Assessing Multimodal Dynamics in Multi-Party Collaborative Interactions with Multi-Level Vector Autoregression

Robert G. Moulder
robert.moulder@colorado.edu
University of Colorado Boulder
Boulder, Colorado, USA

Nicholas D. Duran
nicholas.duran@asu.edu
Arizona State University
Glendale, Arizona, USA

Sidney K. D'Mello
sidney.dmello@colorado.edu
University of Colorado Boulder
Boulder, Colorado, USA

ABSTRACT

Multi-level vector autoregression (mlVAR) is a recently developed dynamic network model for assessing multimodal temporal data streams derived from multiple users over time. Importantly, mlVAR facilitates investigations into highly complex collaborative interactions within a unified framework. In order to demonstrate the utility of mlVAR for understanding the temporal dynamics of multimodal multi-party (MMP) interactions, we apply it to 9 signals measured from 201 users (67 triads) who engaged in a 15-minute collaborative problem solving task. Measured signals reflect participants' affective states (positive valence and negative valence), physiological states (skin conductance and heart rate), attention (gaze fixation duration and gaze dispersion), nonverbal communication (head acceleration and facial expressiveness), and verbal communication (speech rate). Using node-level metrics of in-strength, out-strength, and synchrony, we show that mlVAR is capable of teasing apart complex role-based dynamics (controller, primary contributor, or secondary contributor) between participants. Our findings also provide evidence for a complex feedback system between individuals where internal states (i.e., skin conductance) are influenced by external signals of shared attention and communication (i.e., gaze and speech).

CCS CONCEPTS

• **Applied computing** → **Psychology**.

KEYWORDS

network analysis, time series, collaborative problem solving

ACM Reference Format:
Robert G. Moulder, Nicholas D. Duran, and Sidney K. D'Mello. 2022. Assessing Multimodal Dynamics in Multi-Party Collaborative Interactions with Multi-Level Vector Autoregression. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '22), November 7–11, 2022, Bengaluru, India.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3536221.3556595

1 INTRODUCTION

Collaboration between individuals in specialized roles is a key component of solving complex problems at both large and small scales. Successful large-scale humanitarian efforts may require collaboration between aid workers, local politicians, and funding sources. Product development at a moderate-sized company requires collaboration and effective communication between management, production facility personnel, development teams, and quality assurance experts. Even small-scale collaborations such as students working on a classroom project may require role assignment and work division to achieve high marks. When collaboration efforts succeed, solutions are found to problems that would otherwise not have been found. When collaboration efforts fail, resources are wasted, time and money are spent, and in extreme cases people may be injured or die. Thus, understanding the mechanisms behind maintaining strong and active collaborations is imperative to increasing innovation, stoking creativity, and spurring invention. However, despite the ubiquity and impact of multi-party collaborations, they are notoriously difficult to study due to the inherent complexity of multimodal information streams interacting together as humans collaborate with other humans or machines.

This difficulty should be no surprise to many researchers, as humans are widely differing in their affective, behavioral, and cognitive qualities. Additionally, these qualities are not static within individuals, but are a function of that person's current and past environment, biological state, and psychological state. Each of these qualities may directly or indirectly influence an individual's efficiency and success at solving a problem. It is also no surprise then that successful multi-party collaborations are difficult to maintain [23, 40]. Complex systems of interactions between many agents may weaken or break down entirely due to numerous factors [38]. Because multi-party collaborative dynamics are more than the sum of the qualities of individual interacting agents, studying multimodal and multi-party (MMP) collaboration becomes increasingly difficult. With individual agents constantly creating multimodal associations with other agents (that also have unique and dynamic qualities), the complexity of studying multi-party interactions quickly grows and more complex multimodal data streams are needed to fully represent all of the dynamics involved between and within each agent. As these data streams increase in complexity, so too must analytic methods used to model these data streams.

One approach to drawing meaningful inferences from highly complex systems is with network analysis [42]. Network analysis is a class of statistical and computational analysis for modeling complex systems[1]. Network models are powerful computational tools

ICMI '22, November 7–11, 2022, Bengaluru, India

ACM ISBN 978-1-4503-9390-4/22/11.
https://doi.org/10.1145/3536221.3556595

[1] In this paper we make a clear distinction between network analysis/network models and artificial neural networks (ANNs). Artificial neural networks seek to estimate an underlying functional relationship between a set of inputs and a set of outputs, generally for the purpose of prediction. Network analytic models seek to estimate associations between all variables of interest for the purpose of statistical inference and modeling.

for simultaneously assessing relationships between large numbers of variables/features across multiple domains. In a network analytic framework, a graph is constructed representing underlying connections between units in a system. Variables in this graph are represented as nodes and the connections between those variables are represented as edges. These graphs may be directed or undirected and can estimate relationships between large amounts of data. Because of these qualities, network analytic approaches have been shown to be a useful tool for inferential analyses of complex multimodal data streams [31, 60]. Unlike many common statistical and machine learning techniques, network analytic approaches seek to jointly estimate the mutual influence between large numbers of variables for the purposes of statistical inference in favor of predictive strength [18].

Recently, network-based models have been developed for studying systems that change dynamically over time. These dynamic networks are capable of modeling how the state of a variable (or set of variables) at a given time point affect itself or another variable at a future time point. This allows for the successful inferential modeling of simultaneous and mutually interacting temporal multimodal data streams both within and between agents, making dynamic network models a prime candidate for the study of multi-party collaboration paradigms.

We employ one such dynamic network model (multilevel vector autoregression, a.k.a. mlVAR) for modeling multimodal affective, behavioral, and cognitive data streams derived from a role-based triadic collaborative problem solving task. Specifically, we use mlVAR to model how students' emotional, physiological, attentional, verbal communication, and nonverbal communication dynamics influence each other, and are in-turn influenced by each other, while solving digital physics-based puzzles in teams of three. Insights derived from these modalities show clear patterns in how students taking a lead in a group-based problem solving task influence, and are influenced by, actively engaged and less-actively engaged collaborators. These findings are evidence that dynamic network modalities such as mlVAR are appropriate and invaluable tools for modeling complex multimodal data derived from multi-party collaborations.

The successful adoption of mlVAR models into the study of MMP collaboration tasks may further the development of tools for increasing the collaborative efficiency of teams, identifying unique and important dynamic links between team members, or training artificial intelligence systems to become optimal team members when interacting with humans. Findings from the current study can imply that the multimodal dynamics of individuals in specific roles within a MMP collaborative environment have differential influences on emotional, behavioral, and cognitive signals within their team members. Thus, if altering the dynamics of a MMP collaborative environment is warranted, mlVAR yields specific targets (i.e., what team member and what data stream) on which interested parties may best focus their resources and efforts.

1.1 Related Work

Multi-party Collaboration. Although collaboration is a powerful tool for solving problems, maintaining and managing successful collaborations between agents is difficult. Kerr and Tindale [38] use the term *process loss* to describe the trend of collaboration efforts becoming less effective over time due to breakdowns in social interactions, communications, or a lack of resources necessary to maintain effective collaboration between individuals. Indeed, it appears that some process loss is inevitable as the complexity of the task and the number of agents involved in active collaboration increases exponentially [55].

Despite process loss, successful large-scale collaboration efforts do exist and Barron [7] makes the claim that successful large-scale collaboration efforts persist due to the mutuality of exchanges, the achievement of joint attentional engagement, and the alignment of group members' goals for the problem solving process. Thus, researchers interested in bolstering the effectiveness of collaboration efforts, while staving off process loss, have focused on understanding what psychological, physiological, and structural processes differentiate successful collaboration endeavors from unsuccessful ones. Thus far, researchers have found these differences to be highly complex and dependent on many multimodal signal interactions, team makeup, and situational factors. For example, Vrzakova et al. [62] found that perceived collaboration success in small teams of students was determined by differing patterns of speech and body movement synchronization between each student. Longer periods of silence and less movement were positively correlated with how well students performed on a shared task. Stewart et al. [57] showed that team diversity on a number of metrics (e.g., personality, demographics, and prior experience) was associated with collaboration success of a shared task. Using recurrence quantification analysis, Eloy et al. [28] demonstrated that less regularity in multimodal data streams (i.e., more novel collaboration patterns) was associated with collaboration success.

