Your Actions Talk: Automated Sociometric Analysis Using Kinesics in Human Activities

Cheyu Lin
cheyul@andrew.cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Maral Doctorarastoo
mdoctora@andrew.cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Katherine Flanigan*
kflaniga@andrew.cmu.edu
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA

Abstract

Cyber-physical-social infrastructure systems (CPSIS) are an extension of cyber-physical systems (CPS). In addition to sensing, measuring, interpreting, and optimizing physical attributes of the built environment to improve infrastructure performance, CPSIS also takes into account human-centered—or social—objectives often overlooked by CPS. Although this paradigm shift aims to incorporate the social system supported by infrastructure into CPS, there is still a gap in measuring social objectives in line with the guiding principle of CPSIS. Specifically, the integration of sensing technologies and computation for assessing these social objectives remains largely unaddressed. As a salient example, sociometric tests used ubiquitously to capture the social structure and sociability embedded within a group of individuals still relies on subjects manually answering questionnaires to derive social connectivity. This data collection scheme is, among other things, subject to attribution bias, inefficient, and laborious. Here, reliance on manually-sourced data to inform sociometric tests falls short in leveraging the sensing and automation capabilities inherent in CPSIS. To overcome these challenges, we propose a human activity recognition (HAR) dataset and framework that can help to automate the procedure of sociometric assessment. The design of the dataset takes into account the limitations that hinders the development of the automated sociometric examination in state-of-the-art HAR techniques. The framework adopts a multidisciplinary approach, drawing upon HAR, kinesics, and sociology to efficiently distill the interpersonal relationships within social systems and provide a qualitative and quantitative interpretation of sociability for modeling and optimization in the context of CPSIS.

CCS Concepts

• **Human-centered computing → Social network analysis**; *Social networks.*

Keywords

Cyber-physical-social Infrastructure Systems, Human Activity Recognition, Sociometry, Social Networks.

*Corresponding author.

BUILDSYS '24, November 7–8, 2024, Hangzhou, China

ACM ISBN 979-8-4007-0706-3/24/11
https://doi.org/10.1145/3671127.3699529

ACM Reference Format:
Cheyu Lin, Maral Doctorarastoo, and Katherine Flanigan. 2024. Your Actions Talk: Automated Sociometric Analysis Using Kinesics in Human Activities. In *The 11th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BUILDSYS '24), November 7–8, 2024, Hangzhou, China.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3671127.3699529

1 Introduction

Cyber-physical-social infrastructure systems (CPSIS) have recently evolved as an extension of traditional cyber-physical systems (CPS). While CPS focuses on optimizing infrastructure performance (e.g., efficiency and safety) through sensing, computation, and control, CPSIS also considers how the social system interacts with and benefits from the built environment [3]. Specifically, CPSIS introduces a paradigm that integrates human-centered, or "social," objectives into CPS, measuring, interpreting, and optimizing factors like sociability and productivity for those using the infrastructure, as illustrated in Figure 1. A human-centered framework like CPSIS offers significant advantages for applications such as autonomous vehicles, smart homes, elder care management, and many more.

However, despite its potential across various sectors, a significant gap remains in how to measure and interpret these social objectives in CPSIS [2]. While physical attributes can be continuously sensed and optimized through various technologies, social objectives—such as sociability, cohesion, and productivity—are often less tangible and harder to quantify. Addressing these social aspects of CPSIS requires a more sophisticated approach to measuring and modeling human interactions within the system. Specifically, there is a need to integrate social science methodologies, like sociometric assessments, with sensing technologies to fully capture and respond to the social dynamics within human environments.

Sociability, a key social objective, offers a valuable pathway to understanding group interactions [4]. It refers to the attractions and repulsions within a group of individuals, providing both qualitative and quantitative insights into social dynamics [20]. To systematically analyze and measure sociability, Moreno (1941) developed a sociometric test to uncover the underlying social structure and sociability within groups [20]. Individuals in groups are asked to select others based on specific criteria, such as friendship or collaboration. These tests reveal patterns of attraction, repulsion, social capital, and isolation within groups, helping to map the flow of social interactions. However, sociometric tests have traditionally relied on self-reported data and interviews, which are subject to attribution bias [11] and inefficiency [26]. This reliance on manual data collection hampers their potential in CPSIS, where automated, continuous sensing of both physical and social systems is crucial for near real-time optimization.

