Markov Analysis of Students' Professional Skills in Virtual Internships

Vasile Rus, Dipesh Gautam, [1]Dale Bowman, Arthur C. Graesser, [2]David Shaeffer

Department of Computer Science and Institute for Intelligent Systems, The University of Memphis, Memphis, TN, 38152

[1]Department of Mathematics, The University of Memphis, Memphis, TN, 38152

[2]Department of Educational Psychology, University of Wisconsin-Madison, Madison, WI, 53706

vrus@memphis.edu

Abstract

In this paper, we conduct a Markov analysis of learners' professional skill development based on their conversations in virtual internships, an emerging category of learning systems characterized by the epistemic frame theory. This theory claims that professionals develop epistemic frames, or the network of skills, knowledge, identity, values, and epistemology (SKIVE) that are unique to that profession. Our goal here is to model individual students' development of epistemic frames as Markov processes and infer the stationary distribution of this process, i.e. of the SKIVE elements. Our analysis of a dataset from the engineering virtual internship Nephrotex showed that domain specific SKIVE elements have higher probability. Furthermore, while comparing the SKIVE stationary distributions of pairs of individual students and display the results as heat maps, we can identify students that play leadership or coordinator roles.

Introduction

In virtual internships students play the role of interns in a virtual training environment meant to simulate real internship experiences. The learning that occurs in virtual internships can be characterized by epistemic frame theory. This theory claims that professionals develop epistemic frames, or the network of skills, knowledge, identity, values, and epistemology (SKIVE elements) that are unique to that profession (Chesler et al. 2010). For example, engineers share ways of understanding and doing (knowledge and skills); beliefs about which problems are worth investigating (values), characteristics that define them as members of the profession (identity), and ways of justifying decisions (epistemology).

In this study, we propose a new method to characterize learners' professional skill development in virtual internships in terms of SKIVE elements distributions. The basic idea is to use a Markov process approach to infer the stationary distribution of SKIVE elements based on an analysis of interns'/learners' conversations with other players, e.g. a mentor or intern, in engineering virtual internships. Specifically, we analyze interns' online conversations during the design process, a key activity in the engineering virtual internships such as Nephrotex (NTX) in which students research and create multiple engineering designs (Bagley and Shaffer 2009).

Following prior work, elements of the engineering epistemic frame are operational zed as discourse codes in order to detect when students activate such SKIVE elements during conversations. An example of the identification of SKIVE elements as discourse codes in virtual internship conversations is shown in Table 1 where the student utterance encodes a reference to design ***S****kills*.

Table 1.An example of an utterance and SKIVE codes.

Utterance	S	K	I	V	E
Let me know if you have any questions about their requirements for membrane design.	1	0	0	0	0

While an empirical distribution could be derived by computing the relative proportion of each activated SKIVE elements during conversations, our goal is to infer the true or stationary distribution of SKIVE elements for each student by modeling students' epistemic frames as Markov processes. The stationary distribution is the true distribution of SKIVE elements that would be observed if the student would talk forever (or an extremely long period of time). We designed Markov processes for SKIVE elements in virtual internships as briefly explained next. Markov processes are characterized by a set of states, which in our case are the SKIVE elements, and a set of transition probabilities, which we derive from analyzing the activation of SKIVE elements during the virtual internship conversations. For instance, we consider transitions from SKIVE elements activated by a student in prior dialogue utterances to the SKIVE elements activated by the student in the current utterance. More precisely, we will use a moving window to delimit the number of previous dialogue utterances

to consider when deriving the transitions. The size of the prior context moving window (in terms of number of utterances) can be set by the experimenter, as we will explain later. The larger the window the more likely we will identify transitions between various SKIVE elements, therefore, reducing data sparseness issues. On the other hand, a larger previous context window, i.e. one that includes many previous utterances, will account for long-distance transitions, i.e. between SKIVE elements activated in utterances that are far apart, which may be less relevant.