Multimodal Complexity. It is the clear that maintaining successful collaboration is a difficult problem involving many multimodal processes occurring within and between collaborative agents. In an attempt to simplify this complexity, we discuss the interactions between the multimodal processes occurring within and between agents using the following categorizations: (a) influential processes, (b) influenced processes, and (c) synchronized processes. Influential processes are processes that, when change or are changed, cause other processes to change. Influenced processes are processes that are changed when another process (or set of processes) change. When processes are simultaneously influential and influenced by one another, these processes are synchronized. Each of these types of multimodal processes serve an important purpose in collaborative endeavours as means of information transfer, Figure 1.

Influential and influenced processes facilitate information transfer between collaborative agents. Ochoa and Dominguez [44] showed that automated multimodal training systems that offer immediate feedback for oral presentations showed a significant positive influence on users preparing to give oral presentations to a class such that users were perceived to have improved their oral presentation skills. Boker et al. [12] demonstrated the influential processes of emotional information transfer between facial expressions in a collaborative communication task between digital avatars of pairs of real interacting agents. By dampening the emotional expressions of one digital avatar, they were able to reliably elicit stronger emotional expressions in the person controlling the second avatar. Boker et al. [12] hypothesized this is due to a shared equilibrium

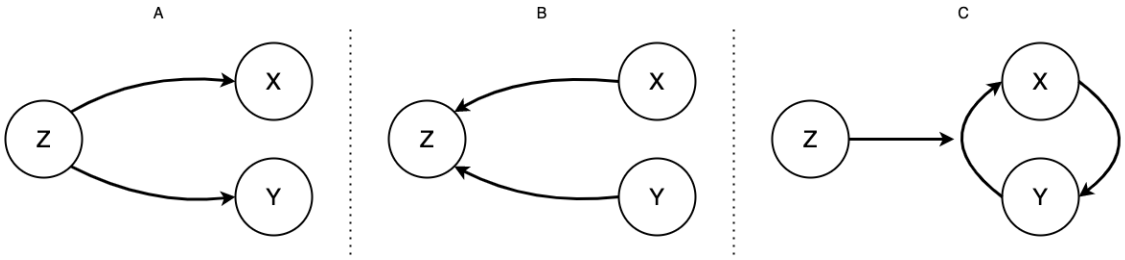

Figure 1: Example network models of variable Z in influential, influenced, and synchronous relationships with variables X and Y. In a network analytic framework, nodes (circles) represent variables and edges (arrows) represent directional influence. In sub-figure (A), node Z is influencing nodes X and Y. In sub-figure (B), node Z is influenced by node X and Y. In sub-figure (C) node Z influences the synchronous/jointly influential relationship between X and Y.

state occurring during conversation which, if violated by one individual, causes a correction by the other. This finding is important to studying multimodal collaboration processes as emotional intensity and affective state predicts the willingness of agents to interact with others for long periods of time [9]. Brennan et al. [14] found that periods of shared gaze between pairs of individuals engaged in a problem solving task were significantly related to collaboration effectiveness. Moulder et al. [43] argues that chaotic behaviors in head motion during collaborative communication tasks form emergent high-dimensional states that facilitate communication. Beyan et al. [10] also argues that such emergent states are necessary for role formation in collaborative tasks, such as leader or supervisory roles. Due to the universality of influenced and influential processes in collaborative paradigms, Cooke and Gorman [22] have developed numerous metrics for determining influence between team members, such as "dominance" and information sharing.

Synchronized states between multimodal signals also facilitate information exchange between collaborative agents, leading to successful multi-party collaboration. Especially in collaborative systems with human agents, synchrony is a core aspect of facilitating social and developmental processes[30, 50]. Synchronization between physiological signals of collaborators (e.g., heart rate and skin conductance) has been shown to be associated with increased success rates in collaborative problem solving paradigms [8]. Similarly, Ashenfelter et al. [4] found that the amount of symmetry between influential and influenced head motions in conversation undergoes periods of building and breaking, and that these periods are necessary for effective information transfer. Chikersal et al. [19] argue that synchronization between facial expressions and skin conductance between collaborating agents is indicative of a group's capacity to perform a wide variety of tasks. Synchronization between emotional states, body motion, and attention has also been shown to facilitate effective communication and understanding between members of a group [1, 32, 48]. It is of note that although synchronization is important in influencing collaborative outcomes (e.g., problem solving ability), synchronization is not always a positive influence on collaborative outcomes and may lead to worse team performance [3].

1.2 Current Directions in Multimodal Multi-Party Data Analysis

Researchers are aware of the need to collect multimodal data streams from multi-party collaboration studies in order to effectively understand multi-party collaborative processes. However, until recently, the majority of research on collaboration has focused specifically on a small number of signals collected from dyads (pairs of collaborating agents). This is due to a combination of difficulties in collecting multimodal signals from multiple agents engaged in shared collaboration and a lack of computational methods geared towards high dimensional multi-party data streams.

The focus on a small number of signals is beginning to change as researchers have begun to advance computational methodologies to be able to handle high dimensional MMP data streams. For instance, Amon et al. [3] developed a means of assessing the collaboration skills of individuals in triads (groups of three) using multidimensional recurrence quantification analysis, while Subburaj et al. [58] used a weighting based approach to quantify collaboration performance in multi-party interactions. Researchers have also developed numerous group level synchrony metrics using complex systems methodologies [34, 52]. Some researchers have proposed generative/predictive models of MMP collaborative processes, while other researchers such as Burk et al. [16] and Rowley [53] argue that network analysis provides the most natural analytic framework for assessing multimodal multi-party data streams.

Generative and Predictive Models. Predictive models such as random forest algorithms and artificial neural networks are powerful tools for deriving useful insights from MMP collaborative tasks. Grafsgaard et al. [33] utilized both feed-forward and long short-term memory (LSTM) artificial neural networks to learn synchronization patterns between romantic couples at greater than chance levels. Olsen et al. [45] also utilized LSTM networks, as well as measures of entropy, to show that multimodal data feeds showed a significant improvement over singular data feeds in predicting collaborative learning outcomes. Researchers have even used predictive modeling in a MMP collaborative context to create generative models of individuals playing sports such as basketball [36]. While these predictive methods are useful, it is difficult to derive specific information about the underlying multimodal dynamics occurring in MMP collaborative tasks due to the "black-box"

nature of such methods. Other methods such as network analysis trade some predictive and generative capability for inferential information.

Network Analysis of Complex Social Systems. Network analysis is a popular tool for assessing large systems of complex relationships across multiple research fields [2, 20]. Network analysis represents variables as nodes in a large graph, with each node being connected by either directed or undirected edges. These edges are defined by the observed relationships between variables. The resulting graph can be a source of rich information about how multimodal data streams are interacting within and between individual agents [6] and have been used in a variety of ways to study complex social systems. For instance, Golino et al. [31] used a dynamic network analysis approach to determine which tweets were likely from troll accounts during the 2016 US presidential election, and Ruis et al. [54] showed that network analysis was able to distinguish error handling patterns between novice and more experienced surgeons. Durugbo et al. [27] has specifically argued that network analysis is both a useful and practical solution for studying multimodal multi-party collaboration dynamics.

Indeed, network analysis has been successfully applied to studying multi-party collaboration efforts. Barabási et al. [5] used network analysis to understand the collaboration patterns of individual scientists and how these collaboration patterns change the scientific inquiry landscape by creating both clusters of habitual collaborators and large networks spanning many labs. Ramasco et al. [49] showed similar findings with similar scientific collaboration networks as well as movie actor collaboration networks. Network analysis has also been used to understand large-scale collaborative efforts from numerous agents editing Wikipedia pages to collaborations between global tech companies [13, 26].

1.3 Research Questions

In using the mlVAR method, we aim to better understand MMP collaboration dynamics between agents engaged in a role-based collaborative problem solving task. Specifically, we seek to use the mlVAR approach coupled with variable influence and synchrony dynamics in order to answer the following research questions: **RQ1** - How does multimodal information flow between individuals in different roles?, **RQ2** - Which multimodal data streams in what context are most influenced by role?, **RQ3** - Which multimodal data streams in what context are most influential by role?, and **RQ4** - Which multimodal data streams are most involved in synchronization between roles?

