Obtaining Relational Data and Activity-Based Social Networks from Track Data: An Example from the WASABI Project

E. El-Mahassni[a] **and G. Burgess**[a]

[a]*National Security and ISR Division,Defence Science and Technology Group*
Email: Edwin.El-Mahassni@dst.defence.gov.au

Abstract: Activity Based Intelligence (ABI) is defined by the Office of the Undersecretary of Defense [US] as "a discipline of intelligence where the analysis and subsequent collection is focused on the activity and transactions associated with an entity, a population, or an area of interest." Functions which are typically coordinated include collection prioritisation, fusion and analysis. Similarly, relational data is concerned with the entities' relationships to other entities, through formal and informal institutions, social networks, and other means. Models from this data can be dynamic and might need constant validation to improve accuracy and usefulness. In other cases, relational data such as kinship, familial ties, leadership or friendship might be more long-lasting and stable. By clustering entities' activities with respect to time and space, relationships can be deduced for a given dataset. The purpose of this paper is to determine whether social networks could be derived from identifying entity interactions obtained from tracks. The work described here clearly relates entities and interactions to other entities by inferring a social network from tracks. Building (or adding to) the social network of a group of individuals can help an analyst understand more about each entity within the group, and conversely, more about the group as a whole. This can improve organisational knowledge about existing persons of interest and it can also enhance the tactical intelligence value of live observations by linking in to what is already known. In addition, this may be useful for an analyst in order to validate or complement an existing representation of a social network.

Keywords: *Activity-based intelligence, relational data, entity interactions, social networks*

1 INTRODUCTION

Data fusion can be defined as the process of integrating data and knowledge, typically from multiple sources and source types. The aim is to develop a consistent, accurate and useful representation from those sources. One common application of data fusion is for intelligence, where the required data is often ambiguous, sparse, incomplete or innacurate (deliberate or not) and generally hard to obtain. This may result in information that contains uncertainty or imprecision. It is an intrinsic goal of intelligence operations to be able to reduce this uncertainty.

Integrating data presents its own challenges. For instance, sensor data is precise while eyewitness accounts provide context. This may introduce the need for a normative or standardised scale which refines data and allows it to be combined with other pieces of information. The need to update new data in a temporal manner as it is being observed, as well as requiring the data to be presented in its semantic form also demands consideration. These are important as we also look to evolve enterprise intelligence systems. The problems are sometimes complex but the key to good data fusion is the ability to utilise all information regardless of its source or form.

Activity Based Intelligence (ABI) is defined by the Office of the Undersecretary of Defense [US] as "a discipline of intelligence where the analysis and subsequent collection is focused on the activity and transactions associated with an entity, a population, or an area of interest"(Phillips, 2012). ABI is seen as one way to orchestrate functions to maximise the effectiveness of intelligence operations. Functions which are typically coordinated include collection prioritisation, fusion and analysis. High level fusion is another function of ABI which can be thought of as the comprehension of a situation under assessment. The challenges considered in this paper lie across the data fusion and analytics functions. Analytics can be categoraised as descriptive (what happened), predictive (what may happen) and prescriptive (what should be done).

In this paper, entity interactions can be defined as exchanges that occur between two or more entities. Similarly, relational data is concerned with the entities' relationships to other entities, through formal and informal institutions, social networks, and other means (Biltgen and Ryan, 2016). Typically, such relational data may be obtained from various means such as social media data, open source data, financial transactions, electronic communications, etc. Relational models can be dynamic and might need constant validation to improve accuracy and usefulness. In other cases, relational information such as kinship, familial ties, leadership or friendship might be more long-lasting and stable. In this paper, we use track data to observe entity interactions and infer relational information. One limitation is that not much can be established on the nature of the existing relationship from track observations.

Of the trials looked at this paper, Jalapeño took place as part of project WASABI (Wide Area Surveillance and Activity Based Intelligence) in November 2014. A primary goal of WASABI was to collect comprehensive data to enable analysis, research and development of techniques. Relational models from this data were developed by clustering single interactions between entities with respect to space and time. These models were then transformed into a network which was visualised with Palintir ^{TM}(Palintir, 2015). Although social media data was artificially generated for analysis, the purpose of this paper is to determine whether social networks could be derived from identifying entity interactions obtained from tracks.