Given the above design, we experimented with several methods of deriving Markov processes to infer SKIVE elements' stationary distributions. For instance, we differentiated between methods that consider utterances from a single player versus all utterances (of all players). Also, we varied the way we derive the state-to-state transition counts, which are used to compute the transition probabilities, from a source state to a destination state: transitions between SKIVE elements/states in any utterance in the moving window will make the same contributions to the final transition count versus a penalizing model in which transitions from SKIVE elements/states farther away in the prior dialogue context are contributing less, e.g. there is a discounting parameter. We also compared models with and without a dummy SKIVE element/state (noSKIVE state) used to characterize utterances in which no SKIVE element is present (i.e., the student is not mentioning any discourse code indicative of a SKIVE element).

Once we inferred the SKIVE epistemic frame in terms of a stationary distribution for each student/intern, we compared students' SKIVE epistemic frames against each other and also against an average epistemic frame distribution obtained by computing an average of the stationary distributions of students' epistemic frames. We compare the epistemic frame distributions using Kullback-Leibler (KL) divergence.

Our work has a merit in the sense that it provides a more rigorous way of describing students' emergence/mastery of SKIVE elements in terms of stationary/true distributions as opposed to empirically derived distributions.

In the next sections, we discuss related work, Markov Chain theory and its use in virtual internship, the proposed conversation models, the engineering virtual internship Nephrotex datasets we used, and experiments and results. The paper ends with Conclusions and Future Work.

Related Works

The epistemic network analysis (ENA) framework was proposed as a way to characterize learning during internship when young apprentices are beginning their professional career by interacting with seasoned professionals (Bagley and Shaffer 2009; Shaffer et al. 2009). ENA is grounded in epistemic frame hypothesis (Shaffer 2006)according to which professionals develop epistemic frames or the network of *skills, knowledge, identity, values, and epistemology*(SKIVE) that are unique to that profession. The network or interconnections between concepts enable the process of assessment of learning progression in context. The ENA framework offers evidence-centered design that provides evidence of learning by systematically linking models of understanding, observable actions, and evaluation rubrics.

Rupp and colleagues (2009) described a method to represent students' epistemic frames using ENA. In their method, the sequence of activities in the Urban Science, a virtual internship game, is divided into time slices and each slice is coded based on whether the slice activates one or more of the SKIVE codes. They then constructed an adjacency matrix of these codes for each slice based on whether any two of the codes co-occur in the slice. A cumulative adjacency matrix is also derived for a player/intern or the mentor from the whole sequence of activities (an accumulation of all activated SKIVE codes over all time slices). They used different statistics such as overall weighted density, absolute and relative centrality and rescaled cumulative association to build a network structure of SKIVE elements based on the adjacency matrices. Students' mastery of SKIVE elements was measured by distance metrics among different players under comparison.

In another work, Bodin (2012) analyzed university physics students' epistemic frames while working on a task in which they were supposed to simulate a particle-spring model system. Students' epistemic frames were analyzed before and after the task using a network analysis approach derived from an analysis of interview transcripts. They found that students change their epistemic frames when switching from a modeling task to a physics task.

Zhu and Zhang (2016) did a pilot investigation to understand the patterns of the communication and connections of engineering professional skills (EPSs). For the pattern of communication, they identified who was talking to whom at different points of time and identified whether one or more EPSs co-occurred in an utterance. They applied social network analysis to the resulting co-occurrence network and found that the high-performing group tend to show denser and more balanced network connections in both the communication and EPS networks.

In our work, we used Markov process theory to characterize students' development of SKIVE epistemic frames in terms of stationary distributions, as described next. That is, we consider students' development of SKIVE elements as a Markov process in which there is a state corresponding to each SKIVE element and the transitions among those SKIVE elements are observed during dialogues. Based on the transitions, we can derive the stationary distribution of an interns' SKIVE elements, which can be regarded as a

reflection of that students' master of their target profession's *skills, knowledge, identity, values, and epistemology*.

Markov Process

A Markov process is a random process characterized by a set of states and transition probabilities among these states. The probability of the Markov process being in a particular state only depends on the previous state. The dependency of the current state only on the previous state, or a limited history of previous states, is called the Markov property of a Markov process. The Markov property is expressed formally using the following equation:

$$P(X_{n+1}|X_1X_2,\dots,X_n) = P(X_{n+1}|X_n)$$

Where $X_1, X_2,\dots, X_n$ is a sequence of random variables.

The transition probability matrix of a Markov process is of the form shown in equation 1, where rows indicates the source state and columns indicate target/destination states. A particular element, e.g. p_{13}, indicates the probability of making a transition from source state, say 1, to a destination state, say, 3.