We use multimodal data streams derived from participants engaged in a triadic collaborative problem solving task to answer each research question. During this task, each participant was assigned to either be a controller (i.e., to control the game while solving a physics puzzle), or a collaborator who can only give verbal assistance to the controller. The collected data streams represent six specific modalities that index key aspects of interpersonal information exchange in the context of this collaborative problem solving task: emotional information, physiological information, attentional information, nonverbal information, and verbal information. Emotional information is represented by each participant's positive and negative valence, physiological information is represented by each participant's heart rate and skin conductance, attentional information is represented by each participant's average length of gaze fixation and gaze dispersion across the computer screen, nonverbal information is represented by each participant's head acceleration and facial expressiveness, and verbal information is represented by each participant's speech rate, yielding 27 time series per triad (9 data streams $\times$ 3 roles).

1.4 Contribution and Novelty

To our knowledge, this is the first successful attempt to explicitly model multimodal and multi-party collaborative dynamics in a holistic network analysis framework, while taking into account triadic interactions, interaction context (i.e., role), nested data structures, and emotional, behavioral, and cognitive data streams. Unlike previous studies that have focused on studying only 1 or 2 signals at a time in the context of dyadic collaboration, we simultaneously assess 9 signals across each of 3 students engaged in triadic collaboration, yielding 27 time series per triad. We propose a novel application of mlVAR, combined with modern network quantification indices of variable influence and synchrony for studying dynamic multimodal multi-party collaborations. The mlVAR model is well grounded in both complexity theory and graph theory, and provides many metrics for assessing dynamics both within and between collaborating agents. We focus on node in-strength, out-strength, and synchrony. Although we demonstrate the mlVAR model on triads, this approach scales to any number of multi-agent systems. Above and beyond more common network approaches, mlVAR allows for the simultaneous estimation of temporal network dynamics while accounting for nesting structures of multimodal data steams nested within collaborating agents. The mlVAR approach, coupled with the quantification of in-strength, out-strength, and synchrony, contributes to the study of collaborative dynamics of MMP interactions.

2 METHOD

2.1 Data Collection

The data used in this analysis is from a larger study on collaborative problem solving. Only variables and teams relevant to the current analysis are described below [see 57]. Figure 2 represents the data collection and analysis pipeline for this study.

Participants. Participants selected for this analysis were 201 students selected from a larger data set of 288 students from 2 large public universities in the United States (average age = 21.77 years) engaged in a collaborative physics game. Of these 201 students 57% were female with a racial makeup of approximately 53% White, 23% Hispanic/Latino, 17% Asian, 3% Black, and 1% Native American, with 3% reporting "Other". Participants were compensated for their time with either course credit or with an Amazon gift card of 50.00 USD if they completed the study.

Collaboration Task. Participants were partnered into teams of 3 based upon their scheduling availability. Each team was then asked to join an in-lab video-conferencing session (Zoom) and work on solving problems in an online Newtonian physics-based game entitled "Physics Playground" [56], Figure 2. All teams engaged three 15-minute long blocks (a warm-up block and 2 experimental blocks). During each block of the study students were randomly

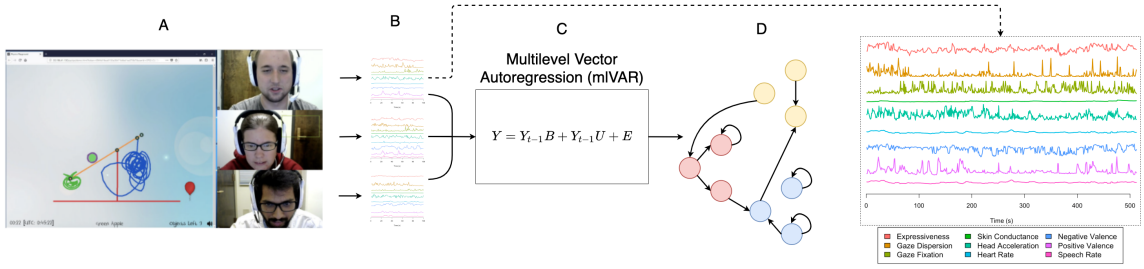

Figure 2: Study pipeline from data collection to final network model. (A) Participants who were assigned to specific roles collaborate on solving a physics based puzzle. (B) During this time, 9 data streams are collected from each participant . A single participant's time series is shown in greater detail to the right of the main figure. (C) All 27 data streams are then assessed with the mlVAR algorithm. (D) Finally the results of the mlVAR algorithm is used to create team-level network models relating variables both within and between team members. In-strength, out-strength, and synchrony scores are then calculated from (D) on a team-by-team basis.

assigned to be either a controller (a person who controlled the mouse and solved the puzzle) or a contributor (a person who could give verbal thoughts and suggestions on the current level). Data for the current paper is taken exclusively from the warm-up block. Students were tasked with using basic principles of Newtonian physics (e.g., gravity, transfer of energy, and leverage/torque) to guide a ball to a goal. The controller's objective was to draw simple machines (e.g., levers) that would interact with the ball on screen and guide the ball to the goal, while the objective of the contributors was to offer their thoughts and suggestions to the controller. Going further, we differentiate the two contributors as the more verbally active contributor across the warm-up block (primary contributor) and the less verbally active contributor across the warm-up block (secondary contributor).

Data Collection and Processing. Collaboration is a multimodal process in which visual, auditory, behavioral, emotional, and attentional information are real-time influences of collaboration dynamics. In order to model this complex process, we collected nine data streams from each participant. Multimodal data streams were extracted at a per-participant level through the use of each participant's webcam, a headset microphone, an eye tracking system, and physiological sensors. All data streams were then down-sampled to a rate of 1Hz for data analysis to correspond to the average length of an utterance. For a triad to be selected for this analysis, each member of the triad must have had observations for each time series. Triads with entirely missing time series were dropped from the analysis. In total, 61 triads (69.7% of the original data set) met our criteria for inclusion in this analysis. Remaining missingness within time series (averaging missingness = 6.36%) was handled with Kalman filtering.

Audio data streams were collected through each participant's headset. These data streams were then fed through IBM Watson's Speech to Text software to yield timestamped transcripts (beginning and end times) of each utterance across the 15-minute warm-up block for each participant. The count of these utterances were then aggregated to 1Hz and defined each participant's speech rate data stream. If an utterance lasted longer than 1 second, it was assigned to the second that it started.

Physiological data streams were collected through the use of Shimmer 3 GSR+ devices. The Shimmer 3 is an unobtrusive wearable device that collects both heart rate (PPG signal) and changes in skin conductance (galvanic skin response) at 51.2Hz. The Shimmer 3 galvanic skin response sensor was placed on each participant's wrist and the heart rate monitor was placed on each participant's earlobe. After data collection, skin conductance was then separated into tonic (slow moving) and phasic (fast moving) components. The current study focuses on the phasic component because the phasic component is sensitive to changes in external stimuli. The Shimmer family of devices have been validated to collect highly accurate and synchronized physiological data streams with minimal error and drift, even on moving participants [17]. Each participant was fitted with a Shimmer 3 during data collection. Both heart rate and skin conductance data streams obtained from the Shimmer devices were down-sampled to 1Hz for analysis using an order eight Chebyshev type I filter.

Emotional data streams (positive valence and negative valence), as well as expressiveness, were collected from videos of participant's faces recorded via webcams attached to each participant's computer. Videos were sampled at 10Hz for the purposes of feature extraction. We utilized the Emotient video analysis software which estimates the likelihoods of the presence of 20 action units relevant to each participant's expressions in each video [41]. These action units were the used to assess participant's positive valence and negative valence using an algorithm developed by Cohn et al. [21]. Expressiveness (overall activity across a given frame) was then calculated as the mean value across the action units.

Motion and attention based data streams were collected with Tobii4C eye tracking devices attached to each participant's computer. Tobii4C devices collected data on each participant's eye gaze and head motion sampled at 90Hz. The Tobii4C devices collected information on the pitch, roll, and yaw of each participant's head. We used participants' visual fixation information (i.e., points where gaze is maintained on a location at a maximum of 25 pixels apart for at least 50ms) to compute fixation dispersion (i.e., the mean Euclidean distance of each raw gaze point in a 1s window from the centroid) and average fixation duration [25]. Fixations longer than 1s were trimmed to 1s and fixations that overlapped boundaries

were assigned to the majority second. Fixations were then averaged over 1s windows. Head pitch, roll, and yaw were converted into X, Y, and Z axis displacement, then into accelerations using two steps of fourth-order central differencing. The magnitude of the X, Y, and Z accelerations was then calculated as a measure of head motion dynamics during the collaborative problem solving task and down-sampled to 1Hz. Head acceleration magnitude has been shown to be a useful data stream for nonverbal information transfer between individuals [e.g., 11].

