Persistent Homology for Capturing Social Structure and Cohesion of F-Formation Groups

Valerie K. Chen* and Claire Liang*

***Abstract*— In social gatherings, people naturally form groups for social interaction. For robots to join and interact with these groups, they need the ability to reason about spatial representations of social configurations. In this paper, we define an abstraction using a 6-dimensional geometric structure towards capturing "stability" – the idea that a group is likely to remain a group – for the entire room. We apply persistent homology, from the field of topological data analysis, to formally assess these structures and provide a quantitative measure of social cohesion. We analyze two case studies for differently-sized groups and show how this 6-D structure can be used to identify and understand levels of cohesion for spatial social groups.**

I. Introduction

Despite great progress in social robot navigation, robots still struggle to operate in dynamic environments where many different groups of people are present [9]. Imagine the "cocktail party" scenario, in which free-standing conversational groups of people organically form, change, and separate over time. A socially-fluent robot would be able to not only anticipate human motion, but also gauge and preserve stability of social grouping while moving around. A current state-of-the-art autonomous robot may be able to weave between people, respecting the individual trajectories of those in motion, but it would not be able to fluidly enter and leave a conversation. A robot needs to be able to gauge the entire room at once and reason about intra-group relationships to do so. With the social fluency to gauge groups, a robot could potentially steer the evolution of social groups, maintain social cohesion, and encourage people into formations conducive to the room's social balance.

We introduce a mathematical structure to capture the spatial configurations of social groups. Given poses of people in a scene, we use techniques from persistent homology to deduce social relationships and rank their cohesion.
The contributions of this work include:

- An abstraction capturing social group cohesion using line of sight, which exposes shared social spaces; and
- A method using persistent homology from the field of topological data analysis to formally analyze and provide a quantitative measure for group social cohesion.

II. Related Work

A. Social Groups: F-Formations and O-Spaces

We build off of Kendon's sociological theory and conceptualization of the F-formation, a social group representation, which "arises whenever two or more people sustain a spatial and orientational relationship in which the space between them is one to which they have equal, direct, and exclusive access" [7]. The "shared inner-space" between those in the F-formation is called the o-space, "surrounded by a narrower one, here called the p-space, which provides for the placement of the participant's bodies" [6]. The remaining space of the room is the r-space, as illustrated in Fig. 1.

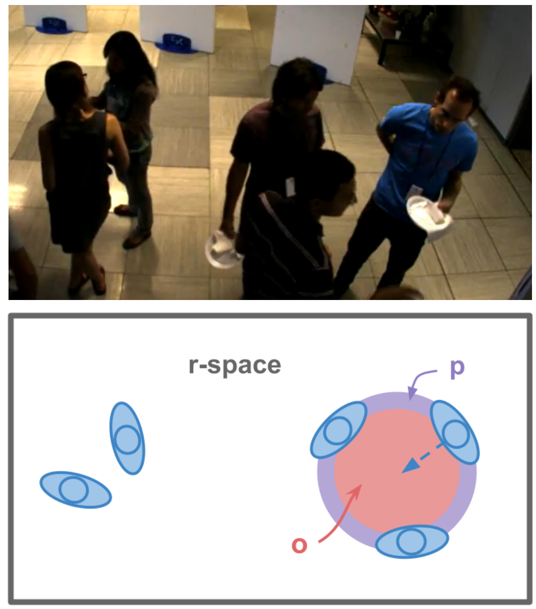

Fig. 1: An example of a three person F-formation alongside a two pair as found in the SALSA dataset [1]. The upper figure shows the camera image, and the lower figure shows the labeled configuration: people are represented in blue, p-space in purple, o-space in pink, and r-space in white.

Since Kendon's original work, a variety of technical methods have employed the concept of F-formations. These include detection algorithms for classifying which agents belong to a single F-formation using techniques from graph methods and voting strategies [5], [12], [13], as well as learned vision approaches [2]. A variety of techniques use other sensing input as well [4], [14], [18]. However, these methods largely focus on detecting the F-formations and classifying each agent's group membership.

