Simplification of Epistemic Networks Using Parsimonious Removal with Interpretive Alignment

Yeyu Wang[1[0000–0003–1978–5453]], Zachari Swiecki[1[0000-0002-7414-5507]], Andrew R. Ruis[1[0000-0003-1382-4677]] and David Williamson Shaffer[1[0000-0001-9613-5740]]

[1] University of Wisconsin – Madison, Madison, WI, USA
ywang2466@wisc.edu

Abstract. A key goal of quantitative ethnographic (QE) models, and statistical models more generally, is to produce the most *parsimonious* model that adequately explains or predicts the phenomenon of interest. In epistemic network analysis (ENA), for example, this entails constructing network models with the fewest number of codes whose interaction structure provides sufficient explanatory power in a given context. Unlike most statistical models, however, modification of ENA models can affect not only the statistical properties but also the *interpretive alignment* between quantitative features and qualitative meaning that is a central goal in QE analyses. In this study, we propose a novel method, Parsimonious Removal with Interpretive Alignment, for systematically identifying more parsimonious ENA models that are likely to maintain interpretive alignment with an existing model. To test the efficacy of the method, we implemented it on a well-studied dataset for which there is a published, validated ENA model, and we show that the method successfully identifies reduced models likely to maintain explanatory power and interpretive alignment.

Keywords: Model Comparison, Model Refinement, Unified Methods, Epistemic Network Analysis (ENA), Interpretive Alignment.

1 Introduction

Quantitative ethnography (QE) is a method for studying cultural and behavioral patterns that facilitates thick description of qualitative data at scale [8]. To do this, QE unifies qualitative and quantitative approaches with the goal of achieving *interpretive alignment* between qualitative meaning-making and the features of quantitative models. For example, *epistemic network analysis* (ENA), a widely used QE technique, models the structure of connections among key concepts, behaviors, or other elements (i.e., Codes) to represent complex phenomena [9]. Critically, ENA maintains linkages between model features (i.e., weighted connections between Codes) and the original qualitative data that produced them, enabling researchers to warrant interpretive alignment.

However, a model that achieves interpretive alignment might not be the most *parsimonious* one. In statistics, model parsimony is often operationalized as the inclusion of the fewest number of variables that sufficiently explain or predict the phenomenon of interest [7]. The variables in an ENA model are the connections among Codes, but because QE unifies qualitative and quantitative perspectives, the most parsimonious ENA model is not necessarily the model with the fewest Codes. Rather, parsimony in ENA involves constructing a model with the fewest Codes that maintains both explanatory power and *interpretive alignment*.

To identify the most parsimonious ENA model while maintaining interpretive alignment, QE researchers typically take a trial-and-error approach. However, this process involves iterative confirmation across qualitative and quantitative perspectives, and there is no reliable method for determining whether two ENA models with different Codes have equivalent explanatory power.

To address this challenge in QE model refinement, we propose a novel method, *parsimonious removal with interpretive alignment* (PRIA), for removing Codes to identify more parsimonious models likely to maintain interpretive alignment. We tested this method on a well-studied dataset for which there are published QE findings, and our results suggest that the PRIA method can reliably find the most parsimonious ENA model that maintains explanatory power. While researchers still need to verify interpretive alignment, PRIA provides a principled method for testing the effects of removing one or more Codes from an existing model.

2 Theory

2.1 Properties of QE Models

Quantitative ethnography unifies ethnographic and computational methods to understand culture, behavior, cognition, and other aspects of human activity at scale. Unlike mixed method studies, in which the qualitative and quantitative analyses are only minimally interdependent, QE studies generate thick descriptions of big data using processes that “inseparably” combine qualitative and quantitative approaches [8].

For example, one frequently-used modeling technique in QE is *epistemic network analysis* (ENA). ENA models the structure of connections among Codes by quantifying the co-occurrence of Codes within recent temporal context. ENA analyses begin when researchers perform a close reading of some discourse data to generate theories about the situated meaning of events, which are then operationalized as Codes. ENA constructs network graphs of the coded data that can be used to explore hypotheses about the relationships among them, as well as summary statistics that can be used to compare the relationships among Codes statistically. Then researchers *close the interpretive loop* [8], validating the features of the model by re-examining the original data that contributed to those features.

In particular, ENA supports the unification of qualitative and quantitative methods in at least five ways. ENA models:

1. Represent the connections among Codes that were developed and validated by researchers grounded in the empirical data.

