

Co-presence Communities: Using pervasive computing to support weak social networks

Jamie Lawrence Terry R. Payne David De Roure

Intelligence, Agents and Multimedia Group
School of Electronics and Computer Science
University of Southampton, UK

E-mail: {je103r, trp, dder}@ecs.soton.ac.uk

Abstract

Although the strongest social relationships feature most prominently in our lives, we also maintain a multitude of much weaker connections: the distant colleagues that we share a coffee with in the afternoon; the waitress at our regular sandwich bar; or the ‘familiar stranger’ we meet each morning on the way to work. These are all examples of weak relationships which have a strong spatial-temporal component but with few support systems available. This paper explores the idea of ‘Co-presence Communities’ - a probabilistic definition of groups that are regularly collocated together - and how they might be used to support weak social networks. An algorithm is presented for mining the Co-presence Community definitions from data collected by Bluetooth-enabled mobile phones. Finally, an example application is introduced which utilises these communities for disseminating information.

1 Introduction

Despite their chaotic appearance, our lives are full of temporal patterns: the morning commute, the normalised work hours, the weekly project meeting, the ‘lunch hour’, and the Friday after-work drinks are just some of the commonly-held patterns. Another attribute of modern life is the ubiquity of weak, incidental, relationships and ‘familiar faces’ we see each day. Where these informal relationships and temporal patterns collide, there exists regular social groupings which are typically unnoticed, unsupported, unexplored and undervalued. For example, in the early 1970’s, Stanley Milgram performed a small study to investigate a phenomenon he called *Familiar Strangers*. Milgram defined a familiar stranger as someone who is observed,

repeatedly for a certain time period and without any interaction [15]. Milgram’s study of a railway station (frequented by commuters) revealed that 89.5% of the participants could identify at least one person which they recognised but had never spoken to. These seemingly trivial relationships are not terribly valuable during the normal course of life but they become essential during a crisis.

The corporate setting can be thought of as a miniature version of the urban environment. Many of the informal relationships that are seen in the city (such as friends, neighbours, and familiar strangers) can also be found within the boundaries of an organisation. There are also informal knowledge-sharing networks called ‘*Communities of Practice*’, which can be thought of as people bound together by their “*common sense of purpose and a real need to know what each other knows*” [2]. These communities of practice provide an invaluable source of knowledge sharing that cuts across the strict boundaries of a hierarchical company structure. An essential ingredient which allows these communities to thrive is the actual physical interaction of its participants. Whether it is technicians swapping stories over lunch or hanging around the coffee pot, the social interaction is a critical component of knowledge generation and dissemination within an organisation.

We define a **Co-presence Community** as a group of individuals who regularly share the same location at the same time, i.e., they are repeatedly collocated together. These communities are a generalisation of other informal social networks based on physical locality (such as the familiar strangers or communities of practice). In order to discover the co-presence communities, it is necessary to capture instances of co-presence between people. As wireless technologies, such as Bluetooth, are becoming more ubiquitous they provide the ideal mechanism by which users’ devices can discover

each other automatically and unobtrusively. In our investigation, we use Bluetooth-enabled devices (such as phones, PDA’s, etc.) to detect ‘encounters’ between people. A data mining algorithm can then reveal the underlying co-presence communities by extracting instances of repeated co-presence. This automated discovery of co-presence communities is particularly valuable because they provide a computational representation of the fundamental social networks which would otherwise be hidden. The co-presence communities would then prove to be a useful means of providing informal social networks with a mechanism for sharing knowledge (in a similar way to a group mailing list).

Our ongoing research effort aims to show how co-presence encounters, detected by wireless devices such as Bluetooth-enabled phones, can be analysed and clustered to identify (and subsequently augment) the sort of informal social networks discussed above. This paper introduces the concept of a Co-presence Community, in Section 2, and illustrates how these communities could be used to provide an knowledge sharing infrastructure for informal social networks (Section 3). A data mining algorithm is presented in Section 4, which can extract the co-presence communities from individual co-presence events. Finally, a brief example from a simulated data source is presented (Section 5) and the paper concludes with a discussion of the future work in Section 6.