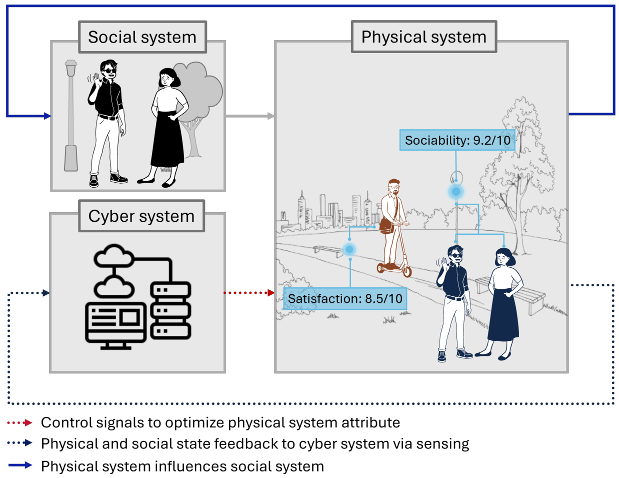

Figure 1: CPSIS aims to interpret the reciprocative influence physical and social systems have on each other at the *behavioral* level. It is distinct from interaction on physiological levels—where the CPS interacts with bio-signals—and also distinct from interaction on population levels—where the focus in on large-scale interactions between netowrked human populations and CPS.

The benefits of automating sociometric tests within CPSIS applications are far-reaching. For example, in education, understanding peer dynamics through sociometric assessments can inform the design of collaborative learning environments, optimizing classroom layouts and peer groupings to enhance both academic performance and social integration [19]. In team-based workspaces, automated sociometric analysis can be used to monitor and enhance team productivity by identifying patterns of collaboration and social cohesion, helping managers foster more effective group dynamics [9]. In public open spaces, tracking sociability through automated methods can offer insights into how different designs or spatial arrangements affect social interaction, leading to better urban planning that promotes community engagement [7]. Additionally, sociometric assessments have the potential to significantly enhance social cohesion in communities. For instance, urban planners could use these insights to design infrastructure that strengthens community ties by facilitating social interactions in public spaces [7]. Likewise, in elder care settings, automated sociometric tests could help identify individuals at risk of social isolation, enabling interventions to maintain social engagement and prevent cognitive decline [5]. By systematically measuring sociability rooted in sociometric analysis, CPSIS can improve both the design and functionality of infrastructure, promoting the well-being of its users.

The current reliance on manually administered sociometric tests, however, falls short in fully leveraging the sensing and automation capabilities of CPSIS. These tests require respondents to reflect on their preferences and relationships, a process prone to attribution bias, as individuals often misjudge the rationale behind their decisions [27]. Additionally, repeated assessments are time-consuming, limiting the practicality of sociometric tests in iterative contexts, such as education or work environments [26]. To overcome these challenges, we propose an approach that automates sociometric assessments by using nonverbal communication and human activity recognition (HAR) to extract social preferences and relationships more accurately and efficiently.

There are many channels to express positive, negative, and neutral emotions, including verbal, nonverbal, visual, and written communications [10]. Nonverbal communication, particularly kinesics—the study of body language—plays a crucial role in human interaction, revealing subconscious cues about the nature of relationships [25]. Kinesics allows for exploration of the disposition embedded in the bodily movements of human interactions—whether the interaction is positive, negative, or neutral. Coupled with HAR, which studies the spatial and temporal coordination of human body movements, we propose the automated extraction of these social cues, providing real-time sociometric data. This approach not only bypasses the attribution bias inherent in self-reported data [22] but also allows for continuous and unobtrusive monitoring of social dynamics, making it ideal for not only CPSIS applications, but also other sociological and psychological examinations. A key contribution of this work lies in the creation of a new dataset, DUET, which supports dyadic HAR and is grounded in a sociability taxonomy. The novel adaptation of the sociability taxonomy connects HAR, kinesics, psychology, and sociology, creating a multidisciplinary bridge. This integration stands to enhance the performance of HAR in inferring dyadic interactions, which is essential for automating sociometric assessments and supporting various sociological and psychological studies within CPSIS and other relevant fields.

The remainder of the paper is organized as follows. First, in Section 2 we discuss in more depth the process of manually-sourced sociometric examination in order to better understand the advantages and disadvantages of this current state-of-the-art process. The identified limitations are used to justify the proposed methodology for automated sociometric analysis. Next, Section 3 delineates the prerequisites for the automation of sociometric analysis, identifies limitations of state-of-the-art technologies and HAR algorithms for dyadic interaction classification needed to support sociometric tests, and presents a new dataset well-suited to overcome these challenges. The procedure for an automated sociometric test that builds upon dyadic HAR is laid out in Section 4. Lastly, Section 5 presents key conclusions of this work and discusses a roadmap for future development.

2 Sociometric Analysis and its Role in CPSIS

Sociometric testing is a well-established method for measuring sociability and understanding the social dynamics within a group of individuals [20]. It provides both qualitative and quantitative insights into social structure, revealing relationships of attraction, repulsion, and isolation. However, the traditional methodology behind sociometric testing and its limitations present challenges when integrated into modern infrastructure systems like CPSIS, which rely on near real-time data and automation. In this section, we describe the sociometric test methodology and explore its inherent limitations.