2. Generate an *ENA score* in a projected metric space. ENA score is defined as a summative measure for an individual network that can be used to conduct statistical tests.
3. Position nodes in the ENA space such that network graphs can be used to interpret the meaning of ENA scores in terms of the network structures they represent.
4. Provide a *goodness of fit* measure that indicates how well the projected metric space and weighted network models are coordinated. High goodness of fit indicates that the network graphs provide a reliable interpretation of the dimensions of the projected metric space.
5. Preserve the *interpretive alignment* between quantitative features (statistical significance between groups and interpretation of the dimensions based on node positions) and a qualitative understanding of the data.

For example, [10] explored collaborative problem-solving in a military training exercise on air defense warfare, using qualitative analysis on the discourse data and ENA to model the patterns in discourse. They compared the behavior of commanders with and without access to a tactical scaffolding system, which provided detailed information about *tracks* (ships and aircraft detected on radar) and a record of actions taken toward them. [10] found that commanders without access to the scaffolding system focused more on seeking information about behaviors of incoming aircraft to understand the tactical situation; commanders using the scaffolding system integrated information about the tactical situation to issue deterrent orders. The ENA space they created showed that network nodes for information seeking were on the negative side of the first dimension of the ENA space, and commanders working without the scaffolding system had low values on that dimension. Network nodes for tactical decision making were on the positive side of the first dimension, and commanders working with the scaffolding system had high values on that dimension. There was a statistically significant difference between commanders in the control and experimental conditions on the first dimension of the ENA space. Thus, the interpretation derived from the quantified ENA model was consistent with the situated hypothesis and Code generation based on thick description of the discourse data.

2.2 Model Parsimony

A common challenge that researchers face when using ENA is deciding whether one or more codes can be removed from a model without affecting interpretive alignment. For example, researchers may construct a model in which one or more Codes—or more specifically, the connections involving one or more Codes—do not provide significant explanatory or discriminative power. That is, the model refinement process may involve exploring whether a more parsimonious model can be constructed that does not change the interpretation of the data, but that tells the story more clearly by removing Codes that do not play a significant role in the model. However, the unification of qualitative and quantitative methods in QE poses a particular challenge for refining QE models in this way.

In statistics, *parsimony*—the idea that a model is better when fewer variables have the same explanatory power—is one of the most important factors in model selection and evaluation. As [11] argue, the primary benefit of a more parsimonious model is that it increases the predictive power by better distinguishing the signal from noise. Parsimony balances variance explained and model complexity, and reduces the likelihood of overfitting, or over-parametrization in model selection [6].

However, [2] examined studies of qualitative parsimony and failed to find a universal definition due to the complexity and diversity of social science research. Although they claimed that "one 'size' of parsimony does not appear to fit all qualitative methods" (p. 1403), parsimony involves the simplest expressions of qualitative findings grounded in thick description.

The problem for QE researchers is that unlike pure statistical models, which can be reduced to the most parsimonious form by systematically removing variables until the amount of variance explained drops too low [7], modification of QE models can affect not only the *statistical* properties of a model but also the *interpretive alignment* between quantitative features and qualitative meaning that is a central goal in QE analyses.

From a QE perspective, then, parsimony needs to unify two points of view. From a quantitative perspective, a parsimonious model needs to involve the fewest parameters or variables to achieve the same level of explanatory power. Connections between Codes are the variables in an ENA model, so the more parsimonious ENA model thus has fewer Codes (and hence, fewer connections). From an ethnographic perspective, QE researchers are concerned with capturing the "right" amount detail to explain the phenomenon at hand. Adding more Codes does not always provide a better or more clear interpretation of events, and in fact can obscure the most central features of a situation.

2.3 Current Approach to Identifying Parsimonious ENA Models

In practice, researchers using ENA take an exploratory approach to refining models, trying to achieve parsimony while maintaining explanatory power and interpretive alignment. Typically, researchers examine their existing "best" model to determine whether one or more Codes are contributing significantly to meaningful interpretation. Codes that are weakly connected, especially those located toward the center of the model, are a common target for removal. A researcher selects one or more Codes to remove, produces a deflated model, and then checks whether the deflated model maintains (a) high goodness of fit, (b) statistical significance, and (c) interpretive alignment. Needless to say, this process is time-consuming, and there is no extant technique for comparing two models with different Codes in a principled way.