Social media data and open source intelligence is always incomplete. Weller and Kinder-Kurlanda (Weller and Kinder-Kurlanda, 2015) recall previous work in this area and note some of the problems lie with ethical considerations, data quality, rate/volume limits, changes to structure, etc. Here, a social network based on entity interactions obtained from tracks is built without the use of social media or open source data. Such a network can be used as a first-estimate for a social network or, based on the connectivity of certain entities and their roles, it can be used to infer the type of activity being planned.

2 PREVIOUS WORK

2.1 PIMENTO TRIAL

Pimento was run as the inaugural trial under the Wide Area Surveillance and Activity Based Intelligence (WASABI) project in May 2014. The goal of this trial was to collect data from a variety of sources, including an airborne sensor platform over a wide geographical area with scenarios undertaken by DST Group volunteers. This data can then analysed for a variety of purposes, such as detecting anomalies behaviour, entity resolution, etc.

Using some of the data from the Pimento trial, several techniques and methods for conducting analysis high-

lighted the importance of ABI and the human element. That is, the importance of determining where a person or persons of interest have been or are going is further augmented by determining any potential interactions with other people or groups of people. Sure, there was a focus on the kinematics such as accurately determining the position, distance or speed of a particular entity, but it was also equally crucial to capture any instances on whether an entity had been in communications with another entity. All the data referenced here were based on the movements of DST Group volunteers who are here after referenced to as entities. However, in this paper it is assumed that SIGINT or data from private communications among the entities is not available.

ENTITY INTERACTIONS. The analysis consisted of establishing whether any two entities had crossed paths, met or travelled together by simply using the GPS tracks. Thresholds were designated for distance, time and speed and potential instances of these interactions were identified.

Entity interactions are a subset of a full list of possible events. But, in ABI, the importance of the human element emphasises these types of interactions. Here, the following pair-wise entity interactions are identified:1) crossed paths, 2) met each other, 3) travelled together and 4) were not close.

The crossing of paths between two entities can be distinguished from a meeting because of their duration. Two entities are deemed to have "met" if they were within a certain distance of one another for a minimum length of time and with both entities being stationary. Similar thresholds can be made for the other interactions.

Using the "met" rule, a graph of potential meetings occurring can be derived as shown in Figure 1. The x-axis indicates time since the trial started and the y-axis is a measure of belief that an interaction took place.
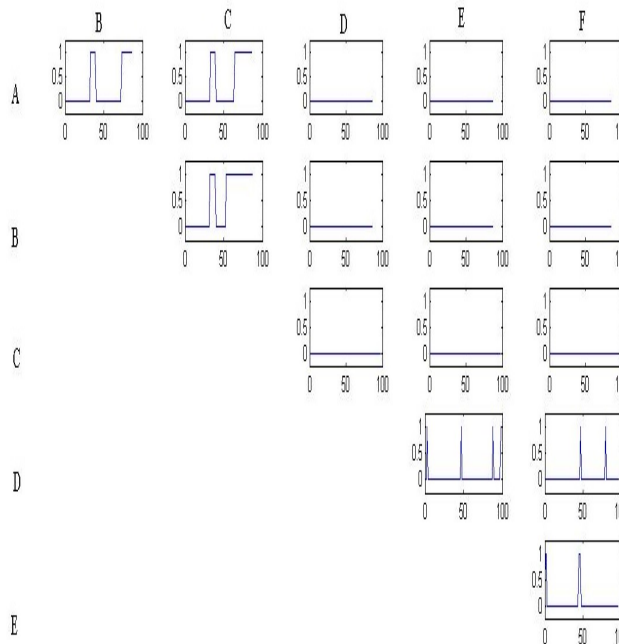
Figure 1. Plots of "meetings" between any two entities

Figure 1 shows that A, B and C met for about 10 minutes, while D, E and F met for about 5 minutes. The plots of Figure 1 show that D also then met with F and a short time later and on a separate occasion with E. This is an accurate assessment of the interactions based on truth. Here truth is determined by the scenarios undertaken by DST Group volunteers.

