Finding f-formations using dominant sets in the Bluetooth proximity data of the Conflab data set

Bram Dikker[1]
Supervisor(s): Stephanie Tan[1], **Hayley Hung**[1]
[1]EEMCS, Delft University of Technology, The Netherlands
b.dikker@student.tudelft.nl

Abstract

When analysing social interactions, manual labour is often required to identify what is happening. An automated method of detecting who is interacting with who would already prove to be a significant help. This paper looks at how automated interaction detecting can be established. We look at methods of detecting proximity and look deeper into detecting F-formations using proximity. An f-formation is a group of people who are standing together with the intention of conversing. We show that it is possible to detect f-formations using a data-set containing proximity information and f-formations as ground truths. Our results show that using only proximity data from this dataset; we can detect f-formations better than the baseline provided in that dataset.

1 Introduction

Smart wearable devices are becoming ever more present in day-to-day life. These wearables have varying sensors that can obtain data about activities the user undertakes. Using this data, we can analyse behaviour that can be relevant, for example, for health care, fitness tracking, and smart assistants [1], [2]. Smart fitness tracking is already widely available in order to improve and track ones' performance [3]. However, these sensors are capable of monitoring much more than would be apparent at first.

One of the recent usages of these sensors is the contact tracing of Covid-19, for this Google and Apple created Decentralized Privacy-Preserving Proximity Tracing [4]. Which uses Bluetooth Low Energy (BLE) to estimate if two phones are closely situated and can anonymously notify users if they have been in contact with someone who has been infected with Covid-19. This shows that BLE proximity data is capable of determining physical proximity.

Research has also been conducted to determine if BLE can determine social interactions between users carrying BLE capable devices [5]. This method allows researchers to capture this data automatically instead of using observations or surveys that can be unreliable because of human errors [6]. Using sensors can provide a quantitative approach to study group dynamics that is also scale-able and ubiquitous, given the rise of smart devices with capable sensors.

The BLE sensors report back the Received Signal Strength Indicator (RSSI) of the BLE devices discovered. It is an estimate of the strength of the signal it's receiving. However, it is dependent on the manufacturer of the device how these values are calculated, including what range the RSSI might have. Therefore the RSSI is the relative signal strength and not a direct measurement. However, universally applicable is that the closer to 0, the better the signal. Because it's a radio signal, it's also susceptible to interference and becomes unreliable after even a short-range [7]. Another aspect of this is that it's Omnidirectional and therefore only approximates the range and not the direction. Adding onto this RSSI value is calculated at the receiving end of the signal, which could mean that the values reported by both devices are different [8].

Other methods of detecting proximity have been investigated in the past. For instance, the use of Radio Frequency Identification (RFID) tags has been studied [9]. These tags can detect proximity on a much lower distance scale than BLE. This is done by tuning the power levels on both receiving and sending ends. Lowering the power levels allows them to assess face-to-face proximity because the tags are localised to detect only tags facing towards them.

A different modality tested is using infrared (IR) sensors. These sensors are directional, do not pass through participants, and can detect participants facing each other [10], [11].

Another method to detect proximity is using Wi-Fi signals. This can be done by using wireless access points as beacons and recording them for each participant [12]. This reveals information about the general whereabouts of participants and not face-to-face interactions. However, it has been shown that using Wi-Fi can be a viable way to detect physical proximity but has not been tested against ground truth data. But has been demonstrated to detect interactions in which a high RSSI between two Bluetooth capable devices is present [13].

Proximity can be used to infer important aspects of interpersonal relationships and character features. We can define this more rigidly by looking at F-formations. An F-formation is a form of focused encounter wherein a group of individuals contribute to maintaining a prolonged conversation. Kendon [14] more clearly defines it as a formation in which people gather that has a convex space, that every member of the F-formation has equal access to. In the past work has been to detect F-formations using a multitude of methods and different modalities [11], [15]–[17]. One of those modalities is proximity, which we can be estimated using RSSI.

One method of approximating f-formations is to formulate it as a graph-theory problem, that of identifying dominant sets (DS) [18]. DS represent maximal clique using edge-weighted cases. All nodes in the graph will represent a person, while the edge weight will represent the affinity or similarity between two pairs. This follows naturally from f-formations because intuitively the affinity between two people inside a f-formations would be higher than that of those outside of the f-formation [16].

