# QUANTIFYING THE EFFECTS OF SPACE ON ENCOUNTER

097

**Vassilis Kostakos**
*Department of Computer Science, University of Bath*
**Eamonn O'Neill**
*Department of Computer Science, University of Bath*

**Keywords:**

Human encounter
Human behaviour
Bluetooth
Visiting pattern

**Vassilis Kostakos**
*Department of Computer Science, University of Bath, Bath BA2 7AY, UK*
*vk@cs.bath.ac.uk*
**Eamonn O'Neill**
*Department of Computer Science, University of Bath, Bath BA2 7AY, UK*
*eamonn@cs.bath.ac.uk*

## Abstract

In this paper we present initial results from our attempts at quantifying the effects of space on human encounter. Our premise is that distinct locations across a city shape and form people's behaviour and encounters. By recording and analysing individual's visiting behaviour and encounters, we can contribute to quantifying our understanding of space and its effect on people.

We were able to record individual visiting patterns by using Bluetooth technology, which is embedded in many mobile handsets. Each phone, if configured appropriately, beams out a unique serial number using Bluetooth, typically within a 10-meter range. Recording this unique serial number, in conjunction with the date and time, can be used as an indicator of when someone visited a specific location.

To analyse our data we look for patterns of co-presence over time. We are interested in identifying patterns in the way people encounter each other at each of the four locations of our study. In our analysis we were able to correlate data features from four distinct locations. Our analysis can be used to explain how the different locations provide different opportunities for encounter.

## Introduction

Our society regenerates through people's movement and encounter (Hillier & Hanson, 1984). A vibrant and sustainable environment gives us the opportunity to meet new people, make new friends. Walking down the street or visiting a pub are opportunities for such encounters. Space syntax maintains that people's movement, and subsequently encounters with others, is affected by the structure of space (Hillier & Hanson, 1984; Hillier et al., 1987). Demonstrating a relationship between structure and movement has been achieved with some success (Hillier et al., 1993), yet a relationship between spatial structure and people's encounters has not been demonstrated at the individual granularity.

Although movement, at the aggregate level, is relatively easy to measure (e.g. using gatecounts), the same is not true of encounters.

Before the advent of mobile technology it was extremely difficult, if not impossible, to measure and record a large number of individuals' encounters with others in the course of everyday life. In this paper we describe how we made use of Bluetooth technology, typically found in mobile devices, to record people's visiting patters and encounters in space. We then show how this data can be analysed to give us insight into the nature of people's encounter and the effect of space on encounter.

## Setup

097-02 In our study we made use of Bluetooth technology typically found in mobile devices. Bluetooth technology has a characteristic that renders it appropriate for study by methods derived from those of space syntax. In contrast to the wireless signals emitted by typically static WiFi access points, the signals emitted by Bluetooth devices map very closely to the movements of people around the city, which in turn are a primary concern of space syntax.

Our basic setup, replicated across 4 sites, involved installing a computer that constantly recorded the number of Bluetooth devices within a 10-meter range. This data allows us to correlate pedestrian movements with Bluetooth device movements, providing baseline data about the penetration of Bluetooth into city life. In previous work, we found that approximately 7.5% of observed pedestrians had discoverable Bluetooth devices (O'Neill et al., 2006). This number most certainly varies between different cities, but still it shows that a considerable portion of the public can be recorded using our method. Beyond simply counting the appearance of Bluetooth devices, we used this method to uncover interesting data on patterns of presence of Bluetooth devices and Bluetooth device names (O'Neill et al., 2006), as well as patterns of encounter, described in this paper.

In our study we considered four locations, which we shall refer to as

- campus
- street
- pub
- office

The first two locations are outdoor pedestrian streets, one on our campus and one in the city of Bath, which connect open spaces and can be considered as gates. The latter two are indoor locations where visitors typically spend some time in them. The pub is open to anyone over the age of 18, while the office is a secure location where only employees and their visitors have access.