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} \quad (1)$$

The transition probabilities can be used to predict the probability of being in a particular state after a number of transitions when starting from a particular state. For instance, the probability of being in state j after 2transitions/steps when starting from state i is shown in equation 3 where superscript two (2) indicates the number of steps.

$$p_{ij}^{(2)} = p_{i1}p_{1j} + p_{i2}p_{2j} + \cdots + p_{in}p_{nj} \quad (2)$$

Furthermore, equation 2 is simply the dot product of the i^{th} row and j^{th} column vectors of the transition matrix. Therefore, we can predict the future state of the Markov process by obtaining the power of the transition matrix. The probability that a Markov process reaching state j from state i after transiting to k-*1*states in between is given by $(i,j)^{th}$ element of the matrix P^k=P*P*…*P (P multiplied k times).

Importantly for our work, the convergence theorem of Markov processes indicates that when k tends to infinity the matrix P^k attains a stationary state in which all rows are equal (Tierney 1994).After reaching convergence, the probability of reaching a state is constant irrespective of the state from which the system had started. In our case, we rely on this convergence theorem to derive the stationary distribution of students' SKIVE elements based on transition probabilities derived from conversations during virtual internships, as explained next.

Markov Processes for Epistemic Frames

Besides content knowledge, students need to master their target profession's *skills, knowledge, identity, values, and epistemology* (SKIVE or epistemic frame elements).We propose here a novel way to monitor and assess students' mastery of the SKIVE elements in terms of stationary distributions of the states of a SKIVE Markov process in which there is a state for each of the SKIVE elements. We rely on students' activation of SKIVE elements during their conversations with other players in the virtual internship to characterize the underlying Markov process and infer the stationary SKIVE distribution for each intern.

For this purpose, every utterance of a conversation is being annotated with binary codes indicating whether a particular SKIVE element is present or absent in the utterance. That is, whether the student activated the corresponding SKIVE elements during his conversational moves.

We model each SKIVE element as a state of an underlying Markov process. Because a student can activate multiple SKIVE elements in the same utterance, another option is to consider as a state a combination of SKIVE elements that students may articulate in any given utterance. We opted for the one-SKIVE-element per Markov process state option because we are interested in characterizing students in terms of SKIVE elements distributions as opposed to combinations of such elements. Secondly, considering all possible combinations SKIVE elements increases the complexity of the Markov process by increasing exponentially the number of states to all possible subsets of SKIVE elements, i.e. to 2^n states where n is the number of SKIVE elements. This will lead to two problems: (i) a data sparseness problem when deriving the transition probabilities and (ii) difficulty with interpreting the outcome.

We derive transition probabilities for each student's Markov process from the sequence of SKIVE elements identified in student's utterances during virtual internship conversations. That is, SKIVE elements activated in previous utterances are considered source states and SKIVE elements activated in the current utterance are considered target states of a state transition. We count each such transition from a source SKIVE state to a target SKIVE state and then normalize the values across all transitions with the same source state to infer the transition probabilities.

There are three important aspects of the way in which we derive the transition probabilities. First, instead of using the full previous dialogue context to detect source states

for the transitions, we use a limited dialogue history in the form of a moving window of *k* previous utterances relative to the current utterance inspired from previous work (Rupp et al. 2009; Rus et al 2014). The larger the window the more likely it is that we will identify transitions between various SKIVE elements, therefore, reducing data sparseness issues. On the other hand, a larger previous context window, i.e. one that includes many previous utterances, will account for long-distance transitions, i.e. between SKIVE elements activated in utterances that are far apart, which may be less relevant.

Second, when analyzing conversations to count the number of transitions from one SKIVE element to another, one can treat transitions from utterances close to each other the same way as transitions from utterances far apart in which case both such types of transitions will contribute an equal count of 1 to the final count. An alternative is to give less weight to transitions derived from utterances far apart. We present results with both these weighting methods.

A third key aspect with respect to deriving the transition probabilities is what utterances in a conversation to consider: the utterances of the student being analyzed or all utterances of all the players. We present results with both models: student-utterances vs. all-utterances.

A formal description of our conversation models and the transition probabilities are derived is presented next.

