Item Design

We first show results of visualizing the process data from the Wells task. As described above, each student has the option to conduct all or a subset of 11 actions to fix the pump. The log data record action sequences generated by student actions. One such example is C4,P,P,R4,P,C5,R5,P. In this action sequence, the student first checked for problem 4, and then checked the pump twice. Next, this student repaired problem 4, checked the pump again, and then checked for and repaired problem 5. Finally, this student checked the pump and ended this task. Using the method introduced above, we can create the corresponding transition network for this student's action sequence. The visualization of this network is shown in Figure 4(a). The visualization was generated using the open-source software program Gephi (Bastian, Heymann, & Jacomy, 2009). The links are curved clockwise to indicate the direction of these links. Because of limitations of the software, the self-loop for node P is shown in a very small size near to the node.

To get an overview of the whole data set, in Figure 4(b) we show the visualization of the aggregated transition network from all 1,318 students. Here, some links are thicker than others, indicating that more students took that transition than other transitions. To indicate student performance in terms of choosing correct actions, the nodes in the graph are colored such that green nodes represent necessary actions in solving the problem (correct actions) and red nodes represent unnecessary actions (incorrect actions). The color of the links carries the color of the beginning and end nodes of the link. In the graph, green links indicate the correct moves from one action to another, while red links indicate incorrect moves.

From this visualization alone, we can make the following basic observations that are closely related to item design. From the action node Start, C4 has the strongest

17453994, 2016, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12107 by Arizona State University Acq & Analysis, Lib Continuations, Wiley Online Library on [04/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2
Descriptive Network Statistics for All Transition Networks ($N = 1,318$)

	Mean	*SD*	Max	Min
Weighted density	.05	.02	.13	.03
Centralization	.44	.09	.76	.01
Reciprocity	.16	.19	.82	.00
003	218.22	19.22	248.00	156.00
012	49.19	20.21	111.00	2.00
102	6.67	7.31	25.00	.00
021D	1.13	1.43	10.00	.00
021U	1.13	1.43	10.00	.00
021C	5.31	3.71	26.00	1.00
111D	1.45	1.78	12.00	.00
111U	1.45	1.78	12.00	.00
030C	.75	.76	4.00	.00
201	.69	1.93	36.00	.00

link, which is a correct and necessary action; however, the other seemingly equivalent action C5 has far less link strength from Start. One potential explanation is that the option for C4 is above C5 on the screen and students tend to click things that appear higher up on the screen for this particular exercise. However, this hypothesis is speculative, and may not bear out in other situations. Another observation is that all sequences end with action P, which can be seen by the exclusive and thick link to the node End. This also might be enforced by the system design. Using network visualizations, we can easily see and discover these interesting patterns without conducting further analysis. The above observations and speculations can be tested using experiments with different option layouts, and/or by surveying participating students on how they make their choices.

Network Statistics and Scoring

We made some observations by visually checking the transition networks, but more rigorous analysis can be done by statistically comparing network measures of the transition networks. For each of the 1,318 transition networks, we calculated the weighted density, centralization, reciprocity, and triad census. The mean, standard deviation, maximum, and minimum are reported in Table 2. Notice here that we show only 10 out of 16 triadic patterns from the triad census. The reason is that the remaining six patterns, including 030T, 120D, 120U, 120C, 210, and 300, were not observed in the current data set.

From the descriptive statistics, we can see that the weighted density is low, which is due to the design of the task that did not allow the students to freely move from one action to another. The centralization scores of the transition networks are high, with a maximum of .76, which indicates that some actions are more popular than others. Again, this result is mainly due to the task design. For instance, the Check

17453994, 2016, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12107 by Arizona State University Acq & Analysis, Lib Continuations, Wiley Online Library on [04/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 3
Descriptive Network Statistics for Different Efficiency Score Categories