Although the first two locations would be considered as gates in the Space Syntax term, the nature of Bluetooth technology mitigates against this. The 10-meter range of our Bluetooth scanner reached beyond walls, and in adjacent offices. Effectively, if our scanner picked up a Bluetooth device, there is no way of knowing if that device was on the street, or in any of the offices. Despite this, on aggregate level we still get quite distinctive patterns of data between the first two and last two locations, as we describe in the next sections. This is because the great majority of devices our scanners picked up was indeed on the street (for the first two locations).

Across the four locations we captured 6 months of data, with approximately 10000 unique devices. In the following section we describe in detail the data we captured and the analyses we carried out.

## Data & Analysis

The method we used to scan for Bluetooth devices generates discrete data about the presence of devices in the environment. A visualisation of our data, which we have termed timeline, can be seen in Figure 1. Here, each dot represents a discovery event, i.e. a point in time (x-axis) when our Bluetooth scanner picked up a specific device in the environment. By applying filters, we can see that, for example, device 16 was present in the environment between approximately 18.5 minutes and 19.5 minutes.

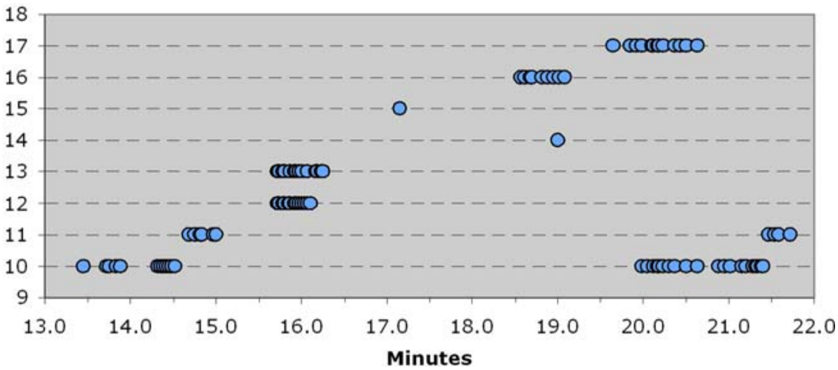



***Figure 1:***

*A timeline visualization of our Bluetooth gatecounts. Each device is given its own timeline (dashed lines) and each discovery event is plotted as a circle on the timeline*

097-03

To study the patterns of co-presence in our data, we first need to identify instances where two or more devices were present at the same place and the same time. For example, in Figure 1 we see that devices 12 and 13 encountered each other. We developed filters that analysed our data and gave us instances of devices encountering each other at each of the four locations in our study. These initial results took the form of records:

device1_id, device2_id, location

At this stage in our analysis we had a long list of such records, describing which devices encountered each other and in which location. For example, in Figure 1 we see that devices 12 and 13 encountered each other at 15.5 minutes and were together for approximately 1 minute. This list of encounters is a textual representation of the patterns of encounter across our four locations. To further study the patterns and structure hidden within this list, we transformed it to four social network graphs, one for each location. Assuming that each device from our dataset becomes a node in the social graph, then the list of encounters indicates which nodes are connected. Proceeding in this manner, we generated four social network graphs, one for each location.

For illustration purposes, in Figure 2 we show the graph from the pub location in our study. In this graph, each device is represented as a node in the graph, and connected nodes indicate that these devices encountered each other at some point. We see that most devices are linked to the main core, whilst some devices are islands. The latter indicates cases where a device was seen only by itself and never in the presence of others. Additionally, the size of nodes represents the total amount of time that a device has spend in this location, while the colour of the nodes (blue to red) indicates the betweeness of a node (from 0 to 1 respectively).