2 Co-presence

In its most general form, co-presence refers to the spatio-temporal conditions under which humans can interact with each other. In 1966, Goffman referred to co-presence as the condition when people “*sense that they are close enough to be perceived in whatever they are doing, including their experiencing of others, and close enough to be perceived in this sensing of being perceived*” [11]. In the telecommunications domain, co-presence typically involves one or more parties collaborating with each other through a virtual representation (e.g. teleconferencing) or within a virtual environment (e.g. a massively multiplayer online role playing game, MMORPG). Zhao [17] attempts to reconcile these differing views by structuring co-presence along two primary axis: proximity (physical or electronic) and corporeal presence (where one, both or neither parties may be present at the physical location). Whilst much of this paper may be applicable to other modes of co-presence, we shall typically be referring to “*corporeal copresence*” from Zhao’s taxonomy — the most natural state of co-presence, where all parties are physically proximate and present at the same site. In our

use of co-presence, we do not require that the parties actually engage in any direct interaction — they may just be walking past each other or sitting on a bus together. The key point is that the spatial and temporal conditions must exist such that it is *possible* for the participants to interact.

2.1 Definitions

When two parties are co-present, a **co-presence relationship** can be said to exist between them. This relationship is symmetric but not transitive, and may exist with multiple parties simultaneously:

- if $a \xrightarrow{\text{copresent}} b$ then $b \xrightarrow{\text{copresent}} a$.
- if $a \xrightarrow{\text{copresent}} b$ and $b \xrightarrow{\text{copresent}} c$ it does not imply that $c \xrightarrow{\text{copresent}} a$
- if $a \xrightarrow{\text{copresent}} b$ it is also possible that $a \xrightarrow{\text{copresent}} c$

A **co-presence event** is simply the instantaneous representation of a co-presence relationship, i.e., the co-presence relationship at an instant of time.

2.2 Co-presence Detection

Corporeal co-presence is typically experienced by humans through their sensory inputs, but for computational systems to understand co-presence, a suitable input technology is required. For co-presence to remain socially meaningful, we take Goffman’s view that the co-present individuals need to be able to directly sense each other. We can draw on the work of Hall [12] in defining the boundaries of the human senses and, in particular, the range at which we can perceive people as individuals: facial expressions, hair styles and age become apparent at a distance of 30m, and at around 20-25m we can determine their feelings and mood. These distances broadly correspond with the effective range of Personal Area Networks (PAN), such as Bluetooth, which makes them an useful technology for detecting co-presence. Other practical approaches to co-presence detection include the use of wearable infra-red transceivers [3], or location sensors and a centralised service to correlate these readings from multiple subjects in space-time. This use of sensors attached to an individual allows the computation into be embodied within the world [5]. This embodiment means that through the medium of real-world interactions, the user can train the system about their co-presence patterns, without resorting to activities outside of their normal daily routine. This aligns well with Dourish’s argument that embodiment “*does not simply mean ‘physical manifestation.’ Rather, it means being grounded in*

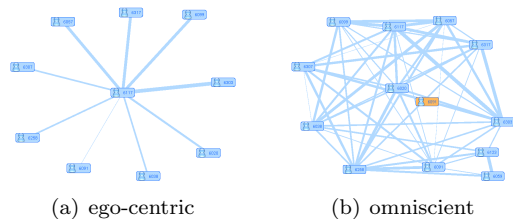


Figure 1. An ego-centric and omniscient views of the co-presence relationships between participants in the Haggie dataset.

everyday, mundane experience... [embodiment] is the property of our engagement with the world that allows us to make it meaningful”.