For this paper, we will be looking at the Conflab data set of the Socially Perceptive Computing Lab (SPCL) at the TU Delft. The SPCL has created a smart wearable badge called a Midge [19] that has sensors for collecting low-frequency audio, BLE proximity and 9-axis Inertial Measurement Unit data. The badges were used to gather data from a mingling event [20]. In this event 48, socially interacting people are all wearing the Midge, recording the sensor data. The Conflab data sets also include manually annotated f-formations for part of the event. For this body of work, we will only be looking at the BLE capabilities of this badge to calculate f-formations. While we know that using BLE can be used to approximate vicinity. What is still unknown is whether using the BLE data from the badges developed at the SPCL can detect f-formations and how accurate they are compared to the manually annotated f-formations.

First, we will explain how the data was extracted from the Conflab data set. We then propose multiple parameters for pre-processing and transforming the data. This data is then used in a clustering algorithm to find f-formations. After which the evaluation method is explained, we discuss the results with the different parameters.

2 Methodology

We used an established method of calculating F-formations. The adaption here is how we pre-process the data before using DS clustering to find the F-formations. First, we describe how we extracted the data from the Conflab data set. Then in the next section, we show how this data is transformed, using different parameters for usage in the DS clustering method. Lastly, our evaluation method is explained. The appendix contains figure 1 in the appendix which shortly explains how and where the data is processed.

2.1 Dataset

We use the Conflab Data set **conflab** containing BLE and ground truth data. The data set consists of 48 Midges which each have separate data files for the different modalities it has recorded. There are around 190.000 data points for each of the separate Midges used for the BLE data.

The ground truth data from the annotations consist of 1277 entries that denote the different F-formations at that second of video data; this comes down to roughly 21 minutes. The F-formations are manually annotated from the video cameras. The ground-truth is synchronised with the proximity readings by using timestamp T.

The Midges broadcast packets every second. However, this does not automatically mean that all Midges will receive every packet from every Midge. There could be multiple reasons for this, one being interference. This is why it's essential to look at what time the RSSI was recorded. The last time Midge A has seen Midge B might show a relatively strong signal; this quickly becomes less interesting as time passes. A time-out parameter K was created to account for this. When the time difference is larger than the parameter, it will replace the RSSI value with the lowest value, in our case -100. So the used RSSI is calculated as follows. Here t_i is the time that the RSSI value v_i was measured for Midge i and T the timestamp of the F-formation in the ground truth.

$$RSSI_i = \begin{cases} v_i & \text{if } \mid T - t_i \mid <= K \\ -100 & \text{else} \end{cases} \tag{1}$$

A problem was that one of the Midges misreported all RSSI values as 1. We used a simple data reconstruction method to overcome this problem. The data were reconstructed using the RSSI values that the other had Midges estimated from the signal it received from this Midge. More on this in the Affinity Matrix section.

Another critical factor is that because the number of participants n differed from frame to frame, we extracted only the data corresponding with Midges used in the annotated scene. This required parsing the ground truths to find who was participating at what timestamp—then extracting only the BLE proximity data from those Midges.

2.2 Dominant Set Clustering

The method we used to find F-formations is DS Clustering, which requires a pair-wise affinity matrix. The scores are a value between 1 and 0, where 1is the maximum. How this was created is explained in the next paragraph. This algorithm is practical because it does not require a preset number of clusters to be found. Because in our case, the proximity data alone does not indicate how many f-formations should be found.

We used Hayley et al. [16] algorithm for detecting F-formations as Dominant sets for this paper.

2.3 Affinity Matrix

The affinity matrix represents a matrix of all pair-wise similarity scores. To use the DS clustering method, we first created a matrix of n by n.

The matrix is then filled with the respective RSSI value received from each Midge. Depending on the parameter set in equation 1.

Because the RSSI signals between two Midges are not symmetrical, we can choose to symmetrise the Matrix to reduce noise. Three methods of symmetrisation were selected. Taking the minimum value, the maximum value or the average.

Here we used a form of feature engineering by transforming the RSSI value depending on a cut-off. This was done because the further two Bluetooth devices become apart, the more susceptible they become to interference and noise. They are showing strong unreliability after even a short-range. To do this, a simple cut-off algorithm was used, setting a threshold C for calculating the affinity a_i for Midge i:

$$a_i = \begin{cases} 1 & \text{if } v_i <= C \\ 0 & \text{else} \end{cases} \quad (2)$$

2.4 Evaluation method

To evaluate our methods, we calculated the precision, recall, and f-measure for each F-formation found using the DS clustering. We chose to calculate the $F1$ score, meaning that for a group to be considered a true positive, all members of that group should be included in the ground truth and no more. A false positive is a group that is detected but does not exist in the ground truth, and false negatives are groups that have not been detected.

$$F_1 = \frac{truepositive}{truepositive + \frac{1}{2}(falsepositive + falsenegative)} \quad (3)$$

Groups of 1 were also included as F-formations in the ground truths. The performance results are categorised with the different parameters it used. For threshold, we used a value of -55, which was calculated by resampling other portions of the data as a test group and a training group to mitigate overfitting.