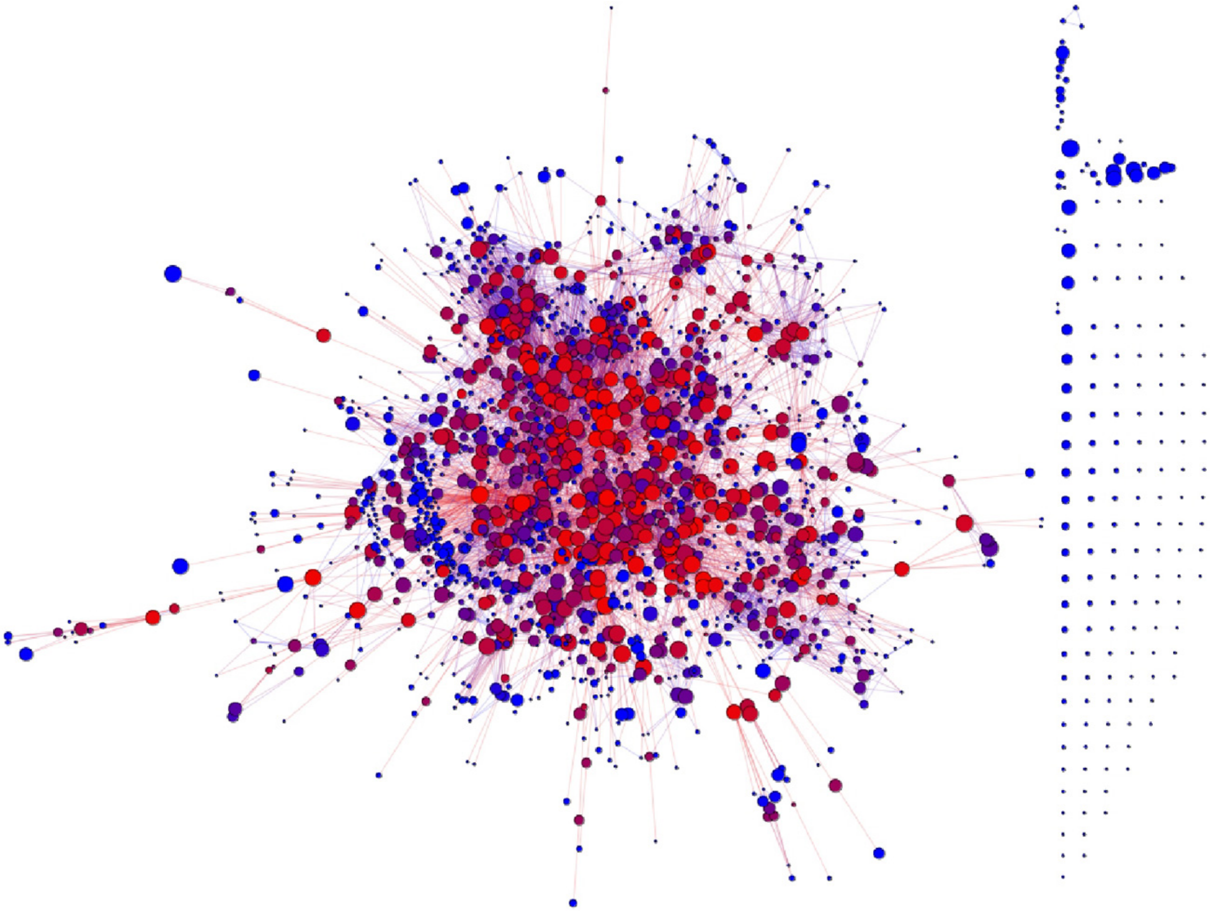
One of our initial observations was that due to the sheer number of nodes in the graphs, the visualisations themselves helped little in analysing our data because of the visual clutter. However, by transforming our data into graph form, we were able to run a number of well-established analysis algorithms using existing software (e.g. Pajek, Ucinet). Specifically, we analysed each of our four graphs in terms of;

- Degree centrality, calculated as the number of neighbours of each node.
- Closeness centrality (access), calculated for any given node as the number of nodes (minus 1) divided by the sum of all distances between the node and every other node.
- Betweeness centrality (control), calculated for any given node as the proportion of shortest paths between all pairs of nodes that include this node.
- Distance, calculated as the probability that the shortest path between a random pair of nodes will be 1, 2, 3, etc.

***Figure 2:***

*A graph visualization of the encounters that we recorded at one of the locations in our study*

097-04



The degree and closeness centrality are measures of the reachability of a node within a network, and describe how easily information can reach a node. Betweeness centrality indicates the importance of a node, and the extent to which it is needed as a link in the chains of contacts that facilitate the spread of information within the network. We should also note that closeness centrality is referred to as "integration" in Space Syntax literature, while betweeness centrality is known as "choice".