Based on observed interactions between pairs of entities throughout the trial, two distinct groups emerge: namely A, B, C and D, E, F. As useful as these images and extracted interactions might be, they still fail to appropiately quantify the importance of each interaction or the context within which they occurred. For example, a particular meeting might have not lasted a long period of time, but if it was followed or proceeded

by several other meetings or entites followed each other, then its significance should be recognised in some manner. This is where our attention now turns.

2.2 JALAPEÑO TRIAL

As in Pimento, this team consisted of DST Group volunteers who set out with a set of prescribed activities. But, unlike the Pimento trial, there was one group (instead of two) working together and cohesively. Here, the task was to coordinate a public nuisance event. This consisted of a movement of a GPS-tracked package which contained offensive material. The data collection for this trial is similar to Pimento. However, in this exercise, there was also the injection of artificial social media feeds as another source of data which could be exploited. One of the objectives of this exercise was to determine which GPS track depicts the movement of the package.This makes it ideal to build Track-Based Social Networks, which will be described in Section 3.3. However, here in this paper, trying to determine the structure of a single social network means that the data from Jalapeño is used, but without any utilising any of the artificially-generated social-media data.

3 ENTITY INTERACTIONS TO TRACK-BASED SOCIAL NETWORKS

Entity interactions are here classified as crossings, travelling together or having met. Just because two entities were in close proximity to one another this does not mean that they may know each other or indeed there was a meeting. It could simply be that two entities happen to be nearby one another but unaware of the other's existence (and determine whether there is a relationship between two entities). Data from other sources or from multiple meetings are used to provide context to the interactions. For instance, in Fig 17.6 of (Biltgen and Ryan, 2016), a social network is constructed from police reports, arrest records, financial records, telephone records and public news. Context can also be obtained if more than one interaction took place, particularly within a short space of time. Clustering this information with the locations of where these physical transactions take place, provides an analyst with the geographical and holistic understanding of several interactions between two entities. Other information such as whether the meeting location was crowded or empty, or what mode of transport was used when traveling together would have potentially strengthened inferences. e.g. traveling together in a public bus vs private car - or both being in a crowded pub at the same time vs in a private house or park bench. Whilst the former instances could indicate good trade-craft the latter are more strongly suggestive of some level of relationship.

3.1 OBTAINING ENTITY INTERACTIONS

As in the Pimento trial, entity interactions in the form of meetings, crossings and travelling together were obtained. The variables that were used to establish an entity interaction are: the distance between two entities, the velocity of both entities and the duration of the entity interaction. Unlike the Pimento trial, all the entities travelled by car which meant that the velocity thresholds could be set at a higher value than would otherwise be the case (if the entities sometimes travelled on foot). Other events, such as two entities being near the same location but at different times of the day, could be considered an interaction. But here, entities are linked by being in close proximity to each other in time and space. The interactions are then listed in chronological order together with their latitudinal and longitudinal positions as shown in Figure 2.

3.2 RELATIONAL DATA

The interactions by themselves do not necessarily mean that two entities know each other or have a relationship. For instance, entities A and B crossed at 11:01:43 but there is no other interaction between them. Hence, it is likely that the entities being in close proximity to one another was purely coincidental and does not represent a relationship of any kind. Conversely, some of the transactions can form clusters such as B and E travelling together at 11:11:44. We might conclude that these were not random events since they also had meeting transactions before and after that time.

For a few of the interactions, in particular the crossings, if no other interactions had occurred, then this would have been removed. Further reductions can be conducted if it is assumed that for two entities to have a relationship, a minimum of $N > 2$ interactions must have taken place between two entities within a short period of time. If this is the case, then the red, green and light blue interactions shown in Figure 3 would be eliminated also. This gives the analyst further assurance that when two entities had a series of interactions within a close period of time, they were not likely to be mere coincidences. This is a balancing act because although what is left would give greater confidence in the existence of a relationship between two entities, some other relationships might be prematurely dismissed. An alternative is to apply a probablity measure to

some tentative interactions which might be strengthened or dismissed as more data is gathered. Yet another parameter, related to time, would also have to be used. For instance two meetings in the space of an hour would be more significant than, say, two meetings within a year.