	1		2		3		4		5	
	67		123		311		76		741	
Category N	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Weighted density	.09	.01	.08	.01	.06	.01	.06	.01	.04	.01
Centralization	.26	.14	.28	.14	.24	.08	.20	.07	.17	.05
Reciprocity	.09	.13	.21	.23	.22	.22	.13	.18	.14	.17
003	176.60	17.31	197.76	17.89	213.99	16.48	214.66	9.58	227.53	11.53
012	77.72	23.33	58.11	26.33	48.63	21.71	52.83	15.53	44.99	15.37
102	4.39	5.91	8.78	8.67	9.18	8.37	6.07	7.35	5.53	6.28
021D	3.15	2.99	1.85	2.09	1.34	1.47	1.17	1.00	.73	.69
021U	3.15	2.99	1.85	2.09	1.34	1.47	1.17	1.00	.73	.69
021C	13.61	5.40	8.42	4.76	5.59	3.37	6.30	2.73	3.82	1.56
111D	2.63	3.18	3.04	2.85	1.97	1.74	1.30	1.49	.88	.97
111U	2.63	3.18	3.04	2.85	1.97	1.74	1.30	1.49	.88	.97
030C	1.04	1.25	.85	1.01	.87	.89	.61	.73	.68	.56
201	1.09	2.73	2.32	4.48	1.10	1.79	.58	1.44	.23	.60

Pump action (P) is one of the most common actions and thus one of the most visited nodes in the network. In comparison, C1 can be visited only once and is one of the least visited nodes. High variance in the node degrees results in a high centralization score. The average value on reciprocity is relatively low and shows that only 16% of the nonnull dyads are reciprocal, even though in our sample the maximum reciprocity is 82%. On triadic measures, 003 and 012 are the top two most observed structures, which are consistent with the observed low density. Even though network statistics are defined to measure different structures, they can potentially be highly correlated to each other. For instance, in our sample, the correlation between reciprocity and centralization is .63. In fact, reciprocity captures whether or not students tend to switch back and forth between two actions. In comparison, centralization captures the dispersion of the attention or attempts given to different actions. A lower centralization score indicates that the student took all actions about the same number of times; while a higher centralization score indicates that the student favored some actions more than others.

Using these network statistics, we want to study whether or not and in what way they are related to efficiency and systematicity. If one or more relationships exist, they can provide another way to score this task or similar tasks, which can be cross-checked with the existing scoring rubrics. We show the results for each of these two scores separately, each with related subcategory networks and then statistical tests on related network measures.

We first look at the efficiency scores. For the five categories, the mean and standard deviation of the three network measures are reported in Table 3. As examples, we report detailed analysis results on the first three network measures. The

17453984, 2016, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12107 by Arizona State University Acq & Analysis, Lib Continuations, Wiley Online Library on [04/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

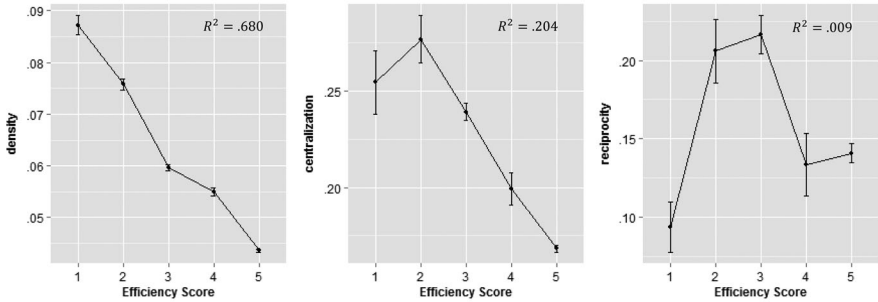

Figure 5. Plot of means of density, centralization, and reciprocity for each efficiency score category.

analysis of variance revealed a significant difference for the five categories for all three network measures with $F\ (4,\ 1,313) =\ 763.30,\ p < .05,\ \eta^2 =\ .70$ for weighted density, $F\ (4,\ 1,313) =\ 91.83,\ p < .05,\ \eta^2 =\ .22$ for centralization, and $F(4,\ 1,313)\ =\ 13.75,\ p < .05,\ \eta^2 =\ .04$ for reciprocity. Post-hoc comparisons using the Tukey HSD test (Miller, 1981) indicate that the mean scores for weighted density for all categories are significantly different from each other, and the differences between increasing score categories are along the same direction. Some pairs of categories for centralization and reciprocity measures are significantly different from each other. We plot the mean with standard error bars in Figure 5. In summary, we conclude that the weighted density captures what the efficiency score measures and might be a good indicator of students' efficiency scores. Looking at the definition of weighted density, this makes intuitive sense. If students are taking a lot of steps to solve the problem, they must conduct more actions, which results in more links in the network. On the other hand, centralization, which measures how students conduct different actions at different frequencies, or reciprocity, which measures how students revisit certain actions, are also related to efficiency but may not be good linear predictors. It is worth noticing that reciprocity is low in both low ends and high for students with medium efficiency scores. One potential explanation is that students with low efficiency scores conducted more actions and may have fewer clues on what might be the correct actions. This kind of exploration pattern results in relatively low reciprocity. On the high end, students with higher efficiency scores conducted mostly necessary actions, and may also have better ideas on the correct actions without testing, which also result in relatively low reciprocity. In comparison, students with medium efficiency scores might be deciding between correct and incorrect ones, which results in higher reciprocity scores. Corresponding observations are shown in the visualizations of the transition networks for students with different efficiency scores in Figure 6.