2.4 The PRIA Approach to Identifying More Parsimonious ENA Models

To address this issue, we propose a method, Parsimonious Removal with Interpretive Alignment (PRIA), to identify deflated models that are likely to maintain explanatory power and interpretive alignment. This method has five steps:

Step 1: Generate deflated models. We construct a set of *deflated models* by generating all possible ENA models with fewer codes using the same parameters as the *original model*.
Step 2: Determine which deflated models have high goodness of fit. We conduct a goodness of fit test on all deflated models, which measures the extent to which the interpretation of dimensions is reliable.
Step 3: Correlate ENA scores. We compute the correlation of ENA scores between each deflated model and the original model. This correlation measures the extent to which the summary statistics for individual units of analysis in the deflated model vary from the original model.
Step 4. Correlate node positions. We compute the correlation of node positions between each deflated model and the original model. Node positions are used to interpret the meaning of ENA scores in terms of the network structures they represent. A high correlation of the node positions between the deflated model and the original model suggests that the underlying patterns described by the networks are aligned.
Step 5: Confirm interpretive alignment for *candidate models*. Deflated models that pass tests for goodness of fit and correlation of ENA scores and node positions become *candidate models*. We sort the set of candidate models by the number of codes removed, from most codes removed to least. Then, we check interpretive alignment one candidate model at a time until we find the most parsimonious model that has good interpretive alignment. To ensure that interpretive alignment is conserved, we examine each candidate model from two perspectives:

1. From the quantitative perspective, we re-run any statistical tests from the original model to determine whether differences remain significant. Additionally, we examine the network graphs to verify that the node positions preserve the interpretation of the dimensions.
2. From the qualitative perspective, we once again close the interpretive loop, re-examining the original data to see if it is aligned with the candidate model.

The candidate model with the most codes removed that meets these alignment criteria becomes the *reduced model*.

2.5 Data and Research Questions

We thus propose five steps for generating a reduced model and confirming interpretive alignment. To examine the feasibility of PRIA, we analyzed a dataset collected from the Tactical Decision Making Under Stress (TADMUS) project [1]. We chose this dataset because prior work has developed and validated a coding scheme and ENA models [10]. In what follows, we use this data to address the following research questions:

1. Can candidate models be constructed that have high goodness of fit, ENA score correlation, and node position correlation?
2. Do any of the candidate models preserve the statistical results of the original?
3. Do any of the candidate models preserve the interpretation of the dimensions of the original?

4. Are any of the candidate models that meet all of the above criteria also well aligned with the original data?

3 Method

3.1 Dataset and Codebook

In this study, we tested the Code removal procedure using data collected in the TADMUS project. The dataset includes the discourse of 16 Navy Air Defense Warfare (ADW) teams working collaboratively to make tactical decisions in a military training simulation [4]. In these training simulations, teams were tasked with identifying, tracking, and assessing potentially hostile air or surface vessels called tracks. Each team consisted of up to six members, two in command roles, and four in supporting roles. Those in command roles made tactical decisions based on the information reported by the supporting members of the team. A decision-support system was developed to scaffold the decision-making process of the commanders by providing detailed information about tracks and a record of actions taken toward them [3]. To test the effectiveness of the system, eight teams were randomly assigned to the experimental group, who had access to the support system, while the other eight teams were assigned to the control group, with access only to the regular watch station. The regular watch station is a device that provides basic tracking and identification information. In this study, we made use of the discourse data for each team in the training project for 94 participants. The transcripts of their discourse are segmented to 12,027 turns of talk in total.

Table 1. Codes, Definitions, and Examples Developed by [10]

Code	Definition	Examples
DETECT/IDENTIFY (DI)	Talk about radar detection of a track or the identification of a track, (e.g., vessel type).	IR/EW NEW BEARING, BEARING 078 APQ120 CORRELATES TRACK 7036 POSSIBLE F-4.
TRACK BEHAVIOR (TB)	Talk about kinematic data about a track or a track's location	AIR/IDS TRACK NUMBER 7021 DROP IN ALTITUDE TO 18 THOUSAND FEET
ASSESSMENT (A)	Talk about whether a track is friendly or hostile, the threat level of a track, or indicating tracks of interest.	TRACKS OF INTEREST 7013 LEVEL 5 7037 LEVEL 5 7007 LEVEL 4 TRACK 7020 LEVEL 5 AND 7036 LEVEL 5