Conversation Model

A conversation is a sequence of utterances $u_1 ... u_T$ where an utterance u_k is coded with a set of binary codes corresponding to each of the SKIVE elements $C_{k1} ... C_{kN}$. We also define a set of weights $w_1 ... w_N$ corresponding to an utterance u_k such that a weight w_j is given by:

$$w_j = \begin{cases} \sum_{p=k-n}^{p=k-1} \alpha^{p-k+1} C_{pj} & \text{if previous utterance window is weighted} \\ 1 & \text{otherwise if } C_{pj} = 1 \text{ for some } p \text{ and previous utterance window is unweighted} \\ 0 & \text{otherwise if } C_{pj} = 0 \text{ for all } p \text{ and previous utterance window is unweighted} \end{cases}$$

where α is a decay factor (set to 2, as explained later)that penalizes the contribution of farther utterances in the window/slice corresponding to the current utterance. C_{pj} is the j^{th} SKIVE element of previous utterance *p* within the moving window of size *n*.

The weighted count transition matrix M of dimension *N*x*N* is obtained by the algorithm listed in Table 2. The above conversation model allows us to derive the transition probabilities among SKIVE elements by constructing an adjacency, binary or weighted, matrix.

Table 2.Algorithm to obtain weighted count matrix

Initialize: M = a zero matrix
do for each utterance u_i:
  U = a NxN zero matrix
  do for each code C_j of u_i:
    $u_{ij} = w_j * C_j$, where u_{ij} is an element of U
  M = M + U

The adjacency matrix for a given student is basically constructed by scanning her conversation utterances and counting the number of times the student activates SKIVE element B in the current utterance while activating SKIVE element A in one of the *k* previous utterances included in our moving window of size k. The adjacency matrix thus obtained cumulatively counts the number of transitions between any pair of SKIVE elements. The result is a state transition matrix whose entries are raw cumulative count of transitions from one SKIVE element to another.

The final state transition probability matrix is obtained by dividing each entry in the adjacency matrix by the sum of the elements in the corresponding row. In order to deal with sparse matrices, which have many zero entries, we do Laplace's add-one smoothing (Lidstone, 1920).

Adjacency Matrix with single player window

In this model, an adjacency matrix is created for each player by considering only utterances of that player. For example in Table 3, for any player pl_i, the utterance for the player is the set of all utterances where pl_i is marked (*x*).

Table 3.Table of utterance and players in a conversation.

	pl_1	**pl_2**	**...**	**pl_n**
utt_1	x			
utt_2		x		
...			…	
utt_k		x		

Adjacency Matrix with multi player window

In this model, an adjacency matrix is created for each player by taking into consideration utterances spoken by all the players in the conversation. In this case, in Table 3, for any player pl_i, the utterances for the player is the set of all the utterances utt_1 through utt_k

The Nephrotex Dataset

We experimented with data from the virtual internships Nephrotex in which groups of students work together on a design problem, e.g. designing filtration membranes for hemodialysis machines, with the help of a mentor.

In the Nephrotex dataset, there are 25 players divided into five groups. Each group is assigned a virtual room to work together on a task.

Table 4. Excerpt of a conversation in Nephrotex (only few SKIVE codes shown).

SN	Player	Content	s.design	s.professional	s.collaboration	s.data
1	4	Let me know if you have any questions about their requirements for membrane design.	1	0	0	0
2	4	At Nephrotex, we have internal consultants who are experts in their fields.	0	0	0	0
3	18	When they say carbon nanotubes, to which surfactant are they referring to?	0	0	0	0
4	16	I believe it's any of them. It's just using the carbon nanotubes in general.	0	0	0	0

Once the task is finished, the players are assigned to other groups and a group is again assigned a room to work on a task. In total, there were 10 unique groups and 19 unique rooms formed. The dataset consists of a total of 2,970 utterances with an average of 37 utterances per room.

Table 4 is an excerpt of a conversation in Nephrotex. The utterances are coded with 20 SKIVE elements. While analyzing the dataset, we found that some of the utterances do not contain any SKIVE elements, hence a row attributed to that utterance has zero counts across all SKIVE elements, i.e. columns in Table 4. We handled such scenario following two different approaches. In a first approach, we discarded all the utterances with all-zero counts. In another approach, we introduced a dummy state, called no-SKIVE state, which indicates a state when no SKIVE element was activated by a student in an utterance.