To explore the predicting power of the network statistics, we conduct a stepwise discriminant analysis (Klecka, 1980) with student efficiency score as dependent variable and all network statistics as independent variables. The overall Chi-square test was significant (Wilks $\lambda = .16$, Chi-square = 2,428.62, $df = 32$, $p <. 001$); the four functions extracted accounted for nearly 84% of between group variability,

17453994, 2016, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12107 by Arizona State University Acq & Analysis, Lib Continuations, Wiley Online Library on [04/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

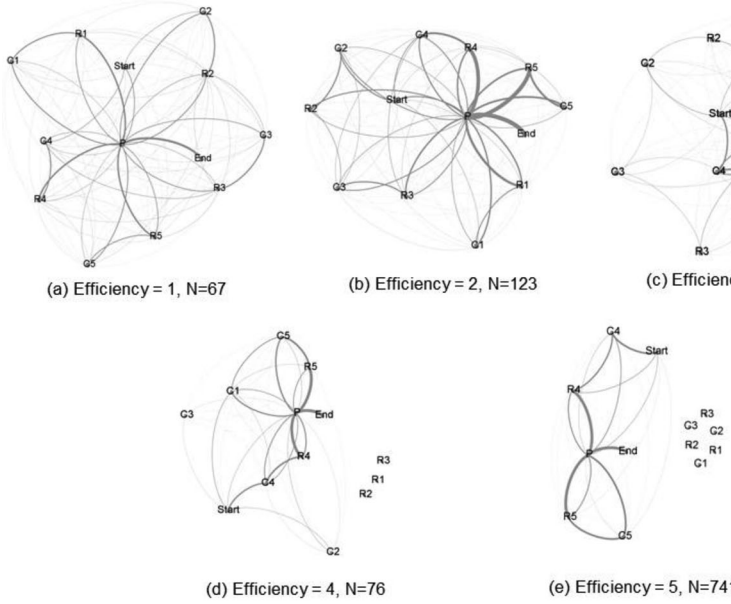

Figure 6. Transition networks for students with different efficiency scores.

Table 4
Standardized Canonical Discriminant Function Coefficients for Efficiency Score

	Function			
	1	2	3	4
Weighted density	−1.86	−1.43	−3.04	−.76
Reciprocity	−1.00	−.18	−2.61	2.40
Centralization	1.32	1.27	4.31	.26
012	.88	3.06	3.64	1.38
102	.62	2.87	2.93	−.14
021C	−.35	−.15	1.34	.68
030C	.40	1.00	−1.46	.53
201	.32	.42	.45	−.83

confirming the predicting power of the network statistics. However, not all network statistics are selected for the final model. Table 4 presents the standardized discriminant function coefficients for all included variables. All dropped variables are measure of triad patterns, including 003, 021D, 021C, 111D, and 111U. As illustrated in Figure 2, the last four structures all capture the transitions of two links in in the triads. For instance, 021D is the case when two actions nodes are both connected to a third node by down-pointing links. The exclusion of these triad measures from the discriminant function is consistent with the rubric of the efficiency score, which only concerns adopted actions and not the specific sequence of conducting these actions.