The sensing technology will usually influence how the co-presence data is scoped. In an *ego-centric* view (Figure 1(a)), only the co-presence data from a single person is available, whereas an *omniscient* view (Figure 1(b)) implies a total knowledge of all co-presence events within the population. Wireless sensors (radio or infra-red) will generally produce ego-centric co-presence data whereas a centralised location correlation service is capable of producing both ego-centric or omniscient data.

One of the limitations of computational co-presence detection is that it requires each person to make themselves visible to the sensors — either by carrying a particular device, activating a software application on their mobile phone or connecting to a centralised service. In most non-trivial experiments, only a very small proportion of the real-world co-presence events can be captured computationally because one or more of the parties are not visible to the sensors. These undetectable people are the ‘*dark matter*’ of the co-presence network. This dark matter is not a critical flaw in the use of co-presence but obviously the higher the density of people carrying sensors, the more accurate and useful a co-presence system will be.

A **Co-presence Community** is a group of individuals that are regularly co-present at approximately the same time. These communities are purely derived from co-presence data and explicitly lack any social context, but it is expected that this contextual information can be added more accurately by the user. However, our hypothesis is that repeated collocations of individuals indicates a valuable underlying social relationship. Even if the underlying cause of the co-presence is just that the participants take the same route to work, they share, if nothing else, a common interest in the functioning of the local area (safety, cleanliness, punctual-

ity) [13]. These co-presence communities can be represented as the set of community members, and the start and end times of the period when the community is active. The stability of a community refers to how predictable the community’s characteristics are. For example, the co-presence community formed by weekly project meetings would have stable membership and temporal dimensions. In contrast, the community formed on a commuter bus might have a stable temporal dimension (because you always catch the same bus) but an unstable membership (because it’s not always the same people on that bus). By obtaining these community memberships via wireless sensors, it should be possible to build a dynamic, self-updating model of the people that a user is typically near.

3 AIDE: Ambient Information Dissemination Environment

The Ambient Information Dissemination Environment (AIDE) is an application that demonstrates the utility of co-presence communities, by allowing users to share content with their Bluetooth-enabled mobile phones. The user can specify different rules for sharing content, depending on the current social context (i.e., the co-presence community). The content could range from recently-read online articles, cartoons and jokes, photos of the local area, new journal publications, or the details of upcoming events. Naturally, not everyone is interested in the same content and, just like human conversations, there is time and place for sharing everything: confidential work discussions aren’t appropriate in the bar and movie reviews aren’t an accepted part of most meetings. Co-presence communities would be utilised to direct and filter the content and bring this into a socially-meaningful context. A similar concept was used in the “Serendipity” application [6], which used co-presence to facilitate interest match-making and introductions¹. The advantage of co-presence communities for this application is that they would allow dissemination in different contexts: sharing jokes with strangers on the commute home (i.e., a time-stable community) or disseminating the latest call-for-proposals to a research group (a fully stable community formed by weekly meetings). AIDE would ambiently determine which community a detected device belongs to and transfer the selected content in the background, based on the source user’s specified preferences. It seems likely that a filtering and ranking mechanism would be required to reduce the informa-

¹However, Serendipity used single incidents of co-presence, rather than historic data, to manage these introductions

tion overload caused by collecting content during each period of co-presence.

The primary goal behind AIDE is to facilitate knowledge sharing within weakly-connected groups that otherwise would not have any means of disseminating information. In essence, it would function like an automatic mailing list where the subscriptions are based on the co-presence relationships rather than an explicitly-defined membership list. It is certainly not designed to replace the social interactions that exist within a community but, instead, augment and encourage them. Other potential co-presence community applications might include: building an online social network using real-world co-presence data [4]; determining the interruptibility of someone based on who they're co-present with (similar to Begole et al's work on rhythm modelling [1]); creating an awareness of the user's social surroundings through introspective visualisations of the communities (like Encounter Bubbles²); or even routing messages between devices using the temporal and membership characteristics of co-presence communities (similar to [14]).