3.3 SOCIAL NETWORKS

INTRODUCTION. A social network can be defined as a social structure which graphically displays the relationships among a finite group of entities, places and/or locations. Social Network Analysis (SNA) seeks to study the patterns that arise from these structures and include determining: the most influential entities and the dynamics of the structure (if observed in a temporal sense). A social network analyst is interested in how an individual is embedded within a structure and the way this structure evolves from the micro-relations of individuals (Haspert, 2000). The types of relationships among entities can be wide-ranging and social networks have been used to represent communications among members of a group, economic transactions between corporations, and trade or treaties among nations (Wasserman, 1994). The networks are self-organizing, emergent and complex; and global coherent patterns appear from the localised interaction of individual elements that make up the system. Surveys or questionnaires can also be conducted to ascertain the type or strength of relationships, however, understandably this might not be feasible in an intelligence domain.

The most common way of annotating relational data is through adjacency matrices. In SNA, a graph might be directed or undirected. A directed graph models the direction of the relationship, whereas an undirected graph only details whether a relationship exists at all. The graph can also be represented by a matrix with either a zero or one in each entry and for an undirected graph the adjacency matrices are always symmetrical. A zero or one in the (i, j) spot would mean that person i does not have or has a tie with person j respectively.

Entity	Entity	Transaction	Duration(s)	Date	Time	Latitude	Longitude
B	E	crossed	2	2-Dec-14	10:46:23	-34.846	138.502736
B	E	travelled	133	2-Dec-14	10:51:17	-34.8403	138.496698
A	C	met	98	2-Dec-14	10:53:17	-34.8447	138.477931
B	E	met	228	2-Dec-14	10:59:53	-34.8397	138.487606
A	E	crossed	2	2-Dec-14	11:01:42	-34.8395	138.4874
A	B	crossed	1	2-Dec-14	11:01:43	-34.8396	138.487484
B	E	travelled	94	2-Dec-14	11:02:52	-34.8397	138.487658
B	E	travelled	176	2-Dec-14	11:03:53	-34.8378	138.484375
B	E	met	299	2-Dec-14	11:06:33	-34.8218	138.490382
B	E	travelled	224	2-Dec-14	11:11:44	-34.824	138.487286
B	E	met	195	2-Dec-14	11:14:30	-34.8349	138.48139
B	C	met	90	2-Dec-14	11:17:02	-34.8422	138.48376
B	C	travelled	192	2-Dec-14	11:19:19	-34.8436	138.483285
B	C	travelled	100	2-Dec-14	11:25:43	-34.8525	138.477153
B	C	met	105	2-Dec-14	11:27:24	-34.8528	138.477258
D	F	crossed	2	2-Dec-14	11:28:38	-34.8493	138.480335
B	D	met	146	2-Dec-14	11:30:05	-34.853	138.477414
B	D	travelled	362	2-Dec-14	11:31:26	-34.8529	138.477461
E	F	crossed	2	2-Dec-14	11:33:43	-34.8518	138.491087
A	E	met	414	2-Dec-14	11:37:49	-34.8528	138.499275
B	D	met	138	2-Dec-14	11:39:05	-34.8467	138.495084
B	D	travelled	250	2-Dec-14	11:40:18	-34.8468	138.49502

Figure 2. Transactions Listed in Chronological Order.

Entity	Entity	Transaction	Duration(s)	Date	Time	Latitude	Longitude
B	E	crossed	2	2-Dec-14	10:46:23	-34.846	138.5027
B	E	travelled	133	2-Dec-14	10:51:17	-34.8403	138.4967
A	C	met	98	2-Dec-14	10:53:17	-34.8447	138.4779
B	E	met	228	2-Dec-14	10:59:53	-34.8397	138.4876
B	E	travelled	94	2-Dec-14	11:02:52	-34.8397	138.4877
B	E	travelled	176	2-Dec-14	11:03:53	-34.8378	138.4844
B	E	met	299	2-Dec-14	11:06:33	-34.8218	138.4904
B	E	travelled	224	2-Dec-14	11:11:44	-34.824	138.4873
B	E	met	195	2-Dec-14	11:14:30	-34.8349	138.4814
B	C	met	90	2-Dec-14	11:17:02	-34.8422	138.4838
B	C	travelled	192	2-Dec-14	11:19:19	-34.8436	138.4833
B	C	travelled	100	2-Dec-14	11:25:43	-34.8525	138.4772
B	C	met	105	2-Dec-14	11:27:24	-34.8528	138.4773
B	D	met	146	2-Dec-14	11:30:05	-34.853	138.4774
B	D	travelled	362	2-Dec-14	11:31:26	-34.8529	138.4775
B	D	met	138	2-Dec-14	11:39:05	-34.853	138.4774
B	D	travelled	250	2-Dec-14	11:40:18	-34.8529	138.4775
A	E	met	414	2-Dec-14	11:37:49	-34.8528	138.4993

Figure 3. Transactions Clustered According to Entity Pairs.