17453994, 2016, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12107 by Arizona State University Acq & Analysis, Lib Continuations, Wiley Online Library on [04/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 5
Descriptive Network Statistics for Different Systematicity Score Categories

	1 401		2 296		3 621	
Category N	Mean	*SD*	Mean	*SD*	Mean	*SD*
Weighted density	.05	.02	.06	.02	.05	.01
Centralization	.23	.10	.20	.10	.18	.07
Reciprocity	.32	.18	.16	.18	.06	.11
003	229.79	20.13	212.52	24.21	213.48	10.96
012	33.29	18.79	53.29	24.07	57.50	11.13
102	12.36	6.60	6.93	7.09	2.88	5.10
021D	.44	.94	1.14	1.65	1.57	1.40
021U	.44	.94	1.14	1.65	1.57	1.40
021C	3.46	3.31	6.40	4.85	5.99	2.78
111D	2.27	1.71	1.65	1.98	.83	1.47
111U	2.27	1.71	1.65	1.98	.83	1.47
030C	.24	.51	.58	.73	1.17	.68
201	1.44	2.77	.73	1.88	.19	.82

Reclassification of cases based on the new canonical variables shows that overall 78.5% of the cases were correctly classified. The reclassification of cases was conducted using the default SPSS procedure, the leave-one-out cross-validation. In this process, one observation was taken out as the validation set and the rest as training set. This was repeated for each observation.

Next, we look at how network measures are related to the systematicity scores. For the three categories, the mean and standard deviation of the network measures are reported in Table 5. The analysis of variance also revealed a significant difference for the three categories for all network measures. We report the details for the first three network measures with $F\ (2,\ 1, 315) = 18.80,\ p < .05,\ \eta^2 = .03$ for weighted density, $F(2,\ 1, 315) = 35.35,\ p < .05,\ \eta^2 = .05$ for centralization, and $F\ (2,\ 1, 315) = 346.90,\ p < .05,\ \eta^2 = .35$ for reciprocity. However, post-hoc comparisons using the Tukey HSD test indicate that the mean scores for weighted density for all categories are significantly different from each other, but run in different directions when systematicity scores increase. For the centralization and reciprocity measures, there are significant differences, and the means decrease with the increase of systematicity scores. We plot the mean with standard error bars in Figure 7. In comparison with the results for efficiency scores, weighted density is not well aligned linearly with the systematicity scores, but centralization and reciprocity are well aligned. Again, to interpret these findings, we revisit how these network measures are defined and how the scoring rubrics are defined. Generally speaking, the systematicity score seeks to measure the C, R, P patterns in the action sequence; the more students following this pattern, the better. Consequently, students with high systematicity scores will be less likely to revisit the immediate previous steps and

17453984, 2016, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12107 by Arizona State University Acq & Analysis, Lib Continuations, Wiley Online Library on [04/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

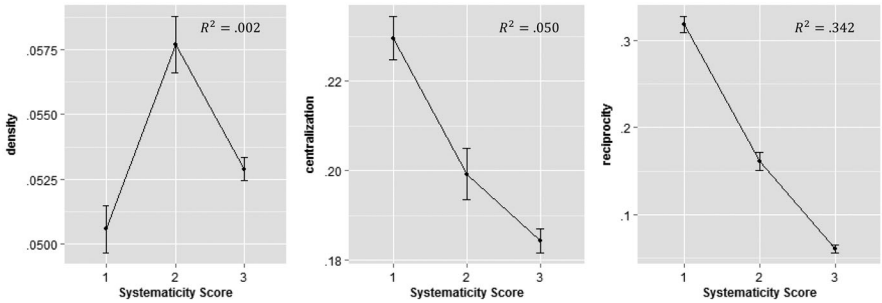

Figure 7. Plot of means of density, centralization, and reciprocity for each systematicity score category.

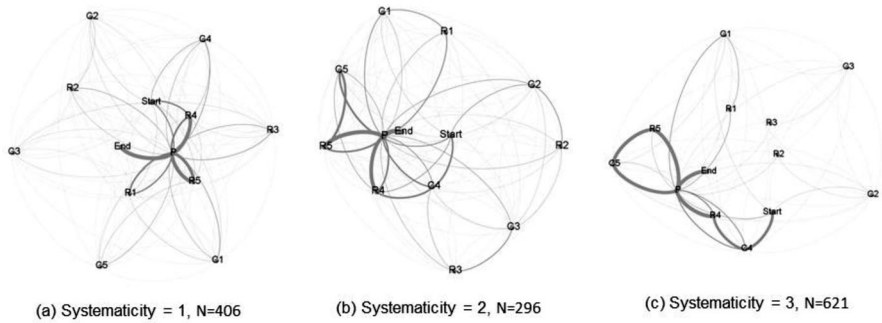

Figure 8. Transition networks for students with different systematicity scores.