Robot navigation in specific applications, such as for mobile museum robots [17] and information presenting robots [16], has also incorporated reasoning about F-formations. Kuzuoka et al. [7] explore how a robot could actively reconfigure human F-formations through changes in its own

*Authors contributed equally. Valerie K. Chen and Claire Liang are with the Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA
vkchen@csail.mit.edu and cyl48@mit.edu

orientation.

Previous work primarily focuses on reasoning about each social group individually. However, social groups in the same space do not operate independently of each other; groups may merge or undergo changes in membership.

Work such as [11] and [15] use learned approaches to investigate the formation and splitting of social groups. We contribute a mathematical abstraction that, unlike these prior works, does not require training data. This abstraction also permits desirable mathematical properties, such as equivalence classes that equate social topologies between scenes.

B. Topological Data Analysis: Persistent Homology

It can be difficult for a robot to reason about social information because the data easily available to the robot is geometric. Meanwhile, humans often reason about social structures using topological representations [8].

In applied mathematics, *topological data analysis (TDA)* is used to capture underlying topological structures from data. A primary component of TDA is *persistent homology*.

Consider a set of points in space. To capture connectivity between points, a ball with radius r is grown around each point, and an (n-1)-simplex is formed when n points' balls connect. For example, two connected points form an edge, and three edges form a face, as shown in Fig. 2. Connected elements form a simplex, and together, the simplices form a complex.

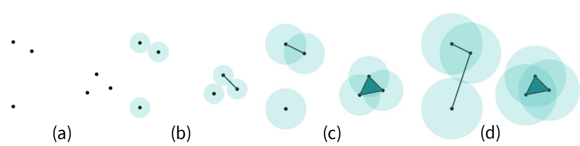

Fig. 2: An example of increasing the radius of balls to determine connectivity. An edge (1-simplex) is formed in (b), and a face (2-simplex) is formed in (c).

Edges and faces form as r does, possibly subsuming others in the process. As a result, we can categorize complexes and their subcomplexes in sequences called *filtrations*. A filtration is a sequence of nested subsimplicial complexes $\mathcal{F} = \mathcal{F}(K)$ of a complex K. Given a filtration function encoding relative distance, each of these complexes is assigned a *filtration value* [3], [9].

In our work, we leverage persistent homology and filtration values to understand the underlying social structures present given geometric positioning of individuals.

III. Methodology

Our goals are to understand geometric spatial information, deduce topological and social relationships, and produce a corresponding mathematical structure suitable for computational methods. We seek to interpret social group information from a set configuration of people.

We first take as input a set of people's positions and orientations. We introduce an abstraction to leverage eye gaze relationships between pairs of people (Section III-A) and create a vector representation to allow computation of a distance metric (Section III-B). Using this metric, we use *persistent homology* and generate alpha complexes (Section III-C) to produce filtration values, allowing for identification and ranking of social cohesion.

A. Social Relationships of Pairs: Pairwise Triangles

Social interaction requires more than one participant; we cannot capture underlying social structure by independently considering each person's position and orientation. Members' relative poses in a group are indicators of cohesion.

Let a represent a person in a scene, defined as a tuple $a = (q, \theta)$, where $q \in \mathbb{R}^2$ represents position, and $\theta \in [0, 2\pi)$ represents orientation. Person a's line of sight is the ray originating from q with angle θ. We first define pairwise geometric relationships through *visual intersection points* (VIPs): a 2-D representation that captures relative position and orientation between two agents without additional parameterization (e.g., of field of view). A VIP lies at the location at which two individuals' lines of sight intersect.

To maintain the physical representation of the two "parent" points that contributed to each VIP, we encode each pairwise parent relationship by considering the full "pairwise triangle" defined by the VIP and its two associated social agents.

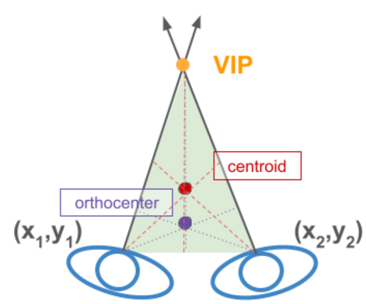

Fig. 3: An example of a pairwise triangle. Lines of sight for parent points (x_1, y_1) and (x_2, y_2) intersect at the VIP. The orthocenter is shown in purple, and the centroid in red.