2.2 Multi-Level Vector Autoregression

Network analytic methods are capable of modeling complex relationships between multiple variables across multiple data clusters (triads in the current case). The mlVAR model is one such network analytic approach that is especially well suited for assessing temporal dynamics in multimodal multi-party data streams [15, 29]. mlVAR accomplishes this task by constructing a series of multiple linear mixed-effects models, each of which predicts a variable at a given time within a cluster $y(t)_{ij}$ by all other variables at time $t-1$:

$$y(t)_{i1} = y(t-1)_{i1}b_{11} + \ldots + y(t-1)_{iJ}b_{1J} + y(t-1)_{i1}u_{1i1} + \ldots + y(t-1)_{iJ}u_{Jij} + e(t)_1,$$
$$y(t)_{i2} = y(t-1)_{i1}b_{21} + \ldots + y(t-1)_{iJ}b_{2J} + y(t-1)_{i1}u_{1i2} + \ldots + y(t-1)_{iJ}u_{Jij} + e(t)_2,$$
$$\vdots$$
$$y(t)_{iJ} = y(t-1)_{i1}b_{J1} + \ldots + y(t-1)_{iJ}b_{JJ} + y(t-1)_{i1}u_{1iJ} + \ldots + y(t-1)_{iJ}u_{Jij} + e(t)_J.$$

where $y(t)_{ij}$ is the value of variable $j \in [1, \ldots, J]$ at time t within cluster i, $y(t-1)_{ij}$ is a time-lagged (lag-1) version of $y(t)_{ij}$, $b_{11} \ldots b_{jj}$ are fixed effects representing average associations between all $y(t)_{ij}$ and $y(t-1)_{ij}$, $u_{1i1} \ldots u_{Jij}$ are random effects representing the cluster-level deviations from $b_{11} \ldots b_{jj}$, and $e(t)_1 \ldots e(t)_J$ are error terms. In the case of the current data, $y(t)_{ij}$ represents positive valence, negative valence, expressiveness, heart rate, skin conductance, speech rate, average gaze fixation duration, gaze dispersion, and head acceleration for controllers, primary contributors, and secondary contributors, as well as a measure of all changes occurring on the screen of the physics game as performed by the controller, for a total of 28 variables per triad. All analyses were conducted with the *mlVAR* R package (version 0.5).

Adjacency matrices can then be constructed from from $b_{11} \ldots b_{JJ}$ and $u_{1i1} \ldots u_{Jij}$ representing general directed connections between nodes and cluster specific connections between nodes respectively:

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1J} \\ b_{21} & b_{22} & \ldots & b_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ b_{J1} & b_{J2} & \ldots & b_{JJ} \end{bmatrix}, \mathbf{U_i} = \mathbf{B} + \begin{bmatrix} u_{1i1} & u_{1i2} & \ldots & u_{1iJ} \\ u_{2i1} & u_{2i2} & \ldots & u_{2iJ} \\ \vdots & \vdots & \ddots & \vdots \\ u_{Ji1} & u_{Ji2} & \ldots & u_{JiJ} \end{bmatrix}$$

Adjacency matrices $\mathbf{B}$ and $\mathbf{U_i}$ form a graph that contains information on the temporal dynamics relating all $y(t-1)_{ij}$ to $y(t)_{ij}$. Elements of these matrices can be read as connections going from columns to rows (e.g., b_{21} represents a connection from node 1 at time $t-1$ to node 2 at time t. It is possible to use the values of $\mathbf{B}$ and $\mathbf{U_i}$ to calculate metrics representing variable influence and synchrony. Here we focus on measures of node degree (in-strength and out-strength) and synchrony.

The total sum of connections (or strength of connections) involving a node within a network is known as the degree of the node. Node degree in directed networks is measured by in-strength (the sum of absolute values of edges going into a node) and out-strength (the sum of absolute values of edges going out of a node) [35]. Node synchrony is a measure of how a set of nodes jointly influence one another [34].

In-strength. Node in-strength represents the total magnitude of connections going from a node (or set of nodes) to a node of interest. For instance, the in-strength of node 1 from nodes 2, 3, and 4 from $\mathbf{B}$ is calculated as:

$$IN_1 = |b_{12}| + |b_{13}| + |b_{14}|.$$

In-strength represents the total amount to which a node is *influenced* within a network. Nodes with higher in-strength compared to nodes with lower in-strength coming from the same set of nodes are therefore more influenced by other nodes within a network. That is, the current dynamics of these nodes are sensitive to changes in other nodes at previous time points.

Out-strength. Node out-strength represents the total magnitude of connections going from a node of interest to a node (or set of nodes). For instance, the out-strength of node 1 to nodes 2, 3, and 4 from $\mathbf{B}$ is calculated as:

$$OUT_1 = |b_{21}| + |b_{31}| + |b_{41}|.$$

Out-strength represents the total amount to which a node *influences* other nodes within a network. Nodes with higher out-strength compared to nodes with lower out-strength going to the same set of nodes are therefore more influential to other nodes within a network. That is, the prior dynamics of these nodes change the current dynamics of other nodes.

Synchrony. Node synchrony may be thought of as the mutual influence of change between sets of nodes within a network. Whereas in-strength represents only the degree to which a node is influenced by changes in other nodes and out-strength represents how influential a change in a node is to other nodes, synchrony is a joint metric representing mutual changes in dynamics. Guastello and Peressini [34] developed a method for modeling synchrony between individual nodes and groups of nodes based upon an optimal linear map of elements of an adjacency matrix.

To illustrate the calculation of node synchrony, consider the following adjacency matrix:

$$\mathbf{B} = \begin{bmatrix} 0.8 & 0.3 & 0.5 & 0.6 \\ 0.2 & 0.7 & 0.4 & 0.3 \\ 0.4 & 0.1 & 0.6 & 0.7 \\ 0.5 & 0.3 & 0.7 & 0.9 \end{bmatrix}, \mathbf{V} = \begin{bmatrix} 0.5 \\ 0.4 \\ 0.7 \end{bmatrix}, \mathbf{M} = \begin{bmatrix} 0.8 & 0.3 & 0.6 \\ 0.2 & 0.7 & 0.3 \\ 0.5 & 0.3 & 0.9 \end{bmatrix}$$

This method begins by choosing a reference node for which to calculate a synchrony score, as well as other nodes with which the reference node may synchronize. In this case we will select node 3 as the reference node and nodes 1, 2, and 4 as the nodes with which node 3 synchronizes. Two additional matrices are then formed from $\mathbf{B}$, Matrix $\mathbf{V}$ is defined as the elements of $\mathbf{B}$ representing node 3 going to all other nodes of interest. Matrix $\mathbf{M}$ is defined as the remaining elements of $\mathbf{B}$ after removing rows and columns associated with node 3. Matrix $\mathbf{V}$ represents the influence of node 3 on all other nodes of interest at a future state. Note that the sum of the absolute values of $\mathbf{V}$ is the out-strength of node 3 to nodes 1, 2, and 4. Matrix $\mathbf{M}$ represents the joint influence of nodes 1, 2, and 4 on themselves at a future state (i.e., the dynamics of these nodes).

The synchrony between these nodes can then be computed as:

$$S_E = \mathbf{V}'\mathbf{M}^{-1}\mathbf{V}$$

In this equation, $\mathbf{M}^{-1}\mathbf{V}$ yields a linear map (i.e., regression weight) from the internal dynamics represented by $\mathbf{M}$ to the external dynamics represented by $\mathbf{V}$. $\mathbf{M}^{-1}$ represents the matrix inverse of $\mathbf{M}$ and $\mathbf{V}'$ represents the matrix transpose of $\mathbf{V}$. A further premultiplication by $\mathbf{V}'$ yields a singular regression weight, S_E, mapping how much the outward dynamics of node 3 ($\mathbf{V}$) influence the association between the outward dynamics of node 3 and the internal dynamics of nodes 1, 2, and 4 ($\mathbf{M}$). That is, S_E represents how much of the joint temporal relationship between nodes 1, 2, 3, and 4 is due to changes in node 3. Higher S_E values represent more synchrony and lower values of S_E represent less synchrony. In this example, $S_E \approx .745$.