thus have low reciprocity scores. In this specific case, since P is the only action that can be conducted more than once, students with low systematicity scores might be visiting P after many of their actions and result in a higher in-degree/out-degree for the action node P, and higher centralization. And students with high systematicity scores follow the C, R, P pattern and thus have more balanced visits to all actions, which result in lower centralization for the transition networks. So, centralization and reciprocity might be good indicators for systematicity. The visualizations of the transitions networks for students with different systematicity scores are in Figure 8. The observations as discussed above are shown in the visualizations.

To explore the predicting power of the network statistics on systematicity score, we conduct a stepwise discriminant analysis again with student systematicity score as a dependent variable and all network statistics as independent variables. The overall Chi-square test was significant (Wilks $\lambda = .46$, Chi-square = 1,021.96, $df = 12$, $p < .001$); the two functions extracted accounted for nearly 54% of between group variability, confirming the predicting power of the network statistics. However, not all network statistics are selected for the final model. Table 6 presents the standardized discriminant function coefficients for all included variables. Among the three network measures we discussed above, only reciprocity was included in the final model, together with five triad census measures. A closer check of the included network measures shows that they are all related to transition patterns in the network,

17453994, 2016, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12107 by Arizona State University Acq & Analysis, Lib Continuations, Wiley Online Library on [04/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 6
Standardized Canonical Discriminant Function Coefficients for Systematicity Scores

	Function	
	1	2
Reciprocity	1.80	.46
003	3.89	.84
021D	−6.07	−.55
021C	7.57	.12
111D	.41	−.24
030C	2.98	1.10

except triad measure 003, which counts the number of isolated triads. Comparing with the discriminant function for the efficiency score, the current function includes more network measures on how students transit from one action to another, especially the ones with links connecting more than two nodes. This is also consistent with the rubric for systematicity, which measures the specific patterns of C, R, P in the sequence. Reclassification of cases based on the new canonical variables was quite successful: overall 71.5% of the cases were correctly classified.

Conclusions and Discussion

In this article, we developed a novel way of using weighted directed networks and related network measures to represent, visualize, and analyze process data collected from educational assessments. We constructed transition networks by modeling student actions extracted from process data as nodes and transitions among the actions as links. We showed that transition networks can represent process data from one or a group of students and that visualization of these networks may assist discovering meaningful patterns in the data. We also introduced related analysis and statistics on the network measures, including global measures, such as weighted density and centralization, and local measures, such as reciprocity and triad census. Using a real data set from a scenario-based task, we showcased the application of the network methods and related analysis.

As process data become available from various assessment or learning environments, we believe that the network approach has a great potential in analyzing such data. Transition networks constructed from process data can represent and visualize the problem-solving processes from one and multiple students. The visualizations make it easy to discover meaningful and interesting patterns from the process data (e.g., Figure 4), and to compare patterns from different groups of students (e.g., Figures 6 and 8), which can be useful at least at the preliminary stage. These findings may provide useful feedback for item design and/or scoring. In addition, related network statistics provide more rigorous measures, which can be used in several different ways. For example, as showcased in this study, some network measures are related to existing scoring rubrics and have some predicting power on student scores.

17453984, 2016, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12107 by Arizona State University Acq & Analysis, Lib Continuations, Wiley Online Library on [04/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Furthermore, since these network measures capture local and global patterns in process data, they capture new evidence from the data and may be used in scoring, such as constructing or validating the rubrics. The network approach also has its limitations. For example, even though transition networks preserve most of the sequential information in the action series, they do not capture all information. For instance, if there are multiple loops in a network going back to the same node, then it will be difficult to tell which loop happened first from just the information preserved in the network. Fortunately, in many cases, as in this current study, this missing information does not impact the analysis results. If it becomes an issue in the future, it can also be addressed by adding more action/status nodes into the transition network to eliminate the ambiguity.