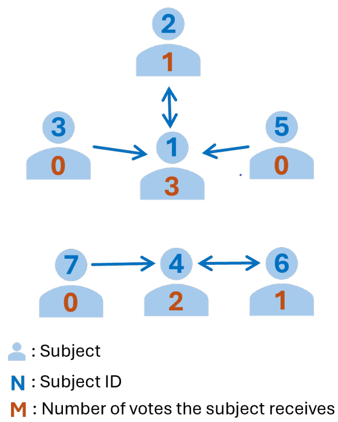

Figure 2: A simple example of a sociogram demonstrates that (1) Subject 1 has the most social wealth and (2) there are two fractions within the group.

2.1 Sociometric Analysis Methodology

Sociometric tests are designed to map the social structure of a group by capturing the social choices individuals make when responding to specific questions. These tests typically involve asking participants to reflect on their preferences for others within a group in hypothetical scenarios. For example, a question might ask, "Who would you like to go on a two-day trip with the most?" Based on the responses, a sociogram is created, which visually represents the social relationships and sociability within the group.

Sociograms provide valuable insights into social structures. For instance, they can reveal which individuals are central to the group's social dynamics (those with high social capital), those who are more isolated, and any subgroups or cliques that may exist. The practical applications of sociometric testing span various fields. Moreno (1941) famously used sociometric testing in penitentiaries to understand inmate relationships and to assign roles based on social capital. In education, sociometric tests are used to identify social dynamics within classrooms, allowing educators to adjust seating arrangements or groupings to foster better collaboration and social integration [19]. Sociometric tests are also employed in psychodrama group therapy, where they help assess the effectiveness of sessions by mapping changes in social relationships before and after therapy [6].

A simple example sociogram inferred from the question "who would you like to spend time with the most" is shown in Figure 2. In this example, there are seven subjects labeled with blue numbers (N), and arrows indicate each respondent's preference for the question asked. For instance, the arrow from Subject 5 to Subject 1 shows that Subject 5 prefers to spend time with Subject 1. A bidirectional arrow means that both subjects have chosen each other. The orange numbers (M) represent the number of votes each subject receives, also referred to as their social wealth or social capital. From this sociogram, we can draw three conclusions: (1) Subject 1 has the highest social capital in the group, (2) the group is divided into two subgroups, and (3) Subjects 3, 5, and 7 have no social capital, which may discourage them from participating in social activities.

These qualitative and quantitative insights derived from sociograms stand to provide valuable information in CPSIS applications. They could serve as a value input and help reduce uncertainty in data-driven agent-based models, allowing for adjustment of physical and environmental features in infrastructure systems to improve social interactions and overall sociability. For example, in an educational setting, the results of routinely updated sociometric tests could be used to dynamically modify the layout of classrooms to foster better peer engagement or support learners with special needs. Similarly, in team-based workspaces, sociometric feedback can be used to design office spaces that promote more effective collaboration by identifying individuals who may benefit from being placed together or apart. The versatility of sociometric testing offers great potential for improving human-centered infrastructure systems.

2.2 Sociometric Analysis Limitations

Despite the widespread use and benefits of sociometric tests described in Section 2.1, this traditional methodology presents several challenges, particularly in the context of modern, automated systems like CPSIS. One of the primary limitations is attribution bias, which affects the accuracy of the data collected. When individuals are asked to make choices regarding their social preferences, they may consciously or unconsciously misrepresent their true feelings. This bias occurs when respondents overanalyze the rationale behind their choices or feel pressured to respond in socially acceptable ways [27]. As a result, the data collected may not accurately reflect genuine social dynamics, reducing the reliability of sociometric tests in providing faithful insights into group interactions.

In addition to attribution bias, the traditional sociometric test is inefficient. Each time the test is administered, participants are required to reflect on their social preferences and respond to a series of questions. This process can be time-consuming and laborious, particularly in settings that require iterative assessments, such as education or workplace environments. Although efforts have been made to expedite the process, such as through the development of web applications for administering sociometric tests [26], these methods still require active participation and are limited by human response time.

These limitations hinder the potential of sociometric testing within CPSIS, where real-time data collection and automation are critical for optimizing infrastructure systems. Manually administered sociometric tests do not align with the continuous sensing and data-driven nature of CPSIS, where human preferences and social dynamics should be captured automatically to allow for timely interventions and adjustments.