B. Encoding Pairwise Triangles As Vectors

We convert the "pairwise triangle" representation into a 6-dimensional vector, where the visual intersection point is (v_x, v_y), the orthocenter of the pairwise triangle is $(orth_x, orth_y)$, the Euclidean distance from the triangle's centroid to the VIP is $d_{centroid}$, and the Euclidean distance between the parent points is d_{parent}, as follows:

$$\begin{bmatrix} v_x & v_y & orth_x & orth_y & d_{centroid} & d_{parent} \end{bmatrix}$$

From this representation, one can uniquely reconstruct the pairwise triangle. The $L2$-norm between two of these vectors can be used as a viable distance metric. We refer to this distance as *social distance*; two vectors closer to each other should be more socially intertwined since the positions, orientation alignment, and shared social space are captured in the pairwise triangle vector representation. We demonstrate how this method captures relevant social fluency information in an analysis using several cases in Section IV.

C. Persistence: Alpha Complexes

We use generation of an alpha complex, a type of simplicial complex, as a technique to relate and interpret pairwise triangles' relationships–and subsequently, those of social groups–via social distance.

Simplices are a generalization of tetrahedral space (e.g., in 2-D, triangles). A simplicial complex is a collection of simplices K where each face of a simplex of K is also in K, and every pair of distinct simplices of K have disjoint interiors [10]. Different types of complexes (e.g., Delaunay, alpha, Cech, Vietoris-Rips) have different properties and bounds on computational complexity [3] .

In this work, we use "alpha complexes" to identify the underlying structure of relative poses. We use alpha complexes because the persistent homology is guaranteed to be equivalent to that of the union of balls (shown in Fig. 2). When we maintain the guarantees of the persistent homology, we inherit properties such as provable stability. An alpha complex can also be interpreted as a subset of a Delaunay complex, which is the dual of a Voronoi diagram.

D. Determining Group Stability: Filtration Values

As noted in Section II, a filtration is a sequence of nested simplicial complexes [3]:

$$\mathcal{F} : \varnothing = K_0 \subseteq K_1 \subseteq \ldots \subseteq K_n = K$$

The filtration of the alpha complex enables analysis of the constituent complexes, and thus the underlying structure of groups of points. Each subcomplex has a *filtration value* given a filtration function. Smaller filtration values indicate complexes with constituent points that are closer to each other; this corresponds to "tighter" notions of social groups.

We use the filtration function provided by [9]. Figure 4 shows an example of an alpha complex with its associated *filtration tree*, which organizes filtration values by the nested sequence of subcomplexes. Filtration values for each level of a branch are not greater than those of the levels above. These filtration values allow us to reason about complexes and their corresponding subsimplicial complexes, which in our work relates social groups' relationships with their subgroups.

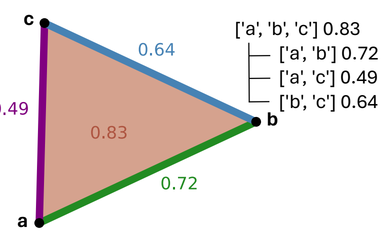

Fig. 4: An example of an alpha complex with labeled filtration values for edges and faces and its associated tree.

E. Implementation

We use the open source library GUDHI (Geometric Understanding in Higher Dimensions) 3.9.0 for topological data analysis [9].

Due to the synthetic nature of our data, we use an alpha value of 10000 to capture complete filtration trees for analysis. As discussed in Section IV, the large jumps in filtration values indicate possibilities for automating choice of alpha to capture inter- vs. intra-group dynamics based on real-world factors, including line-of-sight limitations, room size, etc. Section V further discusses alpha value determination.

IV. Analysis

Our method words towards a mathematical structure that encodes pairwise social relationships. Different components of this abstraction (e.g., edges, faces, dimensionality, and distance) may capture different information about interactions. We begin by examining filtration values with the aim to identify social groups and understand their subcomponents.