## Quantifying Encounter

To gain an overview of the structural properties of the graphs representing encounter, we calculated the metrics shown in Table 1. For each of our locations we calculated the number of unique devices that were recorded by our Bluetooth scanner, the size of the largest core in the encounter graphs, the number of edges in the largest core, the density of the largest core as well as the size of the 2nd largest core. We also calculated some generic centrality measures for each of the largest cores: network degree, closeness and betweeness centralisation. Finally, we measured the maximum and average degree of each graphs, the longest shortest-path distance in each of the graphs, as well as the average shortest-path distance.

***Table 1:***

*Metrics for each of our four graphs*

| | Campus | Street | Pub | Office |
|---|---|---|---|---|
| Unique devices | 1162 | 8450 | 4175 | 329 |
| Largest core | 1028 | 2738 | 4036 | 318 |
| 2nd largest core size | 2 | 4 | 2 | 1 |
| Edges in largest core | 6434 | 5060 | 23919 | 2419 |
| Density | 0.5% | 0.007% | 1.4% | 2.2% |
| Network Degree Centralisation | 0.43 | 0.51 | 0.68 | 0.73 |
| Network Closeness Centralisation | 0.49 | 0.55 | 0.66 | 0.65 |
| Network Betweeness Centralisation | 0.36 | 0.65 | 0.57 | 0.27 |
| Max degree | 454 | 1394 | 2758 | 246 |
| Average degree | 12.26 | 3.70 | 11.85 | 15.21 |
| Max distance (diameter) | 6 | 10 | 9 | 4 |
| Average distance | 2.72 | 2.96 | 2.44 | 2.04 |



In addition to the above metrics, for each of degree, closeness and betweeness centrality measures we generated ranked log-log plots. To do this we attached a value (degree, closeness or betweeness) to each node in the graphs (only the core), and then sorted this list in descending order. We then plotted the sorted lists, resulting in three sets of graphs (degree, closeness, betweeness) for each of our four gates. Additionally, we generated a fourth set of graphs, based on the probable distance between any randomly selected pair of nodes. These graphs are shown in Figures 3 to 6.