3 Results

A summary of the results is shown in table 1. For all of the results in the summarized table we used data reconstruction when the broken Midge was in the annotated frame. the Full table can be found in the appendix in table 2.

In total 1252 seconds of annotated frames were used to calculate all the f-measures. The best results were found when taking a maximum time difference of 30 seconds and symmetrisation using the average.

4 Conclusion and discussion

Detection of f-formations using the BLE proximity in the Conflab data-set shows promising results. Comparing our best $f1$ score to the best $f1$ score in baseline shows we perform 52% higher. This is highly likely when looking at the table 2 because 25167 values were replaced with -100, and then taking average means that it is more conservative when the Midge was last seen longer than 30 seconds ago.

We present many important parameters when using the Conflab data-set to calculate f-formations. The recall significantly decreases when symmetrising using the max function. This is because the number of true positives decreases with more detected f-formations than in the ground truths.

The method also fails heavily as f-formations are broken and created. This is highly likely because of the interference caused by more participants moving throughout the room. It would be interesting if we labelled these events and saw the average f-measure around them.

Another issue we ran into was that parts of the annotated file contained errors. This is why 25 annotated frames of f-formations were discarded.

Timeout	Sym	F1 measure	Precision	Recall
10	avg	0.605	0.679	0.546
	max	0.531	0.645	0.452
	min	0.573	0.636	0.521
20	avg	0.617	0.692	0.557
	max	0.533	0.649	0.452
	min	0.581	0.645	0.529
30	avg	**0.625**	0.696	0.568
	max	0.530	0.649	0.447
	min	0.582	0.648	0.528
40	avg	0.617	0.687	0.559
	max	0.530	0.646	0.448
	min	0.583	0.646	0.532
50	avg	0.619	0.695	0.557
	max	0.526	0.650	0.442
	min	0.577	0.642	0.524
60	avg	0.613	0.691	0.550
	max	0.528	0.650	0.445
	min	0.584	0.646	0.534

Table 1: Summary of the results for the average F1-measure, Precision and recall. The bold value is the highest F1 measure that was found using the DS clustering. The values for the different parameters chosen are also shown in the table.

What could also improve the results is instead of using the RSSI values directly, is using them to calculate vectors and see how they are moved over time. This could increase the f-measure when looking at moments where a lot of movements are happening, this is where our method currently fails.

The most significant improvement that probably can be made is instead of using a simple cut-off algorithm; we could try to map the RSSI values to actual distance and compensate for the noise. This would require altering how the Midges recorded the RSSI values. But would increase the performance of finding the f-formations.

5 Responsible Research

Our results show that the usage of widely available and cheap BLE sensor hardware can be used to estimate f-formations. While this might seem innocent at first, if widely deployed it can be used to determine someones social circle. For this however a lot more information would be required about the devices the BLE sensor is able to observe. The methods we use only works because of the Conflab dataset. For which the decision was made to try and maximize data fidelity while persevering participants privacy.

Given the parameters and access to the Conflab dataset, reproducing the results will not be a problem. As most methods are trivial to implement.