To address these limitations, we propose a privacy-preserving approach that automates the sociometric assessment process by leveraging HAR and kinesics, the study of body language. Nonverbal communication, such as body movements and facial expressions, provides genuine, subconscious cues about social relationships that can bypass the attribution bias inherent in self-reported data [22]. By analyzing these cues, we can automatically infer social preferences and map social structures without requiring participants to manually respond to questions. This automated approach not only eliminates the inefficiency of traditional sociometric tests but also

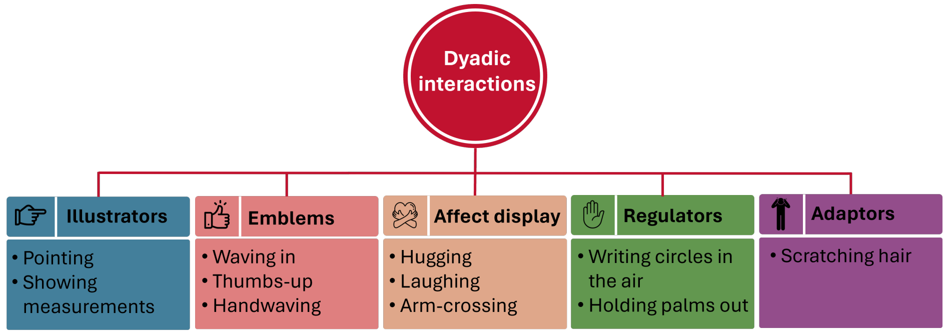

Figure 3: Taxonomy comprising five communication functions.

Table 1: A comparison of existing depth and 3D skeleton-based dyadic datasets demonstrates that the DUET dataset has the highest sample-class ratio, more locations than existing datasets, and is the only one that supports contextualization.

Dataset		Depth	Joints	#Videos	#Classes	#Videos/ #Classes	#Loca- tions	Contextu- alization	Year
UT Interaction	[23]	No	No	160	6	26.7	2	No	2010
SBU Kinect	[29]	Yes	Yes	300	8	37.5	1	No	2012
K3HI	[8]	Yes	No	320	6	53.3	1	No	2013
JPL Interaction	[24]	No	No	399	7	57	5	No	2013
G3Di	[1]	Yes	Yes	168	14	12	1	No	2015
M^2I	[17]	Yes	Yes	1,760	9	195.6	1	No	2015
ShakeFive 2	[28]	No	Yes	153	8	19.1	1	No	2016
NTU RGB+D 120	[16]	Yes	Yes	24,828	26	954.9	-	No	2019
Air Act2Act	[12]	Yes	Yes	500	10	500	2	No	2020
DUET (our dataset)	[14]	**Yes**	**Yes**	**14,400**	**12**	**1,200**	**3**	**Yes**	**2024**

provides nearl real-time insights into social dynamics, making it ideal for integration into CPSIS.

3 HAR for Sociometry

Sociometry distills the social structure embedded within a group by *directly* assessing the interpersonal attractions and repulsions within all subject pairs, emphasizing the need for accurate recognition of two-person (dyadic) interactions for automating sociometric evaluations. However, current progress in dyadic HAR remains inadequate to fully automate sociometry, which relies on the *indirect* inference of these attractions and repulsions.

3.1 State-of-the-art Dyadic HAR

HAR is a branch of artificial intelligence that has achieved significant success in various fields, largely due to factors such as the availability of publicly available datasets that aid in refining data-driven deep learning algorithms. However, while numerous datasets exist, most focus on single-person—or monadic—activities, or single-person human-infrastructure interaction [18]. This disproportionate focus on monadic HAR overlooks the greater complexity involved in the spatial and temporal coordination of dyadic interactions. Lin *et al.* [15] demonstrated that monadic algorithms that have outstanding benchmarking records for single-person activities do not perform nearly as well for dyadic interactions, highlighting the disparity between monadic and dyadic activities. To address this gap—improving the performance of dyadic HAR such that it can support sociometric analysis—there is a need for more datasets specifically designed for dyadic interactions to improve HAR in these contexts.

In addition to increasing the number, quality, and diversity of dyadic datasets for more accurate recognition of bodily movements, automating sociometric assessment requires interpreting dyadic interactions at a higher level of abstraction. Specifically, it must account for the disposition of human interactions—whether they are positive, negative, or neutral—conveyed through the nonverbal communication. Nonverbal communication, particularly kinesics (the study of body language) reveals subconscious inclinations expressed in bodily movements, enabling us to explore the nature of these interactions. By combining kinesics with HAR, we can lay the foundation for automatically extracting subjects' social preferences and generating generating near real-time sociometric data. However, most existing dyadic datasets—such as those listed in Table 1—are insufficient for analyzing kinesics. Some of these datasets focus on healthcare activities, while others merely track bodily movements in specific contexts. To more accurately identify human activities and interpret their dispositions in dyadic settings, a dyadic HAR dataset that supports the analysis of social embeddings of human interactions is needed.