There are also alternatives to network approach. For instance, the n-gram method (Brown, Desouza, Mercer, Pietra, & Lai, 1992) from linguistics can be used to extract adjacent substrings from action sequences. Some of the network measures, such as reciprocity, may be calculated using a combination of these substrings. The process data can also be represented and analyzed using Markov models (e.g. Shu et al., 2014) or Petri net (Howard et al., 2010). As a useful addition to the available tools in analyzing process data, network approach has its own advantages. On the one hand, the visualization and analysis of the transition networks are straightforward and provide useful visual cues on the patterns in the action sequences, which also require relatively less processing compared with Markov models or Petri net. On the other hand, the definitions of nodes and links in the transition network are flexible and generalizable. As introduced in this article, one way is to use the nodes to represent actions and links to represent adjacent transitions. It is also possible to group nodes of similar attributes together to study action transitions at an aggregated level. Links are then beyond direct transitions.

The network approach introduced in this article can be extended in several ways. For example, besides measures and social network generative models, the development of new techniques called exponential random graph models (ERGMs) or p* models (Anderson, Wasserman, & Crouch, 1999; Holland & Leinhardt, 1981; Strauss & Ikeda, 1990) and related network modeling techniques (Snijders, 2011) can be particularly useful in analyzing these network data. ERGMs and other network models offer a method to study the interconnected data and do hypotheses testing the network patterns to find out representative structural patterns for studied networks. In this article, we introduced only two global measures and two categories of local pattern measures, but there are many other network measures (Newman, 2003, 2004; Wasserman & Faust, 1994) that go beyond this list and can be potentially useful in analyzing transition networks generated from process data. Some might be useful when different assessment tasks are studied or when analyzing data scored using some scoring rubrics. A final potential extension of this study is to connect this method with another network analysis technique called *epistemic network analysis* (Shaffer et al., 2009). In this method, instead of focusing on actions as in this current study, the focus is put on the connections among student's developing skills, knowledge, identity, values, and epistemology, as well the dynamic changes of these connections over time. Network measures introduced in this article can be useful in analyzing epistemic networks as well. The idea of longitudinal analysis from

17453984, 2016, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12107 by Arizona State University Acq & Analysis, Lib Continuations, Wiley Online Library on [04/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

epistemic network studies can also be introduced to analyze transition networks from process data if students are observed repeatedly over time.

Appendix: Scoring Rules for the Wells Task

In this section, we provide the detailed scoring rules for the Wells task.

Scoring rules for efficiency. The student should perform troubleshooting and repair actions in an efficient manner. List of student actions:

C1 - Handle moves too easily; no water - Check to see if handle has become disconnected from piston rod
R1 - Handle moves too easily; no water - Repair: reconnect handle to piston rod
C2 - Water has a bad smell - Check for buildup of debris within the chamber
R2 - Water has a bad smell - Repair: flush chamber with water until debris is removed
C3 - Water is very cloudy - Check for buildup of debris within the chamber
R3 - Water is very cloudy - Repair: flush chamber with water until debris is removed
C4 - It is hard to move handle; pump noisy-Check for dirt or rust on external moving parts
R4 - It is hard to move handle; pump noisy—Repair: clean and lubricate moving parts
C5 - No water comes out - Check for tears or cracks in the piston seal
R5 - No water comes out - Repair: Replace piston seal
P - Try the pump

The following are general guidelines used to define *efficiency* as used in the scoring rules.

* The pump is exhibiting problems 4 and 5 (addressed by C4, R4, C5, and R5 in list above). Students should not perform any check or repair actions related to problems that are not exhibited.
* Performing an unnecessary repair is penalized more than performing an unnecessary check, as this is a more inefficient procedure.

