Informatics in Primary Care 2005;13:187–93 © 2005 PHCSG, British Computer Society

Refereed papers

The implementation of electronic services: planned or organic growth?

John Cole
Research Officer, Department of Computing

Malcolm Colledge
Visiting Fellow, Institute for Health Sciences and Public Health Research

University of Leeds, Leeds, UK

Tony Megaw
Head of Primary Care Informatics

Martin Powls
Assistant Head of Primary Care Informatics

Health Informatics Service, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, UK

Seth Bullock
University Research Fellow, Department of Computing

Justin Keen
Professor of Health Politics and Information Management, Institute for Health Sciences and Public Health Research

University of Leeds, Leeds, UK

ABSTRACT

The literature on innovation suggests that projects are successful when rigorous project management is mixed judiciously with 'organic' development. This paper argues that organic growth can play a substantial role in the implementation of electronic services in healthcare settings. Evidence for organic growth is presented, based on a study of email use. Methods are presented for investigating email use in health service settings in the National Health Service (NHS) in Bradford, England. Geographical information systems (GIS) outputs and social network analyses are presented. The results demonstrate a fivefold increase in the use of email over a 13-month period, which is shown to be largely independent of the growth in the number of organisations using the network. They also demonstrate a marked increase in the complexity of the patterns of email use over the period.

Keywords: email, innovation, network

Introduction

The literature on innovation suggests that projects tend to be successful when rigorous project management is mixed judiciously with 'organic' organisational development.[1] In the case of information technology projects, there has long been evidence that rigid project management alone will not lead to success.[2] The responses of users, and their decisions to use rather than reject a technology, are clearly crucial, yet the nature of the 'organic' element in the diffusion of new technologies is poorly understood.

Email is now used routinely in many National Health Service (NHS) settings. It is still not accessible by large numbers of staff, but where it is accessible it is increasingly taken for granted: it is steadily becoming

MICROSOFT EXHIBIT 1099
Microsoft v. X1 Discovery
IPR2025-00255

embedded in routine working practices. There have been a number of studies of the ways in which clinicians and managers use email. Most appear to focus on doctor–patient communication.[3,4] Almost all studies report on analyses of cross-sectional data obtained in a single recruitment period, and the papers do not report on dynamic changes in the patterns of use of systems that might be occurring over weeks or months. Similarly, few studies report on the topographical patterns of email use: they tend to focus on specific subsets of communications such as doctor–patient communications, rather than on overall usage. As a result, we have a relatively limited understanding of the ways in which email is used, and the ways in which it has diffused within the NHS and other healthcare organisations.

This study reports on the growth of email use across the health community in Bradford, England, over a 13-month period. Geographical information system (GIS) outputs are presented, as are the results of social network analyses. The results demonstrate a fivefold increase in the use of email over a 13-month period, which is shown to be largely independent of the growth in the number of organisations using the network. The results also demonstrate a marked increase in the complexity of the patterns of email use over the period.

Background and methods

There have been studies of the topology of email network use in other sectors and these show how email use can be studied in healthcare settings. Newman *et al* and Ebel *et al* both studied the topology of email networks.[5,6] In both papers, the authors conceived individuals' or organisations' email addresses as nodes, which are linked by the messages exchanged between them. If you then draw a diagram of the patterns of emails they send one another, then a small number of people send and receive large numbers of emails, while a much larger number send and receive relatively few. That is, the messages are clustered around relatively few individuals rather than being randomly distributed across users. As we describe below, it is possible to explore the nature and extent of this clustering.

Two datasets were obtained from the Bradford Primary Care Trusts (PCTs), the first from 16 to 30 October 2002 inclusive and the second for the calendar month of November 2003. Both datasets were collected using the email-logging software 'Promodag Reports for Microsoft Exchange Server (Standard Edition)'. The software was installed on the single Bradford Primary Care email server.

Geographical information

Data from the Promodag software were organised as relational data, that is, each record related a sender and recipient of an email, together with contextual data including date and time sent, and size of email, collected as an indicator of the presence or absence of attachments. No data were collected about the contents of the emails. Data were collected and analysed at organisation level – general practice, health centre, hospital, etc. Postcode data were supplied by local NHS organisations. The resulting dataset was used to populate GIS maps using GeoConcept software. Maps were generated which showed the inter-organisational email traffic for individual PCTs in Bradford.

Social networks and small worlds

In contrast to GIS, which are now well established and relatively easy to use, methods for