For **RQ1**, mlVAR was conducted to yield an average temporal network across all triads that accounts for the inherent nesting of participants within teams in the data. Statistically significant paths were assessed at the $p < .01$ level due to the large number of simultaneous estimations occurring within this model. For **RQ2**, **RQ3**, and **RQ4**, we calculate node in-strength (a measured of being influenced), out-strength (a measure of influence), and synchrony from sub-models derived from mlVAR representing the dynamics of each triad.

All analyses relevant to **RQ2** - **RQ4** we assessed using linear mixed effects models. These models were necessary to account for the inherent nesting of individual participant network metrics within teams. In total, six models were conducted each with either in-strength, out-strength, or synchrony as the outcome variable and participant role (ROLE) or data stream (MODALITY) as predictor variables. Models with participant role as a predictor take the form:

$$METRIC_{ij} = f(ROLE) + u_{0j} + e_{ij}$$

Models with data stream as a predictor take the form:

$$METRIC_{ij} = f(MODALITY) + u_{0j} + e_{ij}$$

where $METRIC$ is either in-strength, out-strength, or synchrony score, $f(ROLE)$ and $f(MODALITY)$ are linear models of the form $\mathbf{XB}$ that includes all main effects (3 for ROLE and 9 for MODALITY), u_{0j} represents a random intercept term per team, and e_{ij} is an error term. A false discovery rate correction was then conducted to control for Type-I statistical errors. We report the results of all between-role contrasts and the strongest effects shown for the between-modality contrasts.

3 RESULTS

3.1 Global Network

The mlVAR method estimated network graphs for individual triads (i.e., random effects), as well as a single network representing statistically significant general findings across all estimated networks (i.e., fixed effects). This single graph is shown in Figure 3. Findings from the overall network answer **RQ1**. Although this network was complex given the large amount of significant associations, there was also a large body of qualitative information to glean from this network. For instance, the strongest connections existed between a variable and itself at the next time point, indicating that these variables tend to greatly influence their own dynamics across time. Secondly, there were a relatively equal number of significant connections within any team member (40 for primary contributor, 39 for primary contributor, 40 for secondary contributor) as there are between any member and the other two team members (average number of connections = 35). This indicated that there was indeed information transfer between these multimodal signals occurring during collaborative interaction and that regulatory processes existed both within a single participant and between that participant and both other collaborators. Analyses relating to **RQ2** - **RQ4** go into further detail.

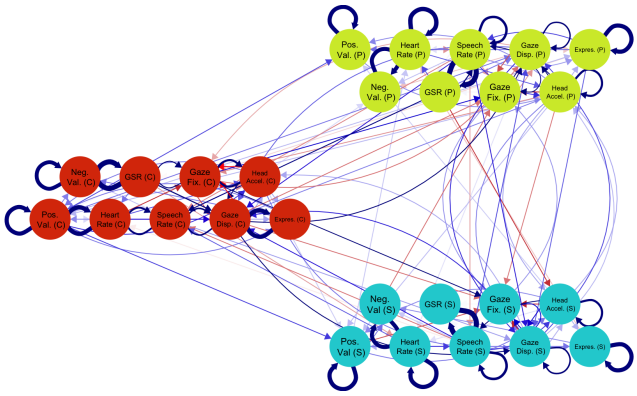

Figure 3: Multilevel vector autoregression network graph of temporal associations between role-based multimodal data streams. Nodes (circles) represent within-person time series variables. Red nodes are from the controller (C), yellow nodes are from the most verbose (i.e., primary) contributor (P), and blue nodes are from the least verbose (i.e., secondary) contributor (S). Within each role, nodes represent positive valence (Pos. Val.), negative valence (Neg. Val.), skin conductance (GSR), heart rate, speech rate, gaze fixation (Gaze Fix.), gaze dispersion (Gaze Disp.), head acceleration (Head Accel.), expressiveness (Expres.). Directed edges (arrows) represent temporal associations between nodes, with red edges indicating significant negative associations and blue edges indicating significant positive associations. Edge thickness indicates connection strength.

3.2 Individual Triad Networks

In addition to the global model, mlVAR estimated networks similar to that of Figure 3 for each triad. In order to understand collaboration dynamics within and across triads for each of multimodal signals (emotional, physiological, attentional, nonverbal, and verbal), in-strength, out-strength, and synchrony was calculated for all nodes. All analyses were conducted at the within-team level using in-strength, out-strength, and synchrony values assessed from a participant in one role to participants in other roles. For instance, controller in-strength is calculated as all edges going to controller nodes from primary and secondary contributor nodes.

Linear mixed-effect models were then used to understand how participant role influenced average in-strength, out-strength, and synchrony scores between an individual in a given role and their

two collaborators. These same models also modeled how different modalities influenced average levels of in-strength, out-strength and synchrony scores. A random effect of team was included to account for the nesting of participants within teams, Figure 4.

In-Strength. Node in-strength represents how much a single data stream from a single participant was jointly influenced by all other data streams from both other participants (**RQ2**, Figure 4-A and 4-D). Controllers were significantly more influenced (i.e., have higher average in-strength) than primary contributors ($p = .001$) or secondary contributors ($p < .001$). Primary contributors were also significantly more influenced than secondary contributors ($p < .001$). Looking more closely at individual data streams, all participants are mostly influenced on their skin conductance (all $ps < .001$). This indicates that influence may be best observed through physiological arousal.

Out-Strength. Node out-strength represents how much a single data stream from a single participant was able to collectively influence all other data streams from both other participants (**RQ3**, Figure 4-B and 4-E). There were no significant role differences in out-strength (all $ps > .836$). However, at a per data stream level, both gaze dispersion and gaze fixation showed significantly larger influence on all other variables compared to all other variables within a role (all $ps < .001$). This indicates that participant's attentional information directed the behavior of other signals within other participants.

Synchrony. Node synchrony represents how much a sets of relationship strengths between data signals are jointly influenced by specific nodes (**RQ4**, Figure 4-C and 4-F). There were no significant role differences in synchrony ($ps > .575$). Between modality, similar to out-strength, both gaze dispersion and gaze fixation showed significantly more influence on the synchronization/relationship between data signals compared to other signals (all $ps < .001$).

4 DISCUSSION

Collaboration is a complex process involving dynamic multimodal multi-party interactions. We have shown that mlVAR is capable of modeling multimodal multi-party data over time. The results of mlVAR are a set of adjacency matrices that can be used to test complex hypotheses regarding both within and between participant dynamics. Although we have shown a specific case of mVAR applied to triads, mlVAR can theoretically be scaled to any number of group members with any number of shared or uncommon variables. This makes mlVAR an invaluable tool for researchers interested in a complex and holistic view of multimodal collaboration processes.

4.1 Main Findings

Our findings emphasize the importance of each role on the dynamics of the team. Team members are in general similar in their ability to influence and synchronize the modalities of other team members (i.e., indistinguishable out-strength and synchrony values across roles). However, the controller and primary contributors have a unique place in the team as their modalities are the most influenced by the dynamics of the team (i.e., controllers show the highest in-strength levels, followed by primary contributors, followed by secondary contributors). This makes sense in the context of the current study as controllers were the only ones able to directly effect the end result of a given physics puzzle.

At an individual modality level, we find that the internal states of participants (as measured by skin conductance) are most influenced by their team members, while participants' attention (as measured by gaze fixation and dispersion) was the most influential modality shared between participants, as well as the modality most influential to team synchrony. Interestingly, while skin conductance was highly influenced by other team members, it is not at all influential to other team members' data streams or the synchronization between those data streams. This may mean that the influential dynamics of collaboration tend to have the highest influence on physiological arousal, and that physiological arousal is a data stream only valuable in information transfer within an individual and not between individuals. Thus, it appears that a main component of team collaboration dynamics involves processes in which overt signals between individuals (e.g., attention and verbal modalities) influence the internal states of singular members. This change in internal state of individuals may then focus the collective behavior of a team toward a common goal.