STATUS UPDATES (SU)	Talk about procedural information, e.g., track responses to tactical actions, or talk about tactical actions taken by the team	TAO ID, STILL NO RESPONSE FROM TRACK 37, POSSIBLE PUMA HELO.
SEEKING INFORMATION (SI)	Asking questions regarding track behavior, identification, or status.	TAO CO, WE'VE UPGRADED THEM TO LEVEL 7 RIGHT?
RECOMMENDATIONS (R)	Recommending or requesting tactical actions	AIR/TIC RECOMMEND LEVEL THREE ON TRACK 7016 7022
DETERRENT ORDERS (DTO)	Giving orders meant to warn or deter tracks.	TIC AIR, CONDUCT LEVEL 2 WARNING ON 7037
DEFENSIVE ORDERS (DFO)	Giving orders to prepare ship defenses or engage hostile tracks	TAO/CO COVER 7016 WITH BIRDS

We adopted this dataset and coding scheme because the Codes were grounded in the qualitative understanding of researchers familiar with the data, and automated classifiers were developed validated by two expert human raters. To understand the differences in team discourse with and without the supporting system, [10] analyzed the transcripts and developed the eight behavior codes in Table 1. For each code, pairwise inter-rater reliability between two human raters and the automated classifier was Cohen's $\kappa > 0.84$ and Shaffer's $\rho(0.65) < 0.05$, indicating that all codes met a minimum threshold of $\kappa = 0.65$.

3.2 Epistemic Network Analysis

Using the same parameters reported in [10], ENA models were constructed with participants subdivided into training scenarios as the units, scenario as the conversation variable, and a moving window of 5 turns of talk, using rENA 0.2.0.1 [5]. We included all eight codes listed in Table 1. To create a projected ENA space, we used a means rotation to maximize the differences between control and experimental conditions, followed by a singular value decomposition, to define the first and second dimensions, respectively, of the projected metric space. On the other hand, we calculate the arithmetic means of the network edges, defined as *centroids*. To test the reliability of the interpretation on the means rotation dimension, we computed the goodness of fit using Pearson's r between the ENA scores with centroids for the same individual (r = 0.95).

3.3 Code removal Procedure and Criteria

As described above in the theory section, we constructed all possible reduced models by extracting potential combinations of code removal[1]. Then, we measured each one on three criteria: (a) its goodness of fit, (b) the correlation of its ENA scores with the original model, and (c) the correlation of its node positions with the original model (see Table 2). Then, we compared the computed correlations (Criteria 1-3 in Table 2) with a given threshold, 0.95. Deflated models significantly above the threshold on all three criteria were classed as candidate models, which were rank-ordered by number of codes, with the model having the least number of codes listed first.

Table 2. Goals and Criteria for Generating Deflated Models

Goals		Criteria
Model Fitting	The interpretation of dimensions is reliable.	Criterion 1: Goodness of Fit for deflated model
Model Alignment	The interpretation of individuals is aligned with the original model.	Criterion 2: Correlation of ENA scores between the original and deflated models
	The interpretation of dimensions is aligned with the original model.	Criterion 3: Correlation of node positions between the original and deflated models

To ensure that the interpretation of dimensions in a deflated model is reliable, we computed the goodness of fit by calculating the pairwise correlation (Pearson's r) of the ENA scores with their corresponding network centroids. We then calculated the 95% confidence interval (C.I.) and compared the lower bound of the C.I. with the threshold of 0.95 (Criterion 1), to maintain the same goodness of fit as the original model. If the lower bound is above the threshold, the fit of the deflated model is significantly higher than the threshold, which indicates that the deflated model is well fitted and reliable for interpretation. Of the deflated models that satisfied Criterion 1, we selected deflated models whose ENA scores and node positions are highly correlated with the original model (Criteria 2 and 3). High correlation of ENA scores indicates that the interpretation of the individuals in the model is consistent with the original model, and high correlation of node positions maintain the interpretation of the dimensions in the network. We calculated the C.I. of correlations for both ENA scores and node positions between each deflated model and original model and returned the deflated models whose lower bounds of the C.I. were above the threshold. The models that satisfied all criteria were considered candidate models, and we sorted them by the number of Codes removed.

[1] In this study, the original model includes eight Codes. We took combination of k Codes (k = 1, 2, …, 5) to be removed from the original model. The maximum value of k is 5 since ENA needs at least three Codes to be constructed.