4 Co-presence Community Miner

The core problem when discovering co-presence communities is one of clustering together historical co-presence events. A relatively simple approach has been taken which combines a feature detection algorithm and COBWEB [8], a well-known incremental clusterer. Essentially, the algorithm creates discrete intervals of co-presence, uses an edge detection technique to find continuous periods when a similar set of devices were co-present, and then clusters these periods across all historic data (outlined in Algorithm 1). The resulting clusters are the definitions of co-presence communities.

Algorithm 1 The Co-presence Community Mining algorithm

```

loop
  batch ← daily batch of co-presence data
  intervals ← discretisation of batch {see 4.1}
  periods ← extract features in intervals {see 4.2}
  for all periods do
    update clusterer with period {see 4.3}
  end for
end loop

```

²<http://www.seansavage.com/encounter-bubbles/>

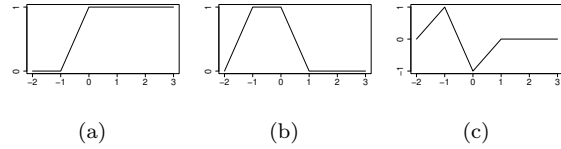


Figure 2. This figure illustrates how the Laplacian operator transforms a signal so that the major transitions (edges) are represented by zero-crossings

4.1 Discretisation

The first process is to segment the daily batches of co-presence data into equally-sized intervals, thereby assigning each co-presence event to a discrete time interval. The purpose of this stage is to normalise the data and provide initial groupings of devices. It is particularly important if the co-presence events were captured on a continuous basis, or from different data sources. I is the sequence of intervals, $i_x, i_{x+1}, \dots, i_{x+n}$ where n is the number of intervals in the batch. Each interval, i_x , consists of the set of devices co-present during the time $[x * interval, (x + 1) * interval]$ where $interval \geq$ to the sampling resolution of the data.

4.2 Feature Extraction

The next stage extracts periods of similar device membership using a combined smoothing and edge detection technique. Essentially, the algorithm needs to smooth the interval's membership (to provide some resistance against measurement errors) and then determine significant changes in membership. This can be succinctly calculated using the *Laplacian of Gaussian* (*LoG*) operator:

$$LoG(x) = \frac{1}{\sigma^2} \left(\frac{x^2}{\sigma^2} - 1 \right) e^{-\frac{x^2}{2\sigma^2}}$$

where σ is the standard deviation used as the Gaussian smoothing parameter. The *LoG* operation combines a Gaussian smoothing process with the second derivative of the signal (the Laplacian) to produce a corresponding signal where each zero crossings represents an edge. For example, Figure 2(a) shows a binary signal with a transition from 0 to 1 between -1 and 0 ; Figure 2(b) shows the first derivative of this signal with the transition represented by the peak between -1 and 0 ; Figure 2(c) is the Laplacian of 2(a) and the transition is represented by the zero-crossing at -0.5 .

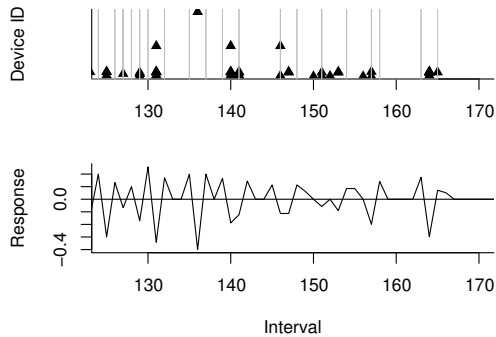


Figure 3. The co-present devices, the period boundaries (in grey) and the corresponding response signal

A ‘response’ for each interval, i , is calculated by averaging the *LoG* of each co-present device, d , across a smoothing window of w :

$$response(i_x) = \frac{\sum_d^{|i_x|} \sum_{j=1-\frac{w-1}{2}}^w \begin{cases} LoG(j), & \text{if } d \in i_{x+j} \\ 0, & \text{otherwise} \end{cases}}{|i_x|}$$

where ($|i_x|$ is the number of devices co-present during interval i_x).