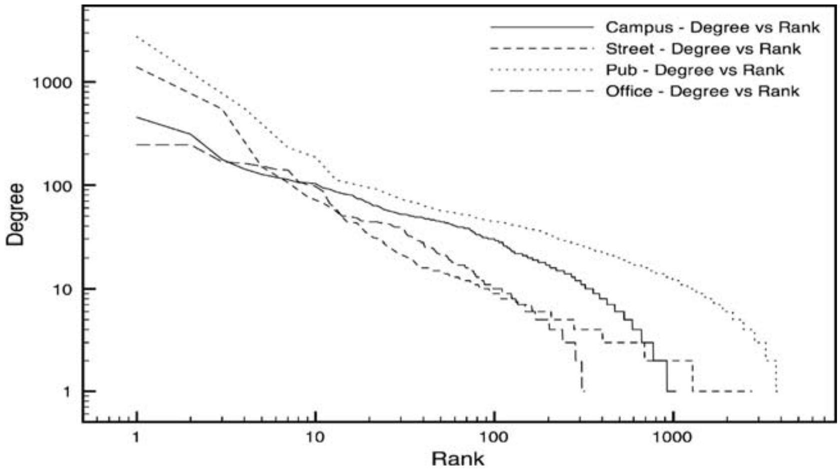



***Figure 3:***

*Ranked log-log plots of degree for each of our four locations*

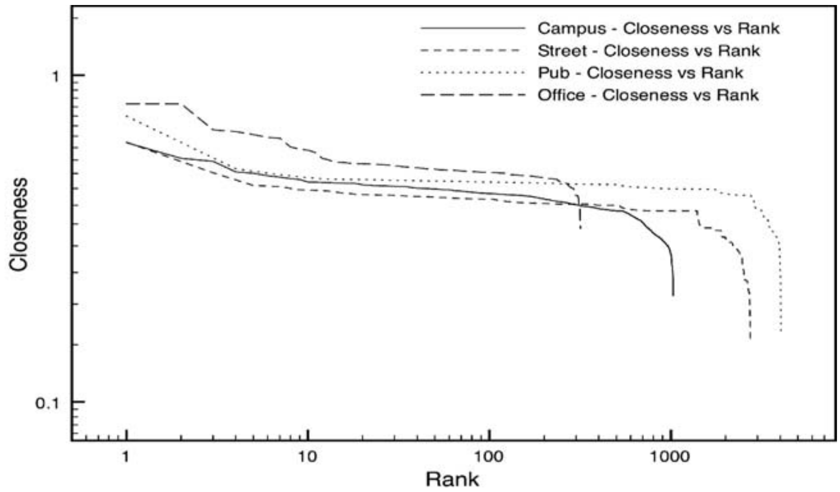



***Figure 4:***

*Ranked log-log plots of closeness for each of our four locations*

*Figure 5:*

*Ranked log-log plots of betweeness for each of our four locations*

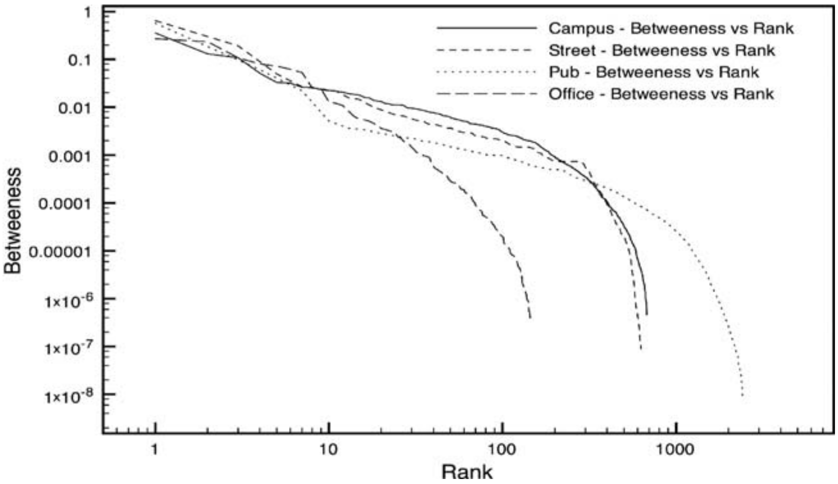
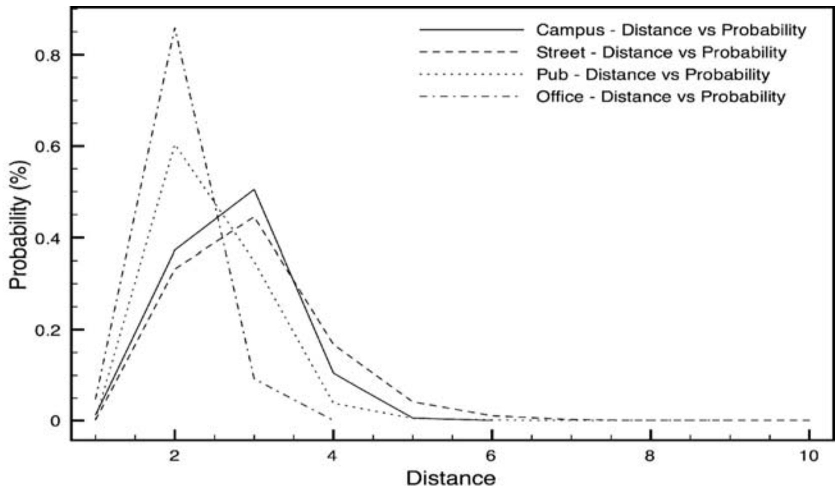





*Figure 6:*

*Probability plots of shortest path distance for each of our four locations*




## Discussion

In our work we set out to measure and quantify the effect of space on encounter. In the previous sections we described our setup, consisting of installing computers that carry out constant Bluetooth scans. We then analysed the captured data to identify instances of co-presence, and thus encounter, within the four locations of our study. These instances of encounter were used to generate social graphs of the community in each of the four locations.