4.2 Applications and Implications

All of these findings would be difficult to ascertain outside of the mlVAR framework. If properly implemented, the mlVAR modality has the potential to offer researchers unparalleled insight into the dynamics of interpersonal collaboration between multiple agents. Possible applications of the findings of the current manuscript exist for both real time team collaboration optimization and recommender systems. Researchers such as Palau et al. [46] have shown that behavioral dynamics estimated through network analysis can be used to create recommender systems as a means of improving collaboration between individuals or to create strategic collaborative groupings of agents [24]. Results from an mlVAR analysis may also be used as a "team fingerprint" in identifying specific teams by their shared multimodal dynamics [47], or allow artificial intelligence agents to have a better understanding of human affective states in order to assess and mitigate collaboration issues [51]. Network perturbation testing also offers valuable "what-if" scenario testing by allowing researcher to change specific parameters or data streams of a learned network model to understand the expected change in other data streams at a later time point [39]. This would be useful in determining specific changes within a given team that might improve team performance over time or keep team performance from falling.

4.3 Limitations/Future Work

Although mlVAR is a well-suited analysis for studying collaboration in a MMP framework, there are limitations to this method that researchers should be aware of. Two primary concerns are the possibility of false positives or false negatives in the network model estimation process. As the mlVAR model is relatively new, little research has been done on the error rates and statistical power of these models. There are suggestions on minimum sample sizes and effect size calculations, but there is little formal analyses done to assess these statistical properties [37, 61]. Although it can be expected that as the number of multimodal signals or number of

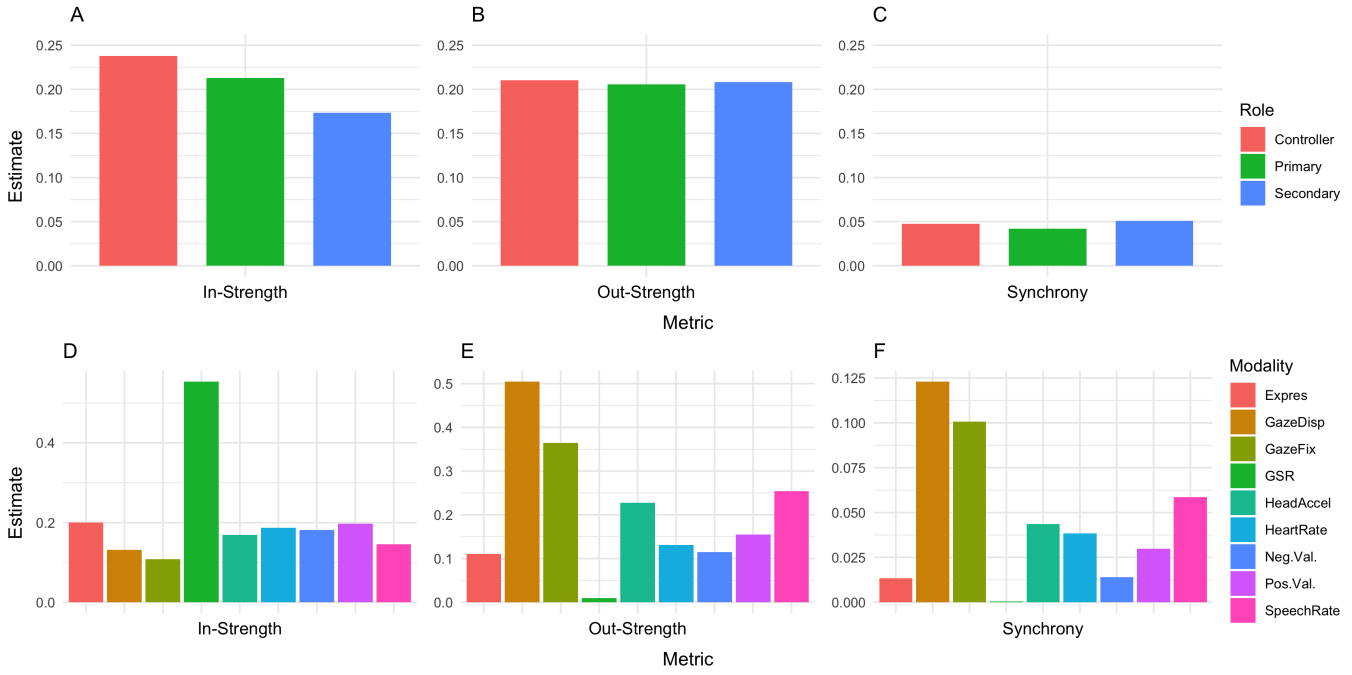

Figure 4: Plots of individual network results. (A) - (C) represent results for each role by in-strength, out-strength, and synchrony respectively. (D) - (F) represent results by modality for in-strength, out-strength, and synchrony respectively.

group members increases, more data will be necessary both at the individual level and the group level.

Additionally, although we have only discussed in-strength, out-strength, and synchrony as being measures of how influenced/influential a variable is and how much synchrony is due to a specific variable respectively, there are dozens more network metrics that can be calculated from mlVAR [35]. Each metric has its own meaning and may show differential interest to different researchers. More metrics are constantly being developed for quantifying network dynamics in order to study specific aspects of network functions and differences between networks [59].

4.4 Concluding Remarks

Network models are an invaluable tool for inferential analysis of multimodal multi-party collaboration processes. We have demonstrated that the mlVAR model is an especially well suited method for uncovering collaboration dynamics occurring within and between collaborative agents. The mlVAR model is uniquely able to handle this large number of variables, as well as a large number of collaborators, making it an indispensable tool for understanding complex collaboration efforts. Additionally, new network metrics are being developed yearly, each with their own specific representation of the complex dynamics occurring in temporal network models. As the mlVAR model is still actively being improved upon, we believe it will continue to increase its utility for studying multimodal multi-party processes and will become a common analysis for individuals interested in studying group dynamics such as collaboration.

REFERENCES

[1] Drew H. Abney, Alexandra Paxton, Rick Dale, and Christopher T. Kello. 2015. Movement dynamics reflect a functional role for weak coupling and role structure in dyadic problem solving. *Cognitive Processing* 16, 4 (nov 2015), 325–332. https://doi.org/10.1007/s10339-015-0648-2

[2] Réka Albert and Albert-László Barabási. 2002. Statistical mechanics of complex networks. *Reviews of Modern Physics* 74, 1 (jan 2002), 47–97. https://doi.org/10.1103/RevModPhys.74.47 arXiv:0106096 [cond-mat]

[3] Mary Jean Amon, Hana Vrzakova, and Sidney K. D'Mello. 2019. Beyond Dyadic Coordination: Multimodal Behavioral Irregularity in Triads Predicts Facets of Collaborative Problem Solving. *Cognitive Science* 43, 10 (2019), 1–22. https://doi.org/10.1111/cogs.12787

[4] Kathleen T Ashenfelter, Steven M Boker, Jennifer R Waddell, and Nikolay Vitanov. 2009. Spatiotemporal symmetry and multifractal structure of head movements during dyadic conversation. *Journal of experimental psychology. Human perception and performance* 35, 4 (2009), 1072–1091. https://doi.org/10.1037/a0015017

[5] A.L Barabási, Hawoong Jeong, Zoltan Néda, Erzsebet Ravasz, Andras Schubert, and Tamas Vicsek. 2002. Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications* 311, 3-4 (aug 2002), 590–614. https://doi.org/10.1016/S0378-4371(02)00736-7

[6] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani. 2004. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences* 101, 11 (mar 2004), 3747–3752. https://doi.org/10.1073/pnas.0400087101 arXiv:0311416 [cond-mat]

[7] Brigid Barron. 2000. Achieving Coordination in Collaborative Problem-Solving Groups. *Journal of the Learning Sciences* 9, 4 (oct 2000), 403–436. https://doi.org/10.1207/S15327809JLS0904_2

[8] F. Behrens, J. A. Snijdewint, R. G. Moulder, E. Prochazkova, E. E. Sjak-Shie, S. M. Boker, and M. E. Kret. 2020. Physiological synchrony is associated with cooperative success in real-life interactions. *Scientific Reports* 10, 1 (dec 2020), 19609. https://doi.org/10.1038/s41598-020-76539-8