The zero-crossings of *response* are detected by scanning across the signal and checking for changes in sign. The actual position of the zero-crossing may fall between interval boundaries so, to account for this, the value which is closest to zero is chosen as the zero-crossing. For example, if $response(x) = 0.3$ and $response(x+1) = -0.2$, the zero-crossing will occur at $x+1$. The locations of these zero-crossings are used as the start and end boundaries for periods, where each period collects together a sequence of similar intervals. Figure 3 shows a small portion of the co-presence events from the Reality Mining dataset [7], the result of the *LoG* transformation and the corresponding period boundaries. As can be seen in this figure, the period boundaries correctly delineate the major changes in co-presence membership.

4.3 Clustering

The previous step has constructed a set of periods, each of which collects together contiguous intervals of similar co-presence events. The aim of this step is to cluster together those periods with a similar start time, end time and membership — thereby constructing the definitions of co-presence communities. The

periods may be clustered on numerical (start and end times), nominal (day of the week) and set (the membership of a period) attribute types. These attribute types, and our requirement for incremental learning, effectively rule out many of the popular clustering algorithms such as *K-means*, *Finite Mixture Models* or *Expectation-Maximisation* approaches. Instead, a class of clustering algorithms known as *Concept Formation* algorithms are more appropriate because of their incremental nature and support for several attribute types. COBWEB is a well-known conceptual learning algorithm that has been applied in numerous domains and has formed the basis for several other clustering algorithms [8, 16]. It incrementally constructs a hierarchy of concepts using four functions (incorporation, creation, merge, split) which are controlled by the *Category Utility* (CU) metric [10]. In practice, maximising the category utility will maximise the similarity of instances within a cluster and the dissimilarity between clusters. As originally proposed, the COBWEB clusterer only operates on nominal attributes but an extension, called CLASSIT, provides support for numerical attributes by calculating the CU of numerical attributes using a normal distribution and standard deviation [9]. A simple modification of the original CU formula is proposed to account for sets of nominal values (i.e., the devices co-present during a period). For a set attribute, S_i , with members m_1, m_2, \dots, m_n , the category utility can be defined as:

$$CU_{set}(C_\ell) = \sum_i \sum_n (Pr[m_n \in S_i | C_\ell]^2 - Pr[m_n]^2)$$

which is simply the nominal category utility summed across all members of a set attribute.

From the input periods, COBWEB will produce a hierarchical tree of clusters. Each cluster provides a probabilistic representation of the co-presence periods and therefore describe the user’s co-presence communities.

5 Validation

To provide a synthetic source of well-behaved co-presence data, an agent-based simulation, PedSim, has been developed. This simulation allows co-presence events to be collected from a simulated population of agents performing various movement behaviours such as a random walk or oscillation pattern. These behaviours can be perturbed by a random noise which will allow the controlled analysis of the Co-presence Community Miner. Figure 4 shows the results of running the algorithm on the co-presence events produced by a simple ‘home-work’ oscillation pattern. This produces

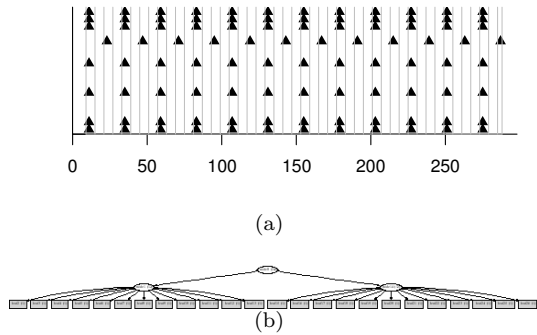


Figure 4. The output from a perfectly regular simulated dataset: (a) The input co-presence data and period boundaries; (b) The cluster hierarchy produced by the algorithm, with the nodes representing the discovered co-presence communities.

a pattern whereby each agent encounters the same two sets of other agents (whilst travelling to ‘work’ and back), 12 times a day. Furthermore, once a full homework cycle has been completed, no new communities should be found (since each day is exactly the same). The resulting COBWEB hierarchy (Figure 4(b)) shows the 24 identified communities directly corresponding to those expected from movement pattern.