This initial analysis is to be followed by further investigation of the complexes themselves, such as the social information captured by an n-simplex. These steps are to be handled in future work, discussed in Section V.

A. Input Data

We generated eight synthetic examples: two configurations of three person groups, three four person groups, one five person group, and two other scattered configurations of five people. Four examples are shown in Fig. 5. Though synthetic, these cases are reflected in real data. For example, the three person group with two others facing away in Fig. 1 and case (a) in Fig. 5 are the same configuration. The synthetic data were generated to model specific group structures, and thus possesses social group membership ground truth.

We examine in detail a pair of three person group cases in Section IV-C and a pair of four person group cases in Section IV-D.

B. Analysis Method

For parent points to be labeled a social group, the simplicial complex containing *all* of their mutual pairwise triangles must have a filtration value that is low given the simplex dimension and relative magnitudes of other filtration values.

We expect that a complex s of dimension d whose pairwise triangles all relate parent points in the same social group will have a lower filtration value than a complex t of dimension d whose pairwise triangles relate parent points not in the same social group. In Section IV-C and Section IV-D, we examine the filtration values for example s and t complexes.

Note that in Fig. 5, filtration values are from the 6-D complex, but the illustrations are exclusively of the original 2-D configurations. The 2-D images capture the people, lines of sight, and pairwise triangles in a human-readable fashion. As 6-D space is difficult to visualize, we use filtration trees rather than images to understand the 6-D objects.

C. Three Person Stable Group: Handling Confounding Information

The three person group in Fig. 5 (a) and (b) has three children VIPs, forming three pairwise triangles that intuitively depict the shared space of the group. Even with confounding information from the non-member individuals pointing towards the group in (b), our method is able to discover the correct social grouping. We observe a gap in