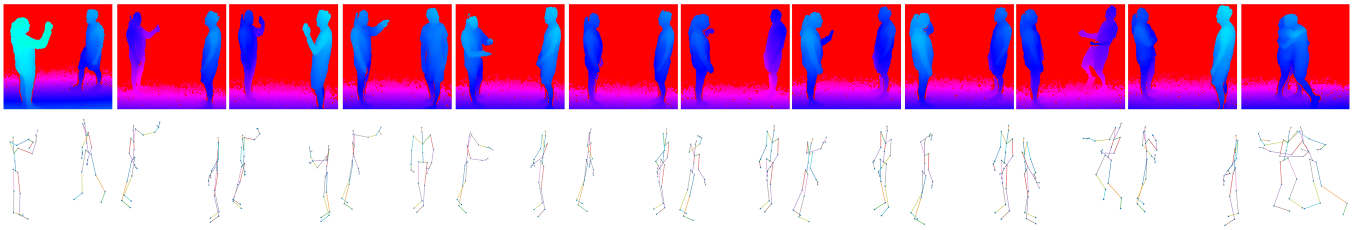

Figure 4: Sample frames of 12 interactions from DUET. The 12 interactions provided are, from left to right: waving in, thumbs up, waving, pointing, showing measurements, nodding, drawing circles in the air, holding palms out, twirling or scratching hair, laughing, arm crossing, and hugging. The modalities are organized in the order of, from top row to bottom row, depth and 3D skeleton joints.

3.2 Privacy Concerns of Data-Driven Sociometry

In human-centered applications like CPSIS, preserving user privacy is paramount, as it helps build trust between stakeholders and users. Modalities commonly used for HAR are RGB, depth (which is used to extract 3D skeleton joints from subjects), and infrared (IR) (Table 1). These modalities each have their own strengths and weaknesses. RGB data captures detailed information in the frame ranging from interactions to locations to characteristic features of subjects, but this richness of detail compromises user privacy. IR data, while less descriptive than RGB, is useful for capturing subjects in low-light conditions and for identifying basic shapes and movements without revealing detailed personal characteristics, offering a middle ground between information value and privacy protection. 3D skeleton joints, on the other hand, represent subjects as nodes and edges in 3D space, stripping away identifying features indicative of the subjects. This sparse representation prioritizes user privacy. Sensors that generate less detailed information than depth sensors, such as passive infrared sensors, typically do not provide enough data to accurately identify and classify the social meaning of interactions [13]. This trade-off between information richness and privacy, as shown in Figure 4, illustrates that the more information is captured, the less privacy is preserved.

To address privacy concerns while supporting future research efforts, our dataset utilizes the high-quality Azure Kinect v2, which offers depth and 3D skeleton joint modalities. Our primary focus for sociometric analysis—and what we discuss in this paper—is on depth-based skeleton data. This modality provides an ideal balance between capturing meaningful interaction data and preserving user privacy, which is essential in social settings. The 3D skeleton joints represent subjects in an abstract form, devoid of identifiable characteristics, making it crucial for sociometric analysis in privacy-sensitive environments such as CPSIS. By focusing on depth and skeleton-based sensing, we ensure that individual identities are protected while still allowing for accurate analysis of social interactions. This approach allows researchers to conduct sociometric assessments and other analyses in privacy-critical contexts without compromising data quality, while still enabling future exploration of multimodal techniques as needed.

3.3 Dyadic User EngagementT (DUET)

To advance the development of dyadic HAR and, by extension, the automation of sociometric assessments, we created a dyadic HAR dataset—Dyadic User EngagementT (DUET) [14]. DUET includes 12 dyadic activities, as shown in Figure 3, which are adapted from a psychological classification system that organizes human interactions into five core communication functions. This integration of psychological taxonomy allows for the contextualization of human activities, revealing the social semantics embedded in bodily movements and adding an additional layer of information to HAR tasks. Contextualization has not only proven effective in enhancing the performance of HAR tasks [21], but also provides a solid foundation for systematically and efficiently classifying the disposition of interactions. The five categories—emblems, illustrators, affect displays, regulators, and adaptors—are closely tied to sociometric assessment as they capture essential nonverbal cues that reveal the nature of interpersonal interactions. Below, we describe how each category contributes to sociometric assessments.

- Emblems: Emblems are gestures with direct verbal translations, often culturally specific (e.g., a thumbs-up). In sociometric assessment, emblems can help assess clear, intentional communication signals between individuals, indicating preferences, agreements, or dislikes. These overt gestures are critical for identifying explicit social dynamics and relationships within groups.
- Illustrators: Illustrators are gestures used to clarify or emphasize verbal communication. In the context of sociometric assessment, illustrators can reveal the level of engagement and emphasis individuals place on their interactions. The frequency or intensity of illustrators may signal how much importance a person places on a relationship, thus offering insight into social cohesion or dominance within a group.
- Affect displays: Affect displays are nonverbal expressions of emotions, such as smiling or frowning. These are key to assessing emotional connections, as they provide subconscious signals about attraction, repulsion, or neutrality in social relationships. Understanding affect displays allows sociometric assessments to gauge emotional dynamics and social bonds within a group.
- Regulators: Regulators are nonverbal cues that govern the flow of conversation, such as nodding to indicate listening or