Efficient actions - E $= \{C4, R4, C5, R5\}$
Unnecessary checks - C $= \{C1, C2, C3\}$
Unnecessary repairs - R $= \{R1, R2, R3\}$
5 - Only actions from set E
4 - Actions from E + 1 action from C
3A - Actions from E + 2−3 actions from C
3B - Actions from E + 0−1 action from C + 1 action from R
2 - Actions from E + 2−3 actions from C + 1−2 actions from R
1 - Actions from E + 3 actions from C + 3 actions from R

Scoring rules for systematicity. The student should perform troubleshooting and repair actions in a methodical, systematic fashion. List of student actions:

C1 - Handle moves too easily; no water - Check to see if handle has become disconnected from piston rod
R1 - Handle moves too easily; no water - Repair: reconnect handle to piston rod

17453984, 2016, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12107 by Arizona State University Acq & Analysis, Lib Continuations, Wiley Online Library on [04/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

C2 - Water has a bad smell - Check for buildup of debris within the chamber
R2 - Water has a bad smell - Repair: flush chamber with water until debris is removed
C3 - Water is very cloudy - Check for buildup of debris within the chamber
R3 - Water is very cloudy - Repair: flush chamber with water until debris is removed
C4 - It is hard to move handle; pump noisy - Check for dirt or rust on external moving parts
R4 - It is hard to move handle; pump noisy - Repair: clean and lubricate moving parts
C5 - No water comes out - Check for tears or cracks in the piston seal
R5 - No water comes out - Repair: Replace piston seal
P - Try the pump

The following are general guidelines used to define sequences used in the scoring rules.

* Students should not perform a repair before checking to verify that the repair they are performing will address the symptom they are attempting to address with the repair.
* Once a student has performed a repair, she should check to see if the problem is solved by trying out the pump (action P). Any additional Ps are irrelevant and will be ignored for scoring purposes.
* Students who use a very inefficient procedure for repairing the pump will receive a low score for systematicity, as students who are performing a lot of unnecessary steps may be following a systematic procedure unrelated to troubleshooting/repair (e.g., pushing all buttons on the interface is in some sense "systematic" but does not provide meaningful evidence of troubleshooting/repair skill).

3 - All checks performed before repairs; pump is checked immediately following each repair.
2a - All checks performed before repairs; pump is not checked immediately following each repair.
2b - One repair is performed before the associated check (or check is omitted); pump may not be checked immediately following each repair.
1 - Two or more repairs performed before the associated check; pump may not be checked immediately following each repair.

References

Almond, R., Deane, P., Quinlan, T., Wagner, M., & Sydorenko, T. (2012). *A preliminary analysis of keystroke log data from a timed writing task.* ETS Research Report RR-12-23. doi:10.1002/j.2333-8504.2012.tb02305.x. Retrieved May 27, 2014, from http://www.ets.org/Media/Research/pdf/RR-12-23.pdf

Anderson, C. J., Wasserman, S., & Crouch, B. (1999). A p* primer: Logit models for social networks. *Social Networks*, *21*, 37–66.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*, *8*, 361–362.

Bergner, Y., Shu, Z., & von Davier, A. A. (2014). Visualization and confirmatory clustering of sequence data from a simulation-based assessment task. In *Proceedings of the*

17453984, 2016, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12107 by Arizona State University Acq & Analysis, Lib Continuations, Wiley Online Library on [04/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

7th International Conference of Educational Data Mining (pp. 177–184). Available at educationaldatamining.org/EDM2014/index.php?page=proceedings

Bondy, J. A., & Murty, U. S. R. (2008). *Graph theory*. New York, NY: Springer.

Brown, P. F., Desouza, P. V., Mercer, R. L., Pietra, V. J. D., & Lai, J. C. (1992). Class-based n-gram models of natural language. *Computational Linguistics*, *18*, 467–479.

Davis, J. A., & Leinhardt, S. (1972). The structure of positive interpersonal relations in small groups. In J. Berger (Ed.), *Sociological theories in progress* (Vol. *2*, pp. 218–251). Boston, MA: Houghton Mifflin.

DiCerbo, K. E., & Behrens, J. T. (2012). Implications of the digital ocean on current and future assessment. In R. Lissitz & H. Jiao (Eds.), *Computers and their impact on state assessment: Recent history and predictions for the future* (pp. 273–306). Charlotte, NC: Information Age Publishing.

DiCerbo, K. E., & Behrens, J. T. (2014). *Impacts of the digital ocean on education*. London, UK: Pearson.

DiCerbo, K. E., Liu, J., Rutstein, D. W., Choi, Y., & Behrens, J. T. (2011, April). *Visual analysis of sequential log data from complex performance assessments*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Freeman, L. (1979). Centrality in social networks I: Conceptual clarification. *Social Networks*, *1*, 215–239.