More complex networks can indicate the strength of each tie. Here, the matrix is not merely bounded to values of 0 and 1, but can take greater integers. These networks are refereed to as weighted social networks and are depicted graphically by adding the weights to each edge. An example of a weighted adjacency matrix below corresponding to a network is shown in Figure 4. Software like Palintir™can also be used to represent this data in a more visually attractive manner. Examples are shown in Figures 5 and 6.

TRACK-BASED SOCIAL NETWORKS. In this instance, using Palintir™, a social network is constructed based on the clustered physical transactions observed earlier (Figure 5).

The factors at arriving at an estimate are the type of interaction that occurs between two entities (for which a weight factor w is applied) and the number of times each takes place. The estimate (E) between two entities (i, j) is given by $E_{i,j} = \sum_k N_k^{(i,j)} w_k$, where $N_k^{(i,j)}$ are the number of interactions between entities i and j for interaction k, and w_k denotes the weight for interaction k. A meeting or travelling together is given a greater

weighting; a value of 1, compared to a crossing which has a weight of 1/2 [1] This is analogous to Table 7.3 in (Antony, 2016) where certain verb phrases were ranked. For instance, for the entitiy pair B and E, there is 1 crossing, 3 meetings and 4 times where they travelled together (Figure 5). This means that their final overall weight is 7.5. Doing this for all entity pairs, the adjacency matrix is thus given as a matrix E with Figure 6 diagramatically showing the social network.

Obtaining a social network graph in this way must be qualified. The graph is merely indicative of the possible strength of the relationship between any two entities for a given activity or task that is being observed. For instance, for the activity observed on the second of December of 2014, entity B was the most connected but this result could be different for a different activity. If the roles of each entity were also known, then this might give some insight into the type of activity that is being planned. In this case the scenario involved the movement of a package between different people. If the package is one of the entities, then it is reasonable to conclude that entity B was the package as it was the most connected. Its movements can be backtracked or the motive of the group might be inferred. [2]
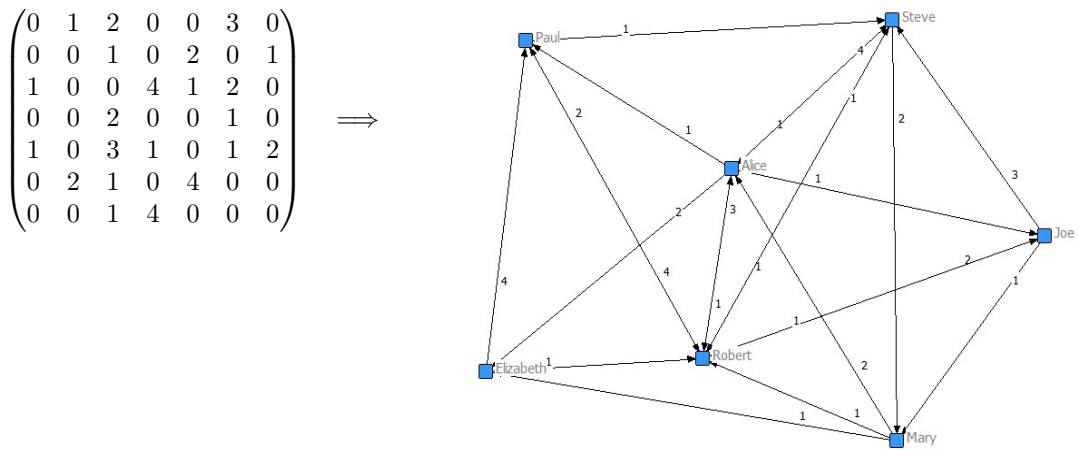
$$\begin{pmatrix} 0 & 1 & 2 & 0 & 0 & 3 & 0 \\ 0 & 0 & 1 & 0 & 2 & 0 & 1 \\ 1 & 0 & 0 & 4 & 1 & 2 & 0 \\ 0 & 0 & 2 & 0 & 0 & 1 & 0 \\ 1 & 0 & 3 & 1 & 0 & 1 & 2 \\ 0 & 2 & 1 & 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 4 & 0 & 0 & 0 \end{pmatrix} \implies$$