After the PRIA algorithm output the candidate models, we confirmed the statistical results and interpretive alignment of the most parsimonious candidate model. To do this, we first determined whether there is a significant difference between the two conditions on the first dimension of the candidate model, using two-sample *t*-test. We then checked the interpretation of the first dimension based on the locations of the nodes to determine whether it was consistent with the original model. Lastly, we compared the transcripts of two teams in the same training scenario that had similar tactical decision-making processes and results, except that one team's discourse contained a removed Code and the other team's discourse did not. The first candidate model to achieve interpretive alignment was confirmed as the reduced model.

4 Results

4.1 Code Removal Results

Of the 218 possible deflated models, the PRIA algorithm returned 3 that satisfied Criteria 1-3 (see Table 3). The three candidate models in Table 3 all have a high goodness of fit and high correlations of ENA scores and node coordinates with the original model. We chose the candidate model with most Codes removed to assess first for interpretive alignment. The model includes five Codes: "Seeking Information", "Detect/Identify", "Track Behavior", "Status Updates", and "Deterrent Orders.

Table 3. Candidate models with correlations (confidence intervals) significantly higher than 0.95

Codes Removed	C1: Goodness of Fit	C2: Correlation of ENA Scores	C3: Correlation of Node Positions	Number of Codes Removed
Assessment, Defensive Order, Recommendation	**0.9562 (0.9556, 0.9567)**	**0.961 (0.953, 0.9677)**	**0.9986 (0.9785, 0.9999)**	**3**
Defensive Order, Recommendation	0.957 (0.9564, 0.9575)	0.9738 (0.9683, 0.9783)	0.9997 (0.9968, 1)	2
Defensive Order	0.9511 (0.9505, 0.9517)	0.9953 (0.9943, 0.9961)	0.9985 (0.9892, 0.9998)	1

4.2 Quantitative Confirmation

To compare the original model and the most parsimonious candidate model, we plotted both models (see Fig. 1). We examined whether the model interpretation changed in the following two respects: 1) statistical comparison of the two conditions; 2) network interpretation based on the node positions.

Statistical Testing between Conditions. In the original model, there is a significant difference between the control condition (Mean = –0.25, N = 211) and the experimental condition (Mean = 0.25, N = 211) according to a two-sample *t*-test ($t = 10.52$, $p < 0.005$, $d = 1.02$) on the first (means rotated) dimension. In the candidate model, there is a significant difference between the control condition (Mean = –0.27, N = 211) and the experimental condition (Mean = 0.27, N = 211) according to a two-sample *t*-test ($t = 9.36$, $p < 0.005$, $d = 0.91$). Thus, the most parsimonious candidate model maintains the significant difference between the two conditions.

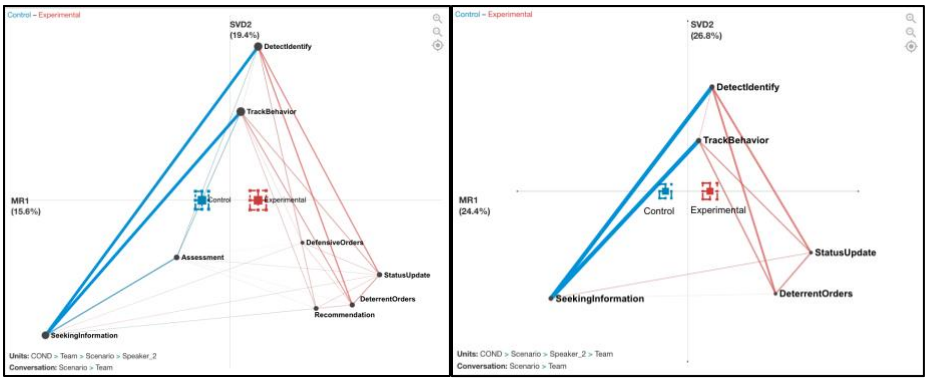

Fig. 1. The Original Model (Left) and the Most Parsimonious Candidate Model (Right)

Node Positions in ENA. In the original network, Codes related to tactical decision making (DETERRENT ORDERS and STATUS UPDATE) are located on the positive side of the first dimension, while SEEKING INFORMATION is located on the negative side. In the most parsimonious candidate model, the relative location of major Codes along the means rotation space are in relative the same positions as the original model.