[9] Diane S. Berry and Jane Sherman Hansen. 1996. Positive affect, negative affect, and social interaction. *Journal of Personality and Social Psychology* 71, 4 (oct 1996), 796–809. https://doi.org/10.1037/0022-3514.71.4.796

[10] Cigdem Beyan, Francesca Capozzi, Cristina Becchio, and Vittorio Murino. 2016. Identification of emergent leaders in a meeting scenario using multiple kernel learning. In *Proceedings of the 2nd Workshop on Advancements in Social Signal Processing for Multimodal Interaction - ASSP4MI '16*. ACM Press, New York, New York, USA, 3–10. https://doi.org/10.1145/3005467.3005469

[11] Laura Bishop and Werner Goebl. 2018. Beating time: How ensemble musicians' cueing gestures communicate beat position and tempo. *Psychology of Music* 46, 1 (jan 2018), 84–106. https://doi.org/10.1177/0305735617702971

[12] Steven M Boker, Jeffrey F Cohn, Barry-John Theobald, Iain Matthews, Timothy R Brick, and Jeffrey R Spies. 2009. Effects of damping head movement and facial expression in dyadic conversation using real-time facial expression tracking and synthesized avatars. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 364, 1535 (2009), 3485–3495. https://doi.org/10.1098/rstb.2009.0152

[13] Ulrik Brandes, Patrick Kenis, Jürgen Lerner, and Denise van Raaij. 2009. Network analysis of collaboration structure in Wikipedia. In *Proceedings of the 18th international conference on World wide web - WWW '09*. ACM Press, New York, New York, USA, 731. https://doi.org/10.1145/1526709.1526808

[14] Susan E. Brennan, Xin Chen, Christopher A. Dickinson, Mark B. Neider, and Gregory J. Zelinsky. 2008. Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition* 106, 3 (mar 2008), 1465–1477. https://doi.org/10.1016/j.cognition.2007.05.012

[15] Laura F. Bringmann, Nathalie Vissers, Marieke Wichers, Nicole Geschwind, Peter Kuppens, Frenk Peeters, Denny Borsboom, and Francis Tuerlinckx. 2013. A Network Approach to Psychopathology: New Insights into Clinical Longitudinal Data. *PLoS ONE* 8, 4 (apr 2013), e60188. https://doi.org/10.1371/journal.pone.0060188

[16] William J. Burk, Christian E.G. Steglich, and Tom A.B. Snijders. 2007. Beyond dyadic interdependence: Actor-oriented models for co-evolving social networks and individual behaviors. *International Journal of Behavioral Development* 31, 4 (jul 2007), 397–404. https://doi.org/10.1177/0165025407077762

[17] Adrian Burns, Emer P. Doheny, Barry R. Greene, Timothy Foran, Daniel Leahy, Karol O'Donovan, and Michael J. McGrath. 2010. SHIMMER: An extensible platform for physiological signal capture. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE, Buenos Aires, Argentina, 3759–3762. https://doi.org/10.1109/IEMBS.2010.5627535

[18] Danilo Bzdok, Denis Engemann, and Bertrand Thirion. 2020. Inference and Prediction Diverge in Biomedicine. *Patterns* 1, 8 (nov 2020), 100119. https://doi.org/10.1016/j.patter.2020.100119

[19] Prerna Chikersal, Maria Tomprou, Young Ji Kim, Anita Williams Woolley, and Laura Dabbish. 2017. Deep Structures of Collaboration: Physiological Correlates of Collective Intelligence and Group Satisfaction. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, New York, NY, USA, 873–888. https://doi.org/10.1145/2998181.2998250

[20] Claire Christensen and Réka Albert. 2007. Using Graph Concepts to Understand the Organization of Complex Systems. *International Journal of Bifurcation and Chaos* 17, 07 (jul 2007), 2201–2214. https://doi.org/10.1142/S021812740701835X

[21] Jeffrey F. Cohn, Laszlo A. Jeni, Itir Onal Ertugrul, Donald Malone, Michael S. Okun, David Borton, and Wayne K. Goodman. 2018. Automated Affect Detection in Deep Brain Stimulation for Obsessive-Compulsive Disorder. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*. ACM, New York, NY, USA, 40–44. https://doi.org/10.1145/3242969.3243023

[22] Nancy J. Cooke and Jamie C. Gorman. 2009. Interaction-Based Measures of Cognitive Systems. *Journal of Cognitive Engineering and Decision Making* 3, 1 (mar 2009), 27–46. https://doi.org/10.1518/155534309X433302

[23] Catherine Durnell Cramton. 2001. The Mutual Knowledge Problem and Its Consequences for Dispersed Collaboration. *Organization Science* 12, 3 (2001), 346–371. https://doi.org/10.1287/orsc.12.3.346.10098

[24] Rob Cross, Stephen P Borgatti, and Andrew Parker. 2002. Making Invisible Work Visible: Using Social Network Analysis to Support Strategic Collaboration. *California Management Review* 44, 2 (jan 2002), 25–46. https://doi.org/10.2307/41166121

[25] Edwin S. Dalmaijer, Sebastiaan Mathôt, and Stefan Van der Stigchel. 2014. PyGaze: an open-source, cross-platform toolbox for minimal-effort programming of eyetracking experiments. *Behavior research methods* 46, 4 (2014), 913–921. https://doi.org/10.3758/s13428-013-0422-2

[26] Giuditta De Prato and Daniel Nepelski. 2012. Global technological collaboration network: network analysis of international co-inventions. *The Journal of Technology Transfer* 39, 3 (dec 2012), 358–375. https://doi.org/10.1007/s10961-012-9285-4

[27] Christopher Durugbo, Windo Hutabarat, Ashutosh Tiwari, and Jeffrey R. Alcock. 2011. Modelling collaboration using complex networks. *Information Sciences* 181, 15 (aug 2011), 3143–3161. https://doi.org/10.1016/j.ins.2011.03.020

[28] Lucca Eloy, Angela E.B. Stewart, Mary Jean Amon, Caroline Reinhardt, Amanda Michaels, Chen Sun, Valerie Shute, Nicholas D. Duran, and Sidney D'Mello. 2019. Modeling Team-level Multimodal Dynamics during Multiparty Collaboration. In *2019 International Conference on Multimodal Interaction*. ACM, New York, NY, USA, 244–258. https://doi.org/10.1145/3340555.3353748

[29] Sacha Epskamp, Lourens J. Waldorp, René Mõttus, and Denny Borsboom. 2018. The Gaussian Graphical Model in Cross-Sectional and Time-Series Data. *Multivariate Behavioral Research* 53, 4 (jul 2018), 453–480. https://doi.org/10.1080/00273171.2018.1454823 arXiv:1609.04156

[30] Ruth Feldman, Romi Magori-Cohen, Giora Galili, Magi Singer, and Yoram Louzoun. 2011. Mother and infant coordinate heart rhythms through episodes of interaction synchrony. *Infant Behavior and Development* 34, 4 (2011), 569–577. https://doi.org/10.1016/j.infbeh.2011.06.008

[31] Hudson Golino, Alexander P. Christensen, Robert Moulder, Seohyun Kim, and Steven M. Boker. 2022. Modeling Latent Topics in Social Media using Dynamic Exploratory Graph Analysis: The Case of the Right-wing and Left-wing Trolls in the 2016 US Elections. *Psychometrika* 87, 1 (mar 2022), 156–187. https://doi.org/10.1007/s11336-021-09820-y

[32] Yulia Golland, Yossi Arzouan, and Nava Levit-Binnun. 2015. The mere Co-presence: Synchronization of autonomic signals and emotional responses across Co-present individuals not engaged in direct interaction. *PLoS ONE* 10, 5 (2015), 1–13. https://doi.org/10.1371/journal.pone.0125804

[33] Joseph Grafsgaard, Nicholas Duran, Ashley Randall, Chun Tao, and Sidney D'Mello. 2018. Generative multimodal models of nonverbal synchrony in close relationships. *Proceedings - 13th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2018*, 195–202. https://doi.org/10.1109/FG.2018.00037

[34] Stephen J. Guastello and Anthony F. Peressini. 2017. Development of a Synchronization Coefficient for Biosocial Interactions in Groups and Teams. *Small Group Research* 48, 1 (feb 2017), 3–33. https://doi.org/10.1177/1046496416675225