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, *21*, 1129–1164. doi:10.1002/spe.4380211102

Hao, J., Shu, Z., Bergner, Y., Zhu, M., & von Davier, A. A. (2014, July). *Assessing students' performances from process data in scenario-based tasks: An edit distance approach*. Paper presented at the 79th annual meeting of the Psychometric Society, Madison, WI, USA.

Holland, P. W., & Leinhardt, S. (1970). A method for detecting structure in sociometric data. *American Journal of Sociology*, *76*, 492–513.

Holland, P. W., & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs (with discussion). *Journal of the American Statistical Association*, *76*, 33–65.

Howard, L., Johnson, J., & Neitzel, C. (2010). Examining learner control in a structured inquiry cycle using process mining. In *Proceedings of the 3rd International Conference on Educational Data Mining*, 71–80. Retrieved May 27, 2014, from http://educationaldatamining.org/EDM2010/uploads/proc/edm2010_submission_28.pdf

Institute of Education Sciences. (2013). *Technology and engineering literacy assessment*. Washington, DC: U.S. Department of Education. Retrieved May 27, 2014, from http://nces.ed.gov/nationsreportcard/pdf/about/schools/grade8_tel_factsheet.pdf.\

Kerr, D., Chung, G., & Iseli, M. (2011). *The feasibility of using cluster analysis to examine log data from educational video games*. CRESST Report No. 790. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Klecka, W. R. (1980). *Discriminant analysis*. Beverly Hills, CA: Sage Publications.

Miller, R. G. (1981). *Simultaneous statistical inference* (2nd ed.). New York, NY: Springer.

Mislevy, R. J., Oranje, A., Bauer, M. I., von Davier, A. A., Hao, J., Corrigan, S., . . . , John, M. (2014). *Psychometric considerations in game-based assessment*. GlassLab Research White Paper. Princeton, NJ: Educational Testing Service.

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, *45*, 167–256. doi:10.1137/S003614450342480

Newman, M. E. J. (2004). Analysis of weighted networks. *Physical Review E*, *70*, 056131. doi:10.1103/PhysRevE.70.056131

17453984, 2016, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12107 by Arizona State University Acq & Analysis, Lib Continuations, Wiley Online Library on [04/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., … Mislevy, R. J. (2009). Epistemic network analysis: A prototype for 21st-century assessment of learning. *International Journal of Learning and Media*, *1*, 33–53. doi:10.1162/ijlm.2009.0013

Shu, Z., Zhu, M., Hao, J., Bergner, Y., & von Davier, A. A. (2014, July). *Using Markov-IRT to characterize process data*. Paper presented at the 79th annual meeting of the Psychometric Society, Madison, WI, USA.

Snijders, T. A. B. (2011). Statistical models for social networks. *Annual Review of Sociology*, *37*, 131–153. doi:10.1146/annurev.soc.012809.102709

Strauss, D., & Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, *85*, 204–212.

Tai, R. H., Loehr, J. F., & Brigham, F. J. (2006). An exploration of the use of eye-gaze tracking to study problem-solving on standardized science assessments. *International Journal of Research and Method in Education*, *29*, 185–208. doi:10.1080/17437270600891614

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

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature*, *393*, 440–442.

Authors

MENGXIAO ZHU is a Research Scientist, Computational Psychometrics Research Center, Educational Testing Service, 660 Rosedale Road, MS-02T, Princeton, NJ, 08541; mzhu@ets.org. Her primary research interests include psychometric models for collaborative problem solving, social network analysis and data mining techniques applied on assessment data, and integration of cognitive science with psychometrics.

ZHAN SHU is a Psychometrician, Educational Testing Service, 660 Rosedale Road, MS-02T, Princeton, NJ, 08541; zshu@ets.org. His primary research interests include characterizing students' behaviors through statistical modeling and/or educational data mining.

ALINA A. VON DAVIER is a Senior Research Director, Computational Psychometrics Research Center, Educational Testing Service, 660 Rosedale Road, MS-12T, Princeton, NJ, 08541; avondavier@ets.org. Her primary research interests include test equating, statistical quality control tools for assessment data, time series, and dynamic models.

17453984, 2016, 2, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/jedm.12107 by Arizona State University Acq & Analysis, Lib Continuations, Wiley Online Library on [04/11/2022]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License