4.3 Qualitative Confirmation

We identified two teams (hereafter, Team 1 and Team 2) working on the same training scenario with a similar distribution of ENA scores in the original model. The distribution of ENA scores were not significantly different according to a Mann-Whitney U Test ($U = 25$, $p = 0.31$, Cliff's $d = 0.39$) and suggest that the two teams made similar connections between codes. However, Team 2 made connections between ASSESSMENT and other codes, but Team 1 did not. The removal of ASSESSMENT in addition to DEFENSIVE ORDER and RECOMMENDATION is what differentiated the most parsimonious candidate model from the next most parsimonious one. In the examples below, we look at each team at the beginning of a training session involving the same potential attack.

Example 1: Decision Making without Explicit Assessment of Threats. At the beginning of the task, Team 1 is told by the SSES that there is an Iranian F-4 in their airspace ("IN VICINITY") on a training mission ("FOR APPARENT LOCAL TRAINING"):

Line	Speaker	Utterance
1	SSES	TAO/SSES WE HAVE INDICATIONS OF AN IRANIAN F-4 AIRCRAFT UP IN VICINTIY BUSHEHR AIRBORNE FOR APPARENT LOCAL TRAINING.

That is, the SSES points out a plane in the area (DETECT/IDENTIFY) and describes its behavior (TRACK BEHAVIOR).

Almost immediately, the team sees radar contacts for another fighter plane:

Line	Speaker	Utterance
2	EWS	TIC/EW I HAVE AN APS-115 CORRELATES TO P-3 BEARING 025.
3	TAO	TRACK 7023 BEARING 025 40M.
4	TIC	EW/TIC WE COPY WE GOT THAT CORRELATED TRACK 7023.
5	IDS	THIS IS IDS, 7023 NO MODES NO CODES.

In line 2, the Electronic Warfare Supervisor (EWS) identifies it as a P-3 (DETECT/IDENTIFY) and gives its bearing (TRACK BEHAVIOR). In line 4, the TIC designates as "TRACK 7023" (DETECT/IDENTIFY), and in line 5, the IDS reports "NO MODES NO CODES" (DETECT/IDENTIFY), indicating that the plane is not responding to automatic tracking. In other words, the plane is not identifying itself, and thus could be hostile.

The team subsequently identifies three other radar tracks, and then almost immediately (line 14) the EWS acquires the Iranian F-4 on radar (DETECT/IDENTIFY and TRACK BEHAVIOR). The TIC marks it as "TRACK 7036" in line 15 and gives its updated bearing (DETECT/IDENTIFY and TRACK BEHAVIOR):

Line	Speaker	Utterance
14	EWS	TIC/EWS APQ-120 BEARING 077 CORRELATES F-1 MIRAGE CORRECTION F-4 PHANTOM.
15	TIC	F-4 CORRELATES TO TRACK 7036 BEARING 078 RANGE 50NM.

Then, the SSES then notifies the team that there are multiple aircraft departing from the Iranian coast on the way toward the ship (DETECT/IDENTIFY and TRACK BEHAVIOR).

Line	Speaker	Utterance

16	SSES	TAO/SSES WE HAVE INDICATIONS OF MULTIPLE AIRCRAFT DEPARTING IRANIAN COAST, EXACT COMPOSITION UNKNOWN, LOCATION UNKNOWN.

In other words, the team has taken a series of actions to DETECT/IDENTIFY tracks and designate track behavior. Collectively, these actions suggest there are hostile aircraft preparing to attack the ship, even though the Iranian F-4 (now TRACK 7036) was supposedly on a training mission:

1. There are multiple planes in the airspace,
2. At least one of them is not identifying itself (TRACK 7023 has "NO MODES NO CODES in line 5), and
3. There are multiple airplanes departing from Iran toward the ship (line 8).

In response to this tactical situation, the TAO issues a level 1 warning (DETERRENT ORDER, giving orders to warn or deter tracks) for the F-4 (TRACK 7036), which is "FEET WET", or flying from land to the water and approaching the ship:

Line	Speaker	Utterance
17	TAO	AWC/TAO TRACK 7036 GOING FEET WET ISSUE LEVEL 1 WARNING PLEASE.

Team 1 thus identified a series of tracks and recorded their behavior, leading the TAO to (a) conclude that the tracks were hostile and (b) issue a deterrent order as a "LEVEL 1 WARNING". However, this process of deciding to issue a deterrent order does not show explicit ASSESSMENT of the threat level of the planes.