[35] London Heathrow and The Roman. 2016. Node Degree and Strength. In *Fundamentals of Brain Network Analysis*. Elsevier, 115–136. https://doi.org/10.1016/B978-0-12-407908-3.00004-2

[36] Boris Ivanovic, Edward Schmerling, Karen Leung, and Marco Pavone. 2018. Generative Modeling of Multimodal Multi-Human Behavior. *IEEE International Conference on Intelligent Robots and Systems* January (2018), 3088–3095. https://doi.org/10.1109/IROS.2018.8594393 arXiv:1803.02015

[37] D. Gage Jordan, E. Samuel Winer, and Taban Salem. 2020. The current status of temporal network analysis for clinical science: Considerations as the paradigm shifts? *Journal of Clinical Psychology* 76, 9 (sep 2020), 1591–1612. https://doi.org/10.1002/jclp.22957

[38] Norbert L. Kerr and R. Scott Tindale. 2004. Group performance and decision making. *Annual Review of Psychology* 55 (2004), 623–655. https://doi.org/10.1146/annurev.psych.55.090902.142009

[39] Gueorgi Kossinets and Duncan J. Watts. 2006. Empirical Analysis of an Evolving Social Network. *Science* 311, 5757 (jan 2006), 88–90. https://doi.org/10.1126/science.1116869

[40] Robert Kraut, Carmen Edigo, and Jolene Galegher. 1988. Patterns of Contact and Communication Collaboration in Scientific Research Collaboration. *CSCW Proceedings ACM Conference* (1988), 1–12.

[41] Gwen Littlewort, Jacob Whitehill, Tingfan Wu, Ian Fasel, Mark Frank, Javier Movellan, and Marian Bartlett. 2011. The computer expression recognition toolbox (CERT). In *Face and Gesture 2011*. IEEE, 298–305. https://doi.org/10.1109/FG.2011.5771414

[42] Melanie Mitchell. 2006. Complex systems: Network thinking. *Artificial Intelligence* 170, 18 (dec 2006), 1194–1212. https://doi.org/10.1016/j.artint.2006.10.002

[43] Robert G. Moulder, Elena Martynova, and Steven M. Boker. 2022. Extracting Nonlinear Dynamics from Psychological and Behavioral Time Series Through HAVOK Analysis. *Multivariate Behavioral Research* 0, 0 (jan 2022), 1–25. https://doi.org/10.1080/00273171.2021.1994848

[44] Xavier Ochoa and Federico Dominguez. 2020. Controlled evaluation of a multimodal system to improve oral presentation skills in a real learning setting. *British Journal of Educational Technology* 51, 5 (2020), 1615–1630. https://doi.org/10.1111/bjet.12987

[45] Jennifer K. Olsen, Kshitij Sharma, Nikol Rummel, and Vincent Aleven. 2020. Temporal analysis of multimodal data to predict collaborative learning outcomes. *British Journal of Educational Technology* 51, 5 (2020), 1527–1547. https://doi.org/10.1111/bjet.12982

[46] Jordi Palau, Miquel Montaner, Beatriz López, and Josep Lluís de la Rosa. 2004. Collaboration Analysis in Recommender Systems Using Social Networks. In *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*. Vol. 3191. 137–151. https://doi.org/10.1007/978-3-540-30104-2_11

[47] Padma Polash Paul, Marina L. Gavrilova, and Reda Alhajj. 2014. Decision Fusion for Multimodal Biometrics Using Social Network Analysis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 44, 11 (nov 2014), 1522–1533. https://doi.org/10.1109/TSMC.2014.2331920

[48] Sami Pietinen, Roman Bednarik, Tatiana Glotova, Vesa Tenhunen, and Markku Tukiainen. 2008. A method to study visual attention aspects of collaboration: Eye-tracking pair programmers simultaneously. *Eye Tracking Research and Applications Symposium (ETRA)* (2008), 39–42. https://doi.org/10.1145/1344471.1344480

[49] José J. Ramasco, S. N. Dorogovtsev, and Romualdo Pastor-Satorras. 2004. Self-organization of collaboration networks. *Physical Review E* 70, 3 (sep 2004), 036106. https://doi.org/10.1103/PhysRevE.70.036106 arXiv:0403438 [cond-mat]

[50] Fabian Ramseyer and Wolfgang Tschacher. 2006. Synchrony: A core concept for a constructivist approach to psychotherapy. *Constructivism in the human sciences* 11, 1 (2006), 150–171. http://www.researchgate.net/publication/215507443_Synchrony_A_Core_Concept_for_a_Constructivist_Approach_to_Psychotherapy/file/3606d2cadc7a8d399c757cbb48c1e8ec.pdf

[51] Pramila Rani, Nilanjan Sarkar, Craig A. Smith, and Leslie D. Kirby. 2004. Anxiety detecting robotic system – towards implicit human-robot collaboration. *Robotica* 22, 1 (jan 2004), 85–95. https://doi.org/10.1017/S0263574703005319

[52] Michael J. Richardson, Randi L. Garcia, Till D. Frank, Madison Gergor, and Kerry L. Marsh. 2012. Measuring group synchrony: A cluster-phase method for analyzing multivariate movement time-series. *Frontiers in Physiology* 3 OCT, October (2012), 1–10. https://doi.org/10.3389/fphys.2012.00405

[53] Timothy J Rowley. 1997. Moving Beyond Dyadic Ties: A Network Theory of Stakeholder Influences. *Academy of Management Review* 22, 4 (oct 1997), 887–910. https://doi.org/10.5465/amr.1997.9711022107

[54] A.R. Ruis, Alexandra A. Rosser, Cheyenne Quandt-Walle, Jay N. Nathwani, David Williamson Shaffer, and Carla M. Pugh. 2018. The hands and head of a surgeon: Modeling operative competency with multimodal epistemic network analysis. *The American Journal of Surgery* 216, 5 (nov 2018), 835–840. https://doi.org/10.1016/j.amjsurg.2017.11.027

[55] David A. Sears and James Michael Reagin. 2013. Individual versus collaborative problem solving: Divergent outcomes depending on task complexity. *Instructional Science* 41, 6 (2013), 1153–1172. https://doi.org/10.1007/s11251-013-9271-8

[56] V Shute, R Almond, and S Rahimi. 2019. *Physics Playground (1.3)[Computer software]*.

[57] Angela E.B. Stewart, Mary Jean Amon, Nicholas D. Duran, and Sidney K. D'Mello. 2020. Beyond Team Makeup: Diversity in Teams Predicts Valued Outcomes in Computer-Mediated Collaborations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376279

[58] Shree Krishna Subburaj, Angela E.B. Stewart, Arjun Ramesh Rao, and Sidney K. D'Mello. 2020. Multimodal, Multiparty Modeling of Collaborative Problem Solving Performance. In *Proceedings of the 2020 International Conference on Multimodal Interaction*. ACM, New York, NY, USA, 423–432. https://doi.org/10.1145/3382507.3418877

[59] Mattia Tantardini, Francesca Ieva, Lucia Tajoli, and Carlo Piccardi. 2019. Comparing methods for comparing networks. *Scientific Reports* 9, 1 (dec 2019), 17557. https://doi.org/10.1038/s41598-019-53708-y

[60] Zhulin Tao, Yinwei Wei, Xiang Wang, Xiangnan He, Xianglin Huang, and Tat-Seng Chua. 2020. MGAT: Multimodal Graph Attention Network for Recommendation. *Information Processing & Management* 57, 5 (sep 2020), 102277. https://doi.org/10.1016/j.ipm.2020.102277

[61] Charlotte Vrijen, Catharina A. Hartman, Eeske van Roekel, Peter de Jonge, and Albertine J. Oldehinkel. 2018. Spread the Joy: How High and Low Bias for Happy Facial Emotions Translate into Different Daily Life Affect Dynamics. *Complexity* 2018 (dec 2018), 1–15. https://doi.org/10.1155/2018/2674523

[62] Hana Vrzakova, Mary Jean Amon, Angela Stewart, Nicholas D. Duran, and Sidney K. D'Mello. 2020. Focused or stuck together. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. ACM, New York, NY, USA, 295–304. https://doi.org/10.1145/3375462.3375467