6 Conclusions and Future Work

This paper has introduced **Co-presence Communities** — a novel concept of regular collocation occurrences between a group of individuals. These communities can be discovered by mining a set of raw co-presence events captured by Bluetooth-enabled mobile phones, and an algorithm for achieving this has been outlined. The generated community definitions are expected to be useful in identifying informal social networks within corporate or urban environments, and providing a means of supporting information dissemination within these groups. The AIDE application is one such service which will utilise the co-presence community definitions to provide an ambient knowledge dissemination channel within regular, but informal, social networks.

Currently, only the most preliminary analysis of the algorithm’s performance has been performed. The next phase will evaluate the algorithm’s performance by using more complicated stochastic simulations and real-world co-presence data. Furthermore, an application trial is scheduled to start in Summer 2006 that will

evaluate the utility and performance of co-presence communities in a real-world scenario.

References

- [1] J. B. Begole, J. C. Tang, and R. Hill. Rhythm modeling, visualizations and applications. In *UIST ’03: Proceedings of the 16th annual ACM symposium on User interface software and technology*, pages 11–20, New York, NY, USA, 2003. ACM Press.
- [2] J. S. Brown and E. S. Gray. The people are the company. *Fast Company*, 1:78, November 1995.
- [3] T. Choudhury and A. Pentland. The sociometer: A wearable device for understanding human networks. In *Computer Supported Cooperative Work - Workshop on Ad hoc Communications and Collaboration in Ubiquitous Computing Environments*, 2002.
- [4] S. Counts and J. Geraci. Incorporating physical co-presence at events into digital social networking. In *CHI, Late Breaking Results*, 2005.
- [5] P. Dourish. *Where the Action Is*. MIT Press, 2001.
- [6] N. Eagle. Can serendipity be planned? *MIT Sloan Management Review*, 46(1):10–14, Fall 2004.
- [7] N. Eagle and A. Pentland. Reality mining: Sensing complex social systems. *Journal of Personal and Ubiquitous Computing*, pages 1–14, February 2006.
- [8] D. Fisher. Knowledge acquisition via incremental conceptual clustering. *Machine Learning*, 2:139–172, 1987.
- [9] J. H. Gennari, P. Langley, and D. Fisher. Models of incremental concept formation. *Artificial Intelligence*, 40:11–61, 1989.
- [10] M. A. Gluck and J. E. Corter. Information, uncertainty, and the utility of categories. In *Proceedings of the Seventh Annual Conference of the Cognitive Science Society*, pages 283–287. Lawrence Erlbaum Associates, 1985.
- [11] E. Goffman. *Behaviour in Public Places*. The Free Press, 1966.
- [12] E. T. Hall. *The Hidden Dimension*. Anchor Books, 1969.
- [13] J. Jacobs. *The Death and Life of Great American Cities*. Jonathan Cape, 1962.
- [14] S. Merugu, M. Ammar, and E. Zegura. Routing in space and time in networks with predictable mobility. Technical Report GIT-CC-04-07, College of Computing, Georgia Institute of Technology, 2004.
- [15] S. Milgram. *The Individual in a Social World*, chapter The Familiar Stranger: An Aspect of Urban Anonymity, pages 51–53. Addison-Wesley, 1977.
- [16] K. Thompson and P. Langley. Concept formation in structured domains. In D. H. Fisher, M. J. Pazzani, and P. Langley, editors, *Concept Formation: Knowledge and Experience in Unsupervised Learning*. Morgan Kaufmann, 1991.
- [17] S. Zhao. Toward a taxonomy of copresence. *Presence*, 12(5):445–455, October 2003.