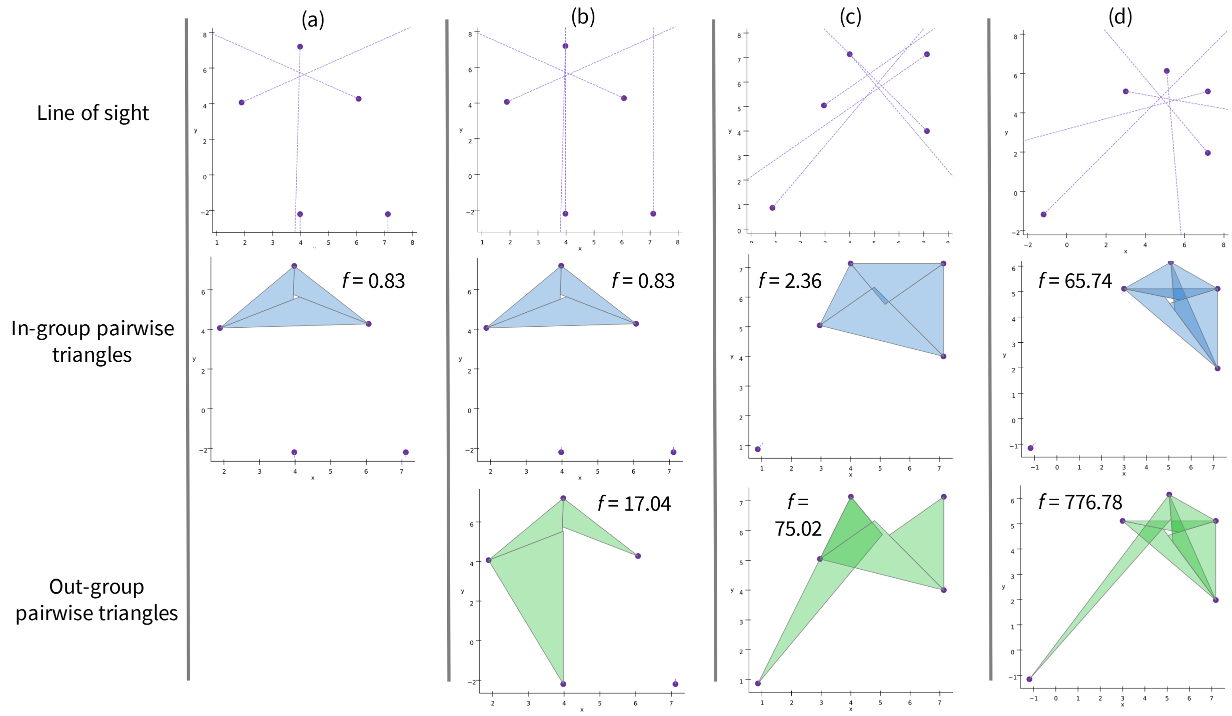

Fig. 5: Configurations of people and their lines of sight (purple, top), pairwise triangles for all dyads in the social group (blue, center), and pairwise triangles that are not exclusively of the social group (green, bottom). 3-group with 2 others facing away (a); same 3-group, but 2 others facing towards the group (b); 4-groups with different member poses (c, d). The filtration value f of simplices from the 6-D vectorized representations is shown next to corresponding sets of triangles.

filtration values f formed by the complex in 6-D space created by in-group pairwise triangles (0.83) and another (same-dimension) complex that includes an out-group pairwise triangle (17.04). Visually, we observe that the pairwise triangle relating to the out-group member is further stretched out and a different size compared to in-group triangles.

D. Four Person Stable Group: Handling Spatial Changes

Fig. 5 (c) and (d) both show social groups of four members with a single distractor non-member oriented toward the group. The group in (c) is more loosely positioned, with a greater area of shared social space than in (d). However, the two people in (d) are oriented less tightly to the group; in a real-world setting, this could represent a willingness to accept the nonmember into the group at a later stage.

Again, our method produces filtration values generally with orders of magnitude separation between the in-group complex and other (same-dimension) complexes that include non-members: 2.36 vs. 75.02 for (c) and 65.74 vs. 776.78 for (d), each corresponding to in-group and out-group complexes, respectively. We note that the increase in overall magnitude of filtration values from (c) to (d) likely is influenced by both the increase in complex dimension (from 4 to 6) and the “openness” of the group in (d) to accepting a new member, signifying decreased stability of the group.

However, we found that the in-group subsimplicial complex for (d) was not the only subset of 6-D points with a filtration value of smaller magnitude; we observed one other complex with a comparable value. Though this was the only outlier found in analysis of all synthetic cases, this finding suggests potential necessity for additional processing. We discuss strategies for addressing this outlier in Section V.

V. Conclusion and Future Work

We apply persistent homology to spatial information in a social environment to extract filtration trees and filtration values that are valuable for understanding the structure of social groups and their stability.

Real Data and Physical Considerations: We have analyzed our generated social complex structure and filtration trees on eight synthetic examples. However, further exploration with real data is necessary to expose a method that can handle outliers such as the one found in Fig. 5 (d), since our synthetic cases may not accurately reflect the complexity of natural social interactions or may carry biases from their hand-generation. Moreover, real-world datasets would enable us to consider natural physical limitations (e.g., room size and limits on vision). Our use of a complex that hinges on the alpha parameter caters well to these considerations; further exploration could incorporate physical properties when selecting alpha.

Conclusion: In this work, we present a structure and measure for cohesion of social groups. We analyze the efficacy of our proposed abstraction using a set of synthetic examples that are generalizations of real world structures. This work takes a step towards the robot that understands the evolution of a room's social dynamics, as well as how it can play a role in influencing the social structure of these groups.

REFERENCES

[1] Xavier Alameda-Pineda, Jacopo Staiano, Ramanathan Subramanian, Ligia Batrinca, Elisa Ricci, Bruno Lepri, Oswald Lanz, and Nicu Sebe. SALSA: A Multimodal Dataset for the Automated Analysis of Free-Standing Social Interactions. In *Group and Crowd Behavior for Computer Vision*, pages 321–340. Academic Press, January 2017.