A Schematic of data flow and processing

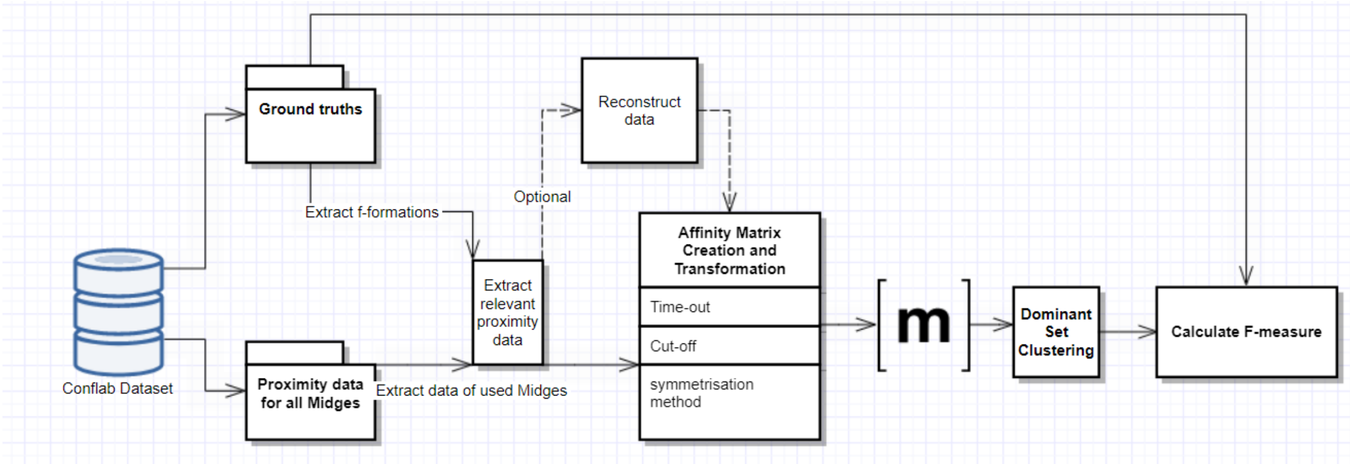

Figure 1: Simple schematic of how the data flows through the various parts of our method and where they are used

B Full table

Timeout	Symmetrisation	Reconstruction	F1 measure	Precision	Recall	Timeouts occurred
10	avg	FALSE	0.010	0.041	0.006	52308
10	max	FALSE	0.505	0.614	0.428	52308
10	min	FALSE	0.572	0.635	0.520	52308
10	avg	TRUE	0.605	0.679	0.546	52308
10	max	TRUE	0.531	0.645	0.452	52308
10	min	TRUE	0.573	0.636	0.521	52308
20	avg	FALSE	0.007	0.028	0.004	32677
20	max	FALSE	0.495	0.606	0.418	32677
20	min	FALSE	0.585	0.649	0.533	32677
20	avg	TRUE	0.617	0.692	0.557	32677
20	max	TRUE	0.533	0.649	0.452	32677
20	min	TRUE	0.581	0.645	0.529	32677
30	avg	FALSE	0.007	0.025	0.004	25167
30	max	FALSE	0.489	0.604	0.411	25167
30	min	FALSE	0.586	0.648	0.535	25167
30	avg	TRUE	0.625	0.696	0.568	25167
30	max	TRUE	0.530	0.649	0.447	25167
30	min	TRUE	0.582	0.648	0.528	25167
40	avg	FALSE	0.007	0.029	0.004	20865
40	max	FALSE	0.487	0.602	0.409	20865
40	min	FALSE	0.578	0.644	0.524	20865
40	avg	TRUE	0.617	0.687	0.559	20865
40	max	TRUE	0.530	0.646	0.448	20865
40	min	TRUE	0.583	0.646	0.532	20865
50	avg	FALSE	0.006	0.024	0.004	17812
50	max	FALSE	0.494	0.605	0.417	17812
50	min	FALSE	0.586	0.647	0.536	17812
50	avg	TRUE	0.619	0.695	0.557	17812
50	max	TRUE	0.526	0.650	0.442	17812
50	min	TRUE	0.577	0.642	0.524	17812
60	avg	FALSE	0.007	0.030	0.004	15446
60	max	FALSE	0.488	0.608	0.408	15446
60	min	FALSE	0.585	0.652	0.530	15446
60	avg	TRUE	0.613	0.691	0.550	15446
60	max	TRUE	0.528	0.650	0.445	15446
60	min	TRUE	0.584	0.646	0.534	15446

Table 2: Summary of the results for the average F1-measure, Precision and recall. The values for the different parameters chosen are also shown in the table. The number of times a value was replaced because of equation 1 is also shown.

References

[1] G. Shin, M. H. Jarrahi, Y. Fei, *et al.*, "Wearable activity trackers, accuracy, adoption, acceptance and health impact: A systematic literature review," *Journal of Biomedical Informatics*, vol. 93, p. 103 153, May 2019, ISSN: 1532-0464. DOI: 10.1016/J.JBI.2019.103153.

[2] A. Ometov, V. Shubina, L. Klus, *et al.*, "A Survey on Wearable Technology: History, State-of-the-Art and Current Challenges," *Computer Networks*, vol. 193, p. 108 074, Jul. 2021, ISSN: 1389-1286. DOI: 10.1016/J.COMNET.2021.108074.

[3] A. Kononova, L. Li, K. Kamp, *et al.*, "The Use of Wearable Activity Trackers Among Older Adults: Focus Group Study of Tracker Perceptions, Motivators, and Barriers in the Maintenance Stage of Behavior Change," *JMIR Mhealth Uhealth 2019;7(4):e9832 https://mhealth.jmir.org/2019/4/e9832*, vol. 7, no. 4, e9832, Apr. 2019, ISSN: 22915222. DOI: 10.2196/MHEALTH.9832. [Online]. Available: https://mhealth.jmir.org/2019/4/e9832.