Experiments and Results

We experimented with a combination of single-player vs multi-player models, weighted vs. non-weighted counts/binary counts, and Markov processes with or without no-SKIVE state. For the weighted window models, we selected a decay factor of $\alpha=2$thus penalizing by a factor of 2 transition counts from previous utterances for each one unit increase in distance from the current utterance.

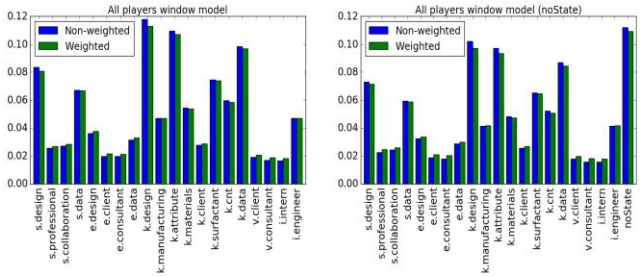

Figure 1. Distribution of probabilities of SKIVE elements for all-player conversation model with non-weighted and weighted window and distributions when noState added.

For each model we ran the Markov process iteratively until it converged, i.e. reaching the stationary state. Figure 1 shows the average stationary distribution of SKIVE elements when the utterances from all the players are considered while Figure 2 shows the average stationary distribution derived using only utterances of one player.

Because the Nephrotex dataset is an engineering design internship, engineering specific components such as design, data, manufacturing, attributes, materials and engineers have higher probabilities compared to others. Also adding a no-SKIVE state in the analysis, shifted a good chunk of probability to the no-SKIVE state. This is the case because a significant part of the conversation consist of short dialogue turns in which the speakers use elliptical responses, in which much is implied from the context, or they focus on general conversation and process topics, e.g. greeting each other or asking about how to use the system.

When comparing the weighted window model to its non-weighted counterpart, the distributions are similar as confirmed by distance measures (KL-divergence) between the

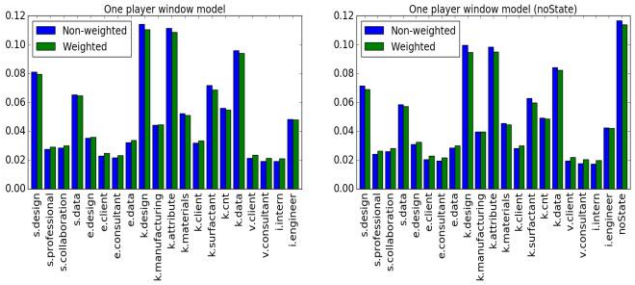

Figure 2. Distribution probabilities of SKIVE elements for single player conversation model with non-weighted and weighted window (top left and top right) and distributions when noState added(bottom left and bottom right)

corresponding distributions of SKIVE elements of the weighted and non-weighted models (KL=0.00031 for all player without no-SKIVE state; KL=0.00035 for one player without no-SKIVE state). However, one can notice that the probability distribution of the least frequent SKIVE elements are boosted. This is in a way a desired effect because the most frequent elements, if not penalized, tend to dominate by the simple fact that they occur in more utterances throughout a conversation and therefore are more

likely to be present in a moving window which in turn leads to increased transition counts. The weighted model penalizes frequent components when occurring in remote utterances relative to the current utterance.

Once the stationary distributions of SKIVE elements were obtained, we conducted an analysis of students' SKIVE profiles by computing KL-divergence (KL) scores between pairs of distributions of SKIVE elements for individual students. A summary of the KL scores is shown as a heat map in Figure 3. Student players are sorted based on their average KL score with other players. The left vertical color bar in the maps show the intensity of the user's average KL score in sorted order.