Example 2: Decision Making with Explicit Assessment of Threats. The second team, working on the same training scenario, receives the same notification from SSES at the beginning of the exercise. The SESS detects an "IRANIAN F-4 AIRCRAFT" (DETECT/IDENTIFY) is "IN VICINITY" (TRACK BEHAVIOR):

Line	Speaker	Utterance
1	SSES	TAO/SSES WE HAVE INDICATIONS OF AN IRANIAN F-4 AIRCRAFT UP IN VICINTIY BUSHEHR AIRBORNE FOR APPARENT LOCAL TRAINING.

The TIC acknowledges the intelligence provided by the SSES and, using that information, asks the AWC to investigate a track that might be the F-4:

Line	Speaker	Utterance
2	TIC	TIC AYE.
3	TIC	AWC/TIC TRACK 7020 WE NEED TO PROBABLY INTERROGATE THEM AND SEE IF WE GET ANY EW, THAT MIGHT BE THE F-4. HE'S AT 3,000FT, UH ACTUALLY I

		TAKE THAT BACK IT'S PROBABLY A HELO AT 80KNTS 3,000FT THERE ABOUT 40M.

In line 3, the TIC initially identifies “TRACK 7020” as the Iranian F-4 (DETECT/IDENTIFY), but quickly changes the identification to a helicopter (“IT’S PROBABLY A HELO”) based on the TRACK BEHAVIOR, including altitude (“3,000FT”), distance (“40M”), and speed (“80KNT”).

Next, the IDS identifies a new track, TRACK 7023:

Line	Speaker	Utterance
4	IDS	TIC/IDC TRACK 7023 NO MODES NO CODES.
5	TIC	TAO/TIC 7023 SPEEDING UP AND TURNING WEST. SHE IS ALSO CLIMBING. SHE IS AT 1800FT RIGHT NOW AND TURNING SOUTH ON 239 250KNTS.
6	IDS	TIC/IDC TRACK 7023 UPDATED UNKNOWN ASSUMED HOSTILE.

The IDS notes (line 4) that the new track has no radar information indicating whether it is friendly or hostile (“NO MODES NO CODES”). Immediately after this DETECT/IDENTIFY from the IDS, the TIC (line 5) reports that the track is exhibiting unusual TRACK BEHAVIOR, including changing speed (“SPEEDING UP”), direction (“TURNING WEST”), and height (“CLIMBING”). Using that information, the IDS makes an explicit ASSESSMENT: TRACK 7023 is “ASSUMED HOSTILE” (line 6).

Soon after, the SSES identifies (line 10) multiple aircraft taking off from the Iranian coast “FEET WET” (DETECT/IDENTIFY and TRACK BEHAVIOR):

Line	Speaker	Utterance
10	SSES	TAO/SSES WE HAVE INDICATIONS OF MULTIPLE AIRCRAFT DEPARTING IRANIAN COAST, EXACT COMPOSITION UNKNOWN, LOCATION UNKNOWN.

Moments later, given the information from the SSES, the TAO issues “LEVEL 1 WARNINGS” to Track 7036 (DETERRENT ORDER):

Line	Speaker	Utterance
15	TAO	AWC/TAO ISSUE LEVEL 1 WARNINGS TO TRACK 7036.

Thus Team 2, like Team 1, identifies a series of tracks and records their behavior, leading the TAO to conclude that the tracks are hostile and issue a DETERRENT ORDER—the same level 1 warning that the TAO from Team 1 ordered.

Along the way, in Team 2 the IDS makes an explicit ASSESSMENT of the tactical situation (line 6). However, this assessment does not appear to have influenced the Team’s tactical decision making or behavior: in both teams, the TAO reaches the same conclusion. In other words, although Team 2 makes an explicit ASSESSMENT of the situation, it appears that the assessment served only to make explicit the interpretations the team was already making—which, in turn, provides qualitative evidence that removing the code from the model does not significantly change the model of team behavior or our interpretation of it.

5 Discussion

In this paper, we proposed the PRIA method and successfully reduced an existing ENA model to a more parsimonious model with fewer Codes but equivalent explanatory power and interpretive alignment. We propose this method based on five properties of ENA that support the unification of qualitative and quantitative approaches. Based on each property, we generated corresponding criteria and confirmatory procedures.