[4] *GitHub - DP-3T/documents: Decentralized Privacy-Preserving Proximity Tracing – Documents*. [Online]. Available: https://github.com/DP-3T/documents.

[5] A. Montanari, S. Nawaz, C. Mascolo, and K. Sailer, "A Study of Bluetooth Low Energy performance for human proximity detection in the workplace," *2017 IEEE International Conference on Pervasive Computing and Communications, PerCom 2017*, pp. 90–99, May 2017. DOI: 10.1109/PERCOM.2017.7917855.

[6] S. R. Corman, E. Steiner, J. D. Proulx, *et al.*, "Revisiting the accuracy problem in network analysis using a unique dataset," *Social Networks*, vol. 66, pp. 1–9, 2021. DOI: 10.1016/j.socnet.2020.12.010. [Online]. Available: https://doi.org/10.1016/j.socnet.2020.12.010.

[7] Q. Dong and W. Dargie, "Evaluation of the reliability of RSSI for indoor localization," *2012 International Conference on Wireless Communications in Underground and Confined Areas, ICWCUCA 2012*, 2012. DOI: 10.1109/ICWCUCA.2012.6402492. [Online]. Available: https://www.researchgate.net/publication/261447858_Evaluation_of_the_reliability_of_RSSI_for_indoor_localization.

[8] *Specifications — Bluetooth® Technology Website*. [Online]. Available: https://www.bluetooth.com/specifications/.

[9] C. Cattuto, W. Van Den Broeck, A. Barrat, V. Colizza, J.-F. Pinton, and A. Vespignani, "Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks," DOI: 10.1371/journal.pone.0011596. [Online]. Available: www.plosone.org.

[10] T. Choudhury and A. Pentland, "Sensing and modeling human networks using the sociometer," *Proceedings - International Symposium on Wearable Computers, ISWC*, pp. 216–222, 2003, ISSN: 15504816. DOI: 10.1109/ISWC.2003.1241414.

[11] X. Alameda-Pineda, J. Staiano, R. Subramanian, *et al.*, "SALSA: A novel dataset for multimodal group behavior analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 8, pp. 1707–1720, Aug. 2016, ISSN: 19393539. DOI: 10.1109/TPAMI.2015.2496269.

[12] A. Madan, M. Cebrian, S. Moturu, K. Farrahi, A. " Sandy, and ". Pentland, "Sensing the "Health State" of a community," Tech. Rep., 2012. [Online]. Available: www.computer.org/pervasive.

[13] P. Sapiezynski, A. Stopczynski, D. K. Wind, J. Leskovec, and S. Lehmann, "Inferring Person-to-person Proximity Using WiFi Signals," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 2, pp. 1–20, Oct. 2016, ISSN: 2474-9567. DOI: 10.1145/3090089. [Online]. Available: https://arxiv.org/abs/1610.04730v1.

[14] J. K. Burgoon and A. Kendon, "Conducting Interaction: Patterns of Behavior in Focused Encounters.," *Contemporary Sociology*, vol. 21, no. 2, 1992, ISSN: 00943061. DOI: 10.2307/2075490.

[15] H. Hedayati, D. Szafir, and S. Andrist, "Recognizing F-Formations in the Open World," *ACM/IEEE International Conference on Human-Robot Interaction*, vol. 2019-March, pp. 558–559, Mar. 2019, ISSN: 21672148. DOI: 10.1109/HRI.2019.8673233.

[16] H. Hung, H. H. Nl, and B. Kröse, "Detecting F-formations as Dominant Sets,"

[17] L. Cabrera-Quiros, A. Demetriou, E. Gedik, L. Van Der Meij, and H. Hung, "The MatchNMingle Dataset: A Novel Multi-Sensor Resource for the Analysis of Social Interactions and Group Dynamics In-The-Wild during Free-Standing Conversations and Speed Dates," *IEEE Transactions on Affective Computing*, vol. 12, no. 1, pp. 113–130, Jan. 2021, ISSN: 19493045. DOI: 10.1109/TAFFC.2018.2848914.

[18] M. Pavan and M. Pelillo, "Dominant sets and pairwise clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 167–172, Jan. 2007, ISSN: 01628828. DOI: 10.1109/TPAMI.2007.250608.

[19] *GitHub - TUDelft-SPC-Lab/spcl_midge_hardware: Hardware design files for the SPC lab MIDGE wearable devices for in-the-wild social data collection.* [Online]. Available: https://github.com/TUDelft-SPC-Lab/spcl_midge_hardware.

[20] "ConfLab: A Rich Multimodal Multisensor Dataset of Free-Standing Social Interactions In-the-Wild."