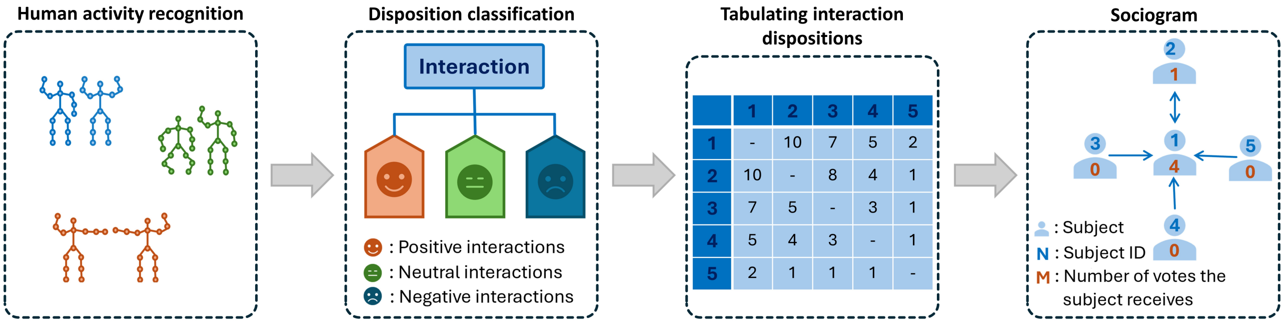

Figure 5: The proposed structure of the automated sociometric examination.

signaling a desire to speak. These cues are crucial for sociometric assessments because they reveal interaction patterns, such as who leads conversations, who defers to others, and how individuals influence group dynamics. This is particularly important for understanding group hierarchies and social influence.

- Adaptors: Adaptors are gestures typically related to self-comfort or emotional regulation, such as scratching one's head or fidgeting. These can provide insight into an individual's emotional state during interactions and may indicate stress, discomfort, or a lack of confidence in specific social contexts. Adaptors can help sociometric assessments by identifying underlying social tensions or insecurities that affect group cohesion.

3.4 Data Collection

DUET is collected at three locations representative of the potential venues for CPSIS at Carnegie Mellon University (Pittsburgh, PA), including a confined indoor space, an open indoor space, and an open outdoor space. At each location, 20 consenting volunteers participated in the data collection process, and were randomly assigned in pairs to perform interactions. In total, 15 males and eight females participated in the experiment. Their ages range from 23 to 42 years old with a mean of 27 years old and standard deviation of 4.01 years. The subject heights span from 165.1cm to 185.4cm with a mean of 172.7cm and standard deviation of 8.46cm. Their weights range from 55kg to 93kg with a mean of 69kg and standard deviation of 10.1kg. With the help from these subjects across three locations, the DUET dataset comprises 14,400 samples across 12 activity classes, yielding the highest sample-class ratio known to date.

4 Automated Sociometric Analysis

The automation of sociometry builds on the research and technologies discussed in Section 3. The first step in automating sociometric assessments is to employ HAR to identify all human interactions within the group, particularly focusing on dyadic interactions. HAR is used to capture the spatial and temporal coordination of body movements, which are essential for understanding the interpersonal relationships that define the social structure of a group. By leveraging HAR as the foundational tool, we can bypass the time-consuming and biased nature of manually-administered sociometric tests, allowing for efficient and objective tracking of human interactions.

Once HAR has captured the necessary interactions, we leverage the established taxonomy of emblems, illustrators, affect displays, regulators, and adaptors to classify the interactions. Each of these nonverbal behaviors is tied to either positive, neutral, or negative dispositions, providing a deeper understanding of the nature of the social exchanges. For example, positive emblems and affect displays often indicate supportive interactions, while neutral adaptors or illustrators may reflect more routine or passive exchanges. Regulators can signal dominance or cooperation, allowing us to infer whether the interaction was collaborative or potentially conflictual.

Using this taxonomy, HAR identifies and categorizes interactions, assigning dispositions to each. Once the interactions are classified as positive, neutral, or negative, points are allocated accordingly: one point for positive interactions, zero for neutral, and negative one for negative interactions. The sum of these points for each subject pair is calculated, representing their social preferences and interaction tendencies. This quantified data is then tabulated to create a comprehensive overview of the social dynamics in the group. As shown in Figure 5, this table captures the cumulative interaction scores between each pair, allowing the generation of a sociogram that visually maps the relationships and social structures embedded within the group. For the question used to derive the sample sociogram in Figure 2, for instance, Subject 1 prefers to spend time with Subject 2 since Subject 2 scores the highest among all subjects in the first row of the table in Figure 5. The synergistic use of HAR, kinesics, and sociology allows us to remove the attribution bias and time inefficiency intrinsic in manually-sourced sociometric tests.