[2] Stefano Alletto, Giuseppe Serra, Simone Calderara, Francesco Solera, and Rita Cucchiara. From Ego to Nos-Vision: Detecting Social Relationships in First-Person Views. In *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 594–599, June 2014. ISSN: 2160-7516.

[3] Tamal Krishna Dey and Yusu Wang. *Computational Topology for Data Analysis*. Cambridge University Press, 1 edition, February 2022.

[4] Tian Gan, Yongkang Wong, Daqing Zhang, and Mohan S. Kankanhalli. Temporal Encoded F-Formation System for Social Interaction Detection. In *Proceedings of the 21st ACM International Conference on Multimedia*, MM '13, pages 937–946, New York, NY, USA, October 2013. Association for Computing Machinery.

[5] Hayley Hung and Ben Kröse. Detecting F-Formations as Dominant Sets. In *Proceedings of the 13th International Conference on Multimodal Interfaces*, ICMI '11, pages 231–238, New York, NY, USA, November 2011. Association for Computing Machinery.

[6] Adam Kendon. *Conducting Interaction: Patterns of Behavior in Focused Encounters*. Cambridge University Press, New York, NY, US, 1990. Pages: xi, 292.

[7] Hideaki Kuzuoka, Yuya Suzuki, Jun Yamashita, and Keiichi Yamazaki. Reconfiguring Spatial Formation Arrangement by Robot Body Orientation. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 285–292, March 2010. ISSN: 2167-2148.

[8] Claire Liang, Ross A. Knepper, and Florian T. Pokorny. No Map, No Problem: A Local Sensing Approach for Navigation in Human-Made Spaces Using Signs. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6148–6155, Las Vegas, NV, USA, October 2020. IEEE Press.

[9] Clément Maria, Jean-Daniel Boissonnat, Marc Glisse, and Mariette Yvinec. The Gudhi Library: Simplicial Complexes and Persistent Homology. In *Mathematical Software – ICMS 2014*, pages 167–174, Berlin, Heidelberg, 2014. Springer.

[10] James R. Munkres and James W. Munkres. *Elements Of Algebraic Topology*. CRC Press, Boca Raton, June 2019.

[11] Navyata Sanghvi, Ryo Yonetani, and Kris Kitani. MGpi: A Computational Model of Multiagent Group Perception and Interaction. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '20, pages 1196–1205, Richland, SC, May 2020.

[12] Francesco Setti, Oswald Lanz, Roberta Ferrario, Vittorio Murino, and Marco Cristani. Multi-Scale F-Formation Discovery for Group Detection. In *2013 IEEE International Conference on Image Processing*, pages 3547–3551, Melbourne, Australia, September 2013. IEEE.

[13] Francesco Setti, Chris Russell, Chiara Bassetti, and Marco Cristani. F-Formation Detection: Individuating Free-Standing Conversational Groups in Images. *PLOS ONE*, 10(5):e0123783, May 2015. Publisher: Public Library of Science.

[14] Mason Swofford, John Peruzzi, Nathan Tsoi, Sydney Thompson, Roberto Martín-Martín, Silvio Savarese, and Marynel Vázquez. Improving Social Awareness Through DANTE: Deep Affinity Network for Clustering Conversational Interactants. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1):20:1–20:23, May 2020.

[15] Allan Wang and Aaron Steinfeld. Group Split and Merge Prediction With 3D Convolutional Networks. *IEEE Robotics and Automation Letters*, 5(2):1923–1930, April 2020. Conference Name: IEEE Robotics and Automation Letters.

[16] F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita. A Model of Proximity Control for Information-Presenting Robots. *IEEE Transactions on Robotics*, 26(1):187–195, February 2010.

[17] Mohammad Yousuf, Yoshinori Kobayashi, Yoshinori Kuno, Akiko Yamazaki, and Keiichi Yamazaki. Development of a Mobile Museum Guide Robot That Can Configure Spatial Formation with Visitors. volume 7389, pages 423–432, July 2012.

[18] Lu Zhang and Hayley Hung. Beyond F-Formations: Determining Social Involvement in Free Standing Conversing Groups From Static Images. pages 1086–1095, 2016.