We focus our discussion on the various properties of the social graphs that we listed in Table 1 and in Figures 3 to 6. The way we captured and analysed our data prohibits us from directly identifying the effect of space on encounters and social networks. However, by comparing the properties the social graphs across our four locations we can begin to draw a picture of the communities that inhabit those locations. Also, it is important to keep in mind that in our observations of the four locations the only parameter we changed was the location itself: the hardware, software and algorithms we used to derive our results are identical for all locations. Although it can be argued that our data are affected by a number of further variables, we consider those as part of the location and the environment.

A notable feature of the graphs is their size. As we expected, the city street had the most "visitors", followed by the pub, the campus and the office. This is quite representative of the populations inhibiting each of the locations, since the street is open to everyone, thus likely to get lots of distinct visitors. The pub is also open to everyone (over 18) and again has a large population of potential visitors. The campus, on the other hand, is mostly visited by students and staff, which amount to about 15,000 students and staff (while the population of Bath is about 86,000). Finally, the office is a secure area where only employees have access, thus a small population of potential visitors.

It is interesting to note, however, that the social network of the street consists of about 2/3 islands, with the core consisting of about 1/3 of the devices. Looking at Table 1 we see that the campus has a much higher density than the street. This indicates that there are more static devices on the campus, such as computers or employees phones, which are likely to act as hubs which connect to the core those single devices that go past in the environment. This is something we can verify from Figure 3, where we see the street graph has a few well connected hubs but then falls quite sharply, as opposed to the campus where there are many more nodes with degree between 100 and 5.



It is interesting to note that both locations where the public can go, the street and the pub, have quite large max-degree (1394 and 2758); yet average degree is much smaller on the street than the pub (3.70 and 11.85). In fact, in Figure 3 we see that the pub completely outperforms the street in terms of degree. This is due to the fact that most people in the pub are co-present, thus they get linked together. In other words, a visit in the pub can give someone much more opportunity for copresence than a visit in the street. This is something we expect, as it is the primary purpose of a pub. Also, we should note that in the pub there are certain devices with extremely high degrees, which we believe are attributed to members of staff or regular customers. These act as central hubs that bring together all the customers of the pub into the central core of the social graph. The same is true in the office, where a number of devices have a relatively high degree, indicating that these people come in frequent contact with others.

In general, across the four locations the "tightness" of the communities varies. Specifically, the office and the pub have shorter average distances between their members (2.04 and 2.44 in Table 1 respectively), and we also see in Figure 6 that the probability curves of these two locations are shifted to the left. This is further enhanced by the relatively high density of the pub and the office, which indicates more interactions between the members of the community.

Another interesting point to note is that although the pub has quite a tight and dense population, it has large diameter (9), which is also true of the street (10). Yet, the pub has a smaller average distance (2.44) as opposed to the street (2.96). Coupled with the density measures, we can describe the pub's network as a large central core, while the street's network more closely resembles a small core with a number of branches and additionally a large number of islands.

Considering the network centralisation measures we can make more inferences about the overall structure of the social networks. These measures range from 0 to 1 and indicate a similarity to a perfect linear-shaped network (0) or to a perfect star-shaped network (1). This is calculated for each of degree (DC), closeness (CC) and betweeness (BC).

The office scores high on DC and CC indicating that some nodes can be reached more easily than others, yet BC is low, indicating that all nodes are more or less equally important in terms control and communication. The opposite is true of the pub, where high DC and CC are coupled with high BC. This indicates that there are certain nodes in the pub that act as hubs of communication and control (most likely the members of staff or regular customers).

097-08

Comparing the campus and street in terms of centralisation measures also yields interesting insights. Both have similar levels of DC and CC, but the campus has low BC while the street has high BC. This indicates that on the street there are a few important nodes, while on campus the nodes are more equal.

Finally, we turn our attention to the graphs shown in Figures 3 to 6. We have found that these are much more useful than a visualisation of the social networks themselves. A really interesting observation is that although in each of the 4 graphs the lines have similar shape, the subtle differences are crucial pointers as to the effect of space on encounter. For instance, the variation in how sharply the values fall is a useful indicator, along with the overall steepness of the graphs.