The use of HAR to classify nonverbal cues not only enhances the accuracy of the sociometric analysis but also provides near real-time insights into the evolving dynamics within the group. This approach eliminates the need for self-reported data, thereby reducing attribution bias and inefficiencies. The automated sociometric framework thus aligns seamlessly with the principles of CPSIS, integrating sensing technologies to deliver continuous, real-time assessments of social interactions. By focusing on the automated recognition and classification of emblems, illustrators, affect displays, regulators, and adaptors, the system offers a robust tool for understanding and optimizing social dynamics in privacy-sensitive environments.

5 Discussion and Conclusion

In this work, we present an automated sociometric approach that extracts the embedded social structure and sociability within groups of individuals. The proposed framework is interdisciplinary, integrating HAR, kinesics, psychology, and sociology to address the attribution bias and time inefficiencies common in manually-sourced sociometric assessments. This approach aligns with the sensing capabilities of CPSIS and creates a strong connection between the sensing and modeling components within CPSIS.

Another key contribution of this work is the dataset, Dyadic User EngagemenT (DUET), presented. DUET incoporates a psychological taxonomy that identifies five fundamental human communication functions. In addition to improving the performance of HAR tasks, this novel integration provides a strong foundation for disposition classification, which is essential for automating sociometric tests. Additionally, the dataset includes multiple modalities (e.g., depth, 3D skeleton joints) that not only support the exploration of individual modalities but also aim to meet the privacy-preserving requirements of CPSIS by combining less detailed modalities to enhance their informational value.

Future developments based on this work can be broadly categorized into three areas: verification of the proposed framework, automatic disposition classification, and the automation of other social objectives. The first area requires comprehensive experimentation, incorporating both manually-sourced and automated sociometric evaluations, to validate the framework and quantify the discrepancy between the two approaches. Once the reliability of the approach is empirically confirmed, the process stands to be streamlined and generalized through disposition classification. Specifically, artificial intelligence can be used to explore the underlying patterns in human interaction data and classify their corresponding dispositions. This research will not only be applied to the DUET dataset but also extended to other interactions. The expansion of the disposition classification will improve the universitality and practicality of this framework. Expanding the disposition classification will enhance the framework's universality and practical applicability. Additionally, this framework paves the way for extracting other social objectives, such as satisfaction and cohesion, and for automating a wider range of sociological and psychological assessments.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant Number 2425121.

References

[1] Victoria Bloom, Vasileios Argyriou, and Dimitrios Makris. 2016. Hierarchical transfer learning for online recognition of compound actions. *Comput Vis Image Underst* 144 (2016), 62–72.

[2] Maral Doctorarastoo, Katherine Flanigan, Mario Bergés, and Christopher McComb. 2023. Exploring the potentials and challenges of cyber-physical-social infrastructure systems for achieving human-centered objectives. In *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation.* November 15-16 2023, Istanbul, Turkey, ACM Inc., New York, NY, 385–389. https://doi.org/10.1145/3600100.3626340

[3] Maral Doctorarastoo, Katherine Flanigan, Mario Bergés, and Christopher McComb. 2023. Modeling human behavior in cyber-physical-social infrastructure systems. In *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation.* November 15-16 2023, Istanbul, Turkey, ACM Inc., New York, NY, 370–376. https://doi.org/10.1145/3600100.3626338

[4] Maral Doctorarastoo, Cheyu Lin, Katherine Flanigan, and Mario Bergés. 2024. A Bayesian network approach for predicting social interactions in shared spatial environments. In *Proceedings of the 11th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation.* November 7-8, 2024, Hangzhou, China, ACM Inc., New York, NY. https://doi.org/10.1145/3671127.3699530

[5] Janet Everly, Jamie Plummer, Matthew Lohman, and Jean Neils-Strunjas. 2023. A tutorial for speech-language pathologists: Physical activity and social engagement to prevent or slow cognitive decline in older adults. *American Journal of Speech-Language Pathology* 32, 1 (2023), 83–95.

[6] Scott Giacomucci. 2021. *Social Work, Sociometry, and Psychodrama: Experiential Approaches for Group Therapists, Community Leaders, and Social Workers.* Springer, New York, NY.

[7] Simon Gurnsey. 2018. Developing community through collective projects and activities: The benefits of using sociometry. *Australian and Aotearoa New Zealand Psychodrama Association Journal* 27 (2018), 63–75.

[8] Tao Hu, Xinyan Zhu, Wei Guo, and Kehua Su. 2013. Efficient interaction recognition through positive action representation. *Math. Probl. Eng.* 2013, 1 (2013).

[9] Diana Jones. 2001. Sociometry in team and organisation development. *British Journal of Psychodrama and Sociodrama* 16, 1 (2001), 10.

[10] Ray T Kest. 2006. Principles of leadership: Leadership management. *Futurics* 30 (2006), 52–71.

[11] Taemie Kim, Erin McFee, Daniel Olguin Olguin, Ben Waber, and Alex Sandy Pentland. 2012. Sociometric badges: Using sensor technology to capture new forms of collaboration. *Journal of Organizational Behavior* 33, 3 (2012), 412–427.