1. ENA models generate the ENA scores, a summative measure for individuals, projected as points in the ENA space. PRIA correlates the ENA scores between the original model and the reduced model, to ensure the interpretation of individuals in the space is maintained.
2. ENA models provide an interpretation of the dimensions of the ENA space based on the node positions. PRIA correlates the original node position between the original and the reduced, to ensure the interpretation of dimension is maintained.
3. ENA models generate a goodness of fit measure to test how well the projected metric space and the weighted network models are coordinated—and thus whether the interpretation of the dimension is reliable. PRIA computes the goodness of fit test on the reduced model to ensure the qualified candidate models are also reliable.

Based on these properties and criteria, PRIA deflates the original model and generates the candidate models with Codes removed that maintain both good model fit and alignment with the original model.

PRIA then has two additional confirmatory procedures that validate the interpretive alignment between quantitative metrics (statistical tests and node positions) and qualitative analyses.

4. ENA performs the statistical tests on the ENA scores and enables interpretation of those results based on node positions. PRIA performs the same tests on the reduced model to confirm significance remained and checks that the node positions in the reduced model do not result in a different interpretation from the original model.
5. ENA closes the interpretive loop by enabling researchers to confirm that the results of a model conform to the original qualitative data. PRIA re-closes that loop to ensure that the reduced model is consistent with the qualitative data as well.

We tested the PRIA method on a data from a military training project that already had validated Codes and an existing ENA model aligned with qualitative analyses. For this data, PRIA suggested it was possible to construct a reduced model with three Codes removed that met all of these criteria.

The PRIA method has some limitations. First, *this approach requires an existing ENA model, with validated Codes and good interpretive alignment*, to generate and evaluate more parsimonious candidate models. That is, the PRIA method provides a principled way for researchers to *refine an existing validated model*, rather than providing a technique to automatically generate an optimal model from a large set of Codes. Second, we only tested the PRIA method on one dataset. To test the generalizability of PRIA method, we will need to conduct the same procedure on other datasets with different properties. Finally, we chose an arbitrary threshold of 0.95 for our correlation criteria. Methods to determine appropriate minimum thresholds need to be developed.

Despite these limitations, the PRIA method shows that it is possible to generate parsimonious reduced models that have good interpretive alignment and explanatory power equivalent to the original model. This suggests that the PRIA approach can provide a principled method for identifying more parsimonious ENA models.

Acknowledgments. This work was funded in part by the National Science Foundation (DRL-1661036, DRL-1713110), the Wisconsin Alumni Research Foundation, and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

References

1. Cannon-Bowers, J.A., Salas, E.E.: Making Decisions Under Stress: Implications for individual and team training. Acta Psychologica. 103, 337–341 (1999).
2. Cutcliffe, J.R., Harder, H.G.: The perpetual search for parsimony: Enhancing the epistemological and practical utility of qualitative research findings. International Journal of Nursing Studies. 46, 10, 1401–1410 (2009).
3. Hutchins, S.G. et al.: Principles For Aiding Complex Military Decision Making. In: Proceedings of the second international command and control research and technolo-gy symposium. p. 22 (1996).
4. Johnston, J.H. et al.: Application of Cognitive Load Theory to Develop a Measure of Team Cognitive Efficiency. Military Psychology. 25, 3, 252–265 (2013).
5. Marquart, L.C. et al.: rENA, https://cran.r-project.org/web/packages/rENA/citation.html, last accessed 2020/06/29.
6. Mees, A.: Parsimonious dynamical reconstruction. International Journal of Bifurcation and Chaos. 3, 03, 669–675 (1993).
7. Raykov, T., Marcoulides, G.A.: On desirability of parsimony in structural equation model selection. Structural Equation Modeling: A Multidisciplinary Journal. 6, 3, 292–300 (1999).
8. Shaffer, D.W.: Quantitative Ethnography. Cathcart Press (2017).
9. Shaffer, D.W., Ruis, A.R.: Epistemic Network Analysis: A Worked Example of Theory-Based Learning Analytics. In: Handbook of Learning Analytics. pp. 175–187 Society for Learning Analytics Research (SoLAR) (2017).
10. Swiecki, Z. et al.: Assessing individual contributions to Collaborative Problem Solving: A network analysis approach. Computers in Human Behavior. 104, 105876 (2020).
11. Vandekerckhove, J. et al.: Model Comparison and the Principle of Parsimony. Oxford University Press (2015).