It can be seen that some players have similar distributions, e.g. those shown in the lower left corner of the heat

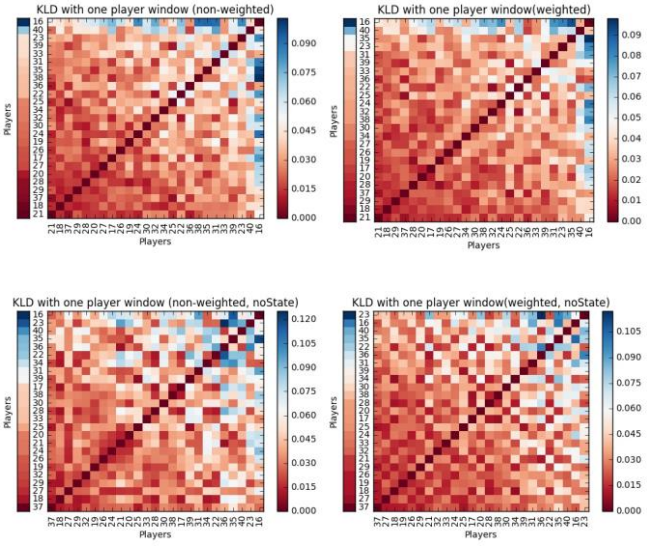

Figure 3. KL-divergence of distribution SKIVE elements of players for window with single player utterances (top two are for weighted and non-weighted and bottom two are for weighted and non-weighted with noState)

maps in Figure 3. The lower left corner corresponds to lower divergence scores. When using a no-SKIVE state and weights, the distributions of SKIVE elements between players seem to be more similar as shown in the heat maps on the lower right hand side of Figure 3. Furthermore, adding a no-SKIVE state (bottom left and bottom right) revealed that some of the players move from an upper position, corresponding to a higher average divergence score, to lower-divergence positions in the heat map. Those that move to lower-divergence positions are more likely to have utterances in which no SKIVE elements are activated. They may correspond to students playing more leadership/coordinator roles as their utterances focus more on process and conversational management topics and less on SKIVE elements. We only show results for single-player utterances models due to space reasons.

Conclusion and Future Work

We conducted a Markov process analysis of students' mastery of epistemic frames which is generally applicable to any epistemic frame. We have experimented and validated our method on data from an engineering epistemic frame using eight different ways to model Markov processes for each student participating in engineering virtual internships. The comparison of the distribution of SKIVE elements for individual students in models with noState revealed that some students may play more managerial or coordinator roles than others.

In future work, we plan to use the stationary distribution of SKIVE components obtained from this analysis to better understand students' effectiveness of acquiring much needed skills to be successful professionally. Furthermore, we plan to develop a dual Markov process to infer stationary distributions of states and transitions.

Acknowledgments. This work was funded in part by the National Science Foundation (NSF; DRL-1418288). The opinions, findings, and conclusions do not reflect the views of the NSF and cooperating institutions or individuals.

References

Bagley, E., & Shaffer, D. W. (2009). When people get in the way: Promoting civic thinking through epistemic game play. *Int. Journal of Gaming and Computer-mediated Simulations*, 1, 36-52.

Bodin, M. (2012). Mapping university students' epistemic framing of computational physics using network analysis. *Physical Review Special Topics-Physics Education Research*, 8(1).

Chesler, N. C., Bagley, E., Breckenfeld, E., West, D., & Shaffer, D. W. (2010, June). A virtual hemodialyzer design project for first-year engineers: An epistemic game approach. In *ASME 2010 Summer Bioengineering Conference* (pp. 585-586). American Society of Mechanical Engineers.

Rupp, A. A., Choi, Y., Gushta, M., Mislevy, R., Bagley, E., Nash, P., ... & Shaffer, D. (2009). Modeling learning progressions in epistemic games with epistemic network analysis: Principles for data analysis and generation. In *Proceedings from the Learning Progressions in Science Conference* (pp. 24-26).

Rus, V., Stefanescu, D., Niraula, N., & Graesser, A. C. (2014, March). Deeptutor: Towards macro-and micro-adaptive conversational intelligent tutoring at scale. In *Proceedings of the first ACM conference on Learning@Scale* (pp. 209-210). ACM.

Shaffer, D. W. (2006). Epistemic frames for epistemic games. *Computers & education*, 46(3), 223-234.

Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., ... & Mislevy, R. (2009). Epistemic network analysis: A prototype for 21st-century assessment of learning.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, 1701-1728.

Zhu, M., & Zhang, M. (2016, March). Examining the patterns of communication and connections among engineering professional skills in group discussion: A network analysis approach. *In 2016 IEEE Integrated STEM Education Conference (ISEC)* (pp. 181-188). IEEE