Figure 4. Example of Adjacency Matrix and Corresponding Weighted Social Network.
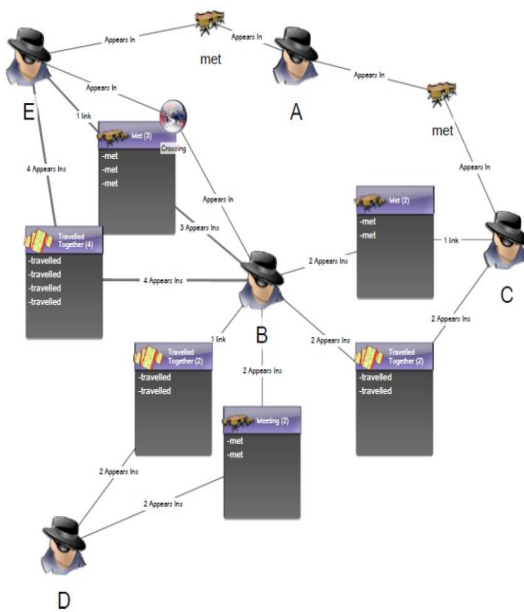

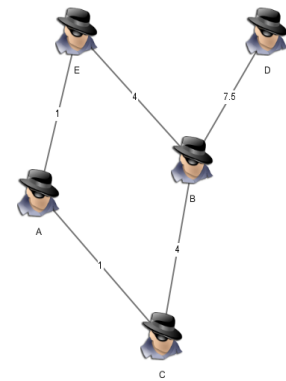**Figure 5.** Graphical Representation of Entities and Their Interactions.

Figure 6. Representation of Entities' Relationships with One Another with Their Weights.

---

[1] It could also be argued that travelling together should carry a greater weight than a meeting; because being in close proximity to another subject for two minutes would generally seem a weaker indicator than traveling together, which implies some degree of common purpose.

[2] Another hint to determine which entity was the package is the fact that it only moved when someone was traveling with it .

4 CONCLUSION

In Pimento, the aim was to develop and formulate some concepts that could be useful in understanding ABI. However, in that trial, there were two distinct social network groups. In Jalapeño, the scenario consisted of a single group of people carrying out a set of activities. Taking only the GPS measurements of the Jalapeño data, the goal was to determine interactions like potential meetings and cases where two or more entities might have travelled together so that a social network could be calculated. In this regard, the techniques showed here were successful in determining which GPS unit corresponded to the package, as it was assumed that this would be the most connected. This can be taken as first approximation of a social network. Here GPS tracks were used so that not only basic kinematic information was derived ("What") but also some contextual ("Why") data could be gleaned. Future work could incorporate other sources of data like social media, phone transactions, etc which would help refine this first social network approximation.

REFERENCES

Antony, R. T. (2016). Data Fusion Support to Activity-Based Intelligence. Artech House.

Biltgen, P. and Ryan, S. (2016). Activity-Based Intelligence: Principles and Applications. Artech House.

Haspert, J. K. (2000). Optimum ID Sensor Fusion for Multiple Target Types. *IDA Document D-2451*.

Palintir (2015). https://www.palintir.com. Visited 01 Mar 2016.

Phillips, M. *A Brief Overview of ABI and Human Domain Analytics*, Trajectory magazine, NGA case number 12-463, Sept. 28, 2012. Online: http://trajectorymagazine.com/web-exclusives/item/1369-human-domain-analytics.html.

Wasserman, S. and Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge University Press.

Weller, K. and Kinder-Kurlanda, K. (2015). Uncovering the Challenges in Collection, Sharing and Documentation: The Hidden Data of Social Media Research?, *In Standards and Practices in Large-Scale Social Media Research: Papers from the 2015 ICWSM Workshop*. Proceedings Ninth International AAAI Conference on Web and Social Media, 28-37.