When considering the whole range of values, degree graphs are overall more close to a power law distribution. Closeness graphs have short sharp tails, with a body that approximates a power law extremely well. Similarly, betweeness graphs have long sharp tails, while their body approximates a power law. The distance probability graphs can be approximated by a Poisson distribution.

The graphs we found in analysing our Bluetooth data, point to power-law distributions ($\gamma \approx 0.6$-$1.1$ for degree, $\gamma \approx 1.2$-$1.4$ for betweenness, $\gamma \approx 0.1$ for closeness) that are characteristic of scale-free, or self-similar networks. Such networks imply infinite variance, and usually in such networks there are a few nodes with extremely large number of links. Barabási et al. (1999a) have dubbed such networks 'scale-free', by analogy with fractals, phase transitions and other situations where power laws arise and no single characteristic scale can be defined. These characteristics can be found in kinship networks, physical and biological systems, and economic systems.

Scale-free networks have stimulated a great deal of theorizing. The earliest work is due to Simon (1955), independently rediscovered by Barabási et al. (1999a; 1999b). They show that scale-free networks emerge automatically from a stochastic growth model in which new nodes are added continuously and attach themselves preferentially to existing nodes, with probability proportional to the degree of the target node. Effectively, the richly connected nodes get richer.

We believe that our scanners recorded a phenomenon and process which is quite similar to the "rich getting richer" model, and which explains the presence of power laws in our data. In terms of encounters, those people who have more links and encounters are the ones who are present more in the environment. When a new person comes along, chances are that they are going to encounter the regular customers or the employees. Thus, they share an encounter with an already well-connected person in the graph. It is this exact process that has been shown to result in power-law distributions.

## Conclusion and Ongoing Work

In this paper we describe our attempts to measure and quantify the effect that space has on people's encounters, and ultimately their behaviour. We present a study where four distinct locations were chosen for installing Bluetooth scanners that monitor the presence, and thus encounter, of people in those spaces. Our scanners generated a very rich data set that we used to derive social graphs for each of the four locations.

In our analysis we focused on the derived social graphs, and were able to compare various well-established properties and measurements of social graphs across the four locations. We found that the graphs exhibit power-law distributions when plotting their properties in rank-ordered graphs. These are characteristic of scale-free networks that can be found in kinship networks, physical and biological systems, and economic systems.

As part of our ongoing work we are interested in exploring further our data sets. For example, we are interested in experimenting with different rules for generating the social graphs from the Bluetooth data. Also, we are in the process of running emulations of our data to explore ways in which information is diffused and spreads across the social networks.

***Acknowledgements;*** We thank Alan Penn, Ava Fatah gen. Schieck, Shinichi Iida and George Roussos for their insightful comments and help in carrying out our analyses.



## References

Barabási, A.L., Albert, R., 1999a, "Emergence of Scaling in Random Networks", *Science*, 286: 509-512.

Barabási, A.L., Jeong, R., 1999b, "H. Mean-Field Theory for Scale-Free Random Networks", *Physica A*, 272: 173-197.

Hillier B., Penn A., Hanson J., Grajewski T., Xu J., 1993, "Natural Movement; Or, Configuration and Attraction in Urban Space Use", *Environment and Planning B: Planning and Design*, Pion, 20: 29-66.

Hillier, B., Burdett, R., Peponis J., Penn, A., 1987, "Creating Life: Or Does Architecture Determine Anything?", *Architecture and Behaviour*, 3(3): 233-250.

Hillier, B., Hanson, J., 1984, *The Social Logic of Space*, Cambridge University Press, Cambridge

O'Neill, E., Kostakos, V., Kindberg, T., Fatah Schiek, A., Penn, A., Stanton Fraser, D., Jones, T., 2006, "Instrumenting The City: Developing Methods for Observing and Understanding the Digital Cityscape", UbiComp 2006, *Lecture notes in Computer Science 4206*, Springer, pp. 315-332.

Simon, H.,A., 1955, "On a Class of Skew Distribution Functions", *Biometrika*, 42,425–440.

097-10