[12] Woo-Ri Ko, Minsu Jang, Jaeyeon Lee, and Jaehong Kim. 2021. AIR-Act2Act: Human–human interaction dataset for teaching non-verbal social behaviors to robots. *Int. J. Rob. Res.* 40, 4-5 (2021), 691–697.

[13] Cheyu Lin and Katherine Flanigan. 2023. Human trajectory estimation using analog privacy-preserving urban sensing technologies. In *Proc. SPIE 12486, Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems.* April 18 2023, Long Beach, California, USA, 1248604.

[14] Cheyu Lin and Katherine Flanigan. 2024. DUET Repository. https://huggingface.co/datasets/saluslab/DUET.

[15] Cheyu Lin, John Martins, and Katherine A Flanigan. 2024. Read the room: Inferring social context through dyadic interaction recognition in cyber-physical-social infrastructure systems. In *ASCE International Conference on Computing in Civil Engineering 2024 (i3ce2024).* July 28-31 2024, Pittsburgh, PA., ASCE.

[16] Jun Liu, Amir Shahroudy, Mauricio Perez, Gang Wang, Ling-Yu Duan, and Alex C Kot. 2019. Ntu rgb+ d 120: A large-scale benchmark for 3d human activity understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42, 10 (2019), 2684–2701.

[17] Tingting Liu, Zengzhao Chen, Hai Liu, Zhaoli Zhang, and Yingying Chen. 2018. Multi-modal hand gesture designing in multi-screen touchable teaching system for human-computer interaction. In *Proceedings of the 2nd International Conference on Advances in Image Processing (ICAIP '18).* ACM Inc., New York, NY, USA, 198–202.

[18] John Martins, Katherine Flanigan, and Christopher McComb. 2023. Skeleton-based human action recognition in a thermal comfort context. In *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation.* November 15-16 2023, Istanbul, Turkey, ACM Inc., New York, NY, 377–384. https://doi.org/10.1145/3600100.3626339

[19] Luis Miguel Mazón, Datzania Villao, Teresa Guarda, Linda Núñez, María Muñoz, Manuel Serrano, and Divar Castro. 2018. Sociometry: A University Tool to Facilitate the Cohesion of Academic Groups. In *Developments and Advances in Defense and Security.* April 18-20 2018, Salinas, Ecuador, Springer, New York, NY, 351–359.

[20] Jacob Levy Moreno. 1941. Foundations of sociometry: An introduction. *Sociometry* 4, 1 (1941), 15–35.

[21] Friedrich Niemann, Stefan Lüdtke, Christian Bartelt, and Michael Ten Hompel. 2021. Context-aware human activity recognition in industrial processes. *Sensors Technology and Machine Learning for Human Activity Recognition* 22, 1 (2021), 134.

[22] Allan Pease. 1994. *Body language.* Camel Publishing Company, Kenmore, Washington.

[23] Michael S Ryoo, Chia-Chih Chen, JK Aggarwal, and Amit Roy-Chowdhury. 2010. An overview of contest on semantic description of human activities (sdha) 2010. In *Recognizing Patterns in Signals, Speech, Images and Videos (ICPR 2010).* August 23-26 2010, Istanbul, Turkey, Spirnger, New York, NY, 270–285.

[24] Michael S Ryoo and Larry Matthies. 2013. First-person activity recognition: What are they doing to me?. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* June 23-28, Portland, Oregon, IEEE, New York, NY, 2730–2737.

[25] Navya N Sharan, Alexander Toet, Tina Mioch, Omar Niamut, and Jan BF van Erp. 2021. The relative importance of social cues in immersive mediated communication. In *Human Interaction, Emerging Technologies and Future Systems V:*

Proceedings of the 5th International Virtual Conference on Human Interaction and Emerging Technologies, IHIET 2021. August 27-29 2021, France, Springer, New York, NY, 491–498.

[26] Chutima Suraseth and Prakob Koraneekij. 2022. Development of a sociometric web application to study the relationship among secondary school learners. *Heliyon* 8, 8 (2022).

[27] Philip E Tetlock and Ariel Levi. 1982. Attribution bias: On the inconclusiveness of the cognition-motivation debate. *Journal of Experimental Social Psychology* 18, 1 (1982), 68–88.

[28] Coert Van Gemeren, Ronald Poppe, and Remco C Veltkamp. 2016. Spatio-temporal detection of fine-grained dyadic human interactions. In *Human Behavior Understanding: 7th International Workshop, HBU 2016*. October, 16, 2016, Amsterdam, The Netherlands, Springer, New York, NY, 116–133.

[29] Kiwon Yun, Jean Honorio, Debaleena Chattopadhyay, Tamara L Berg, and Dimitris Samaras. 2012. Two-person interaction detection using body-pose features and multiple instance learning. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. June 16-21 2012, Providence, RI, IEEE, New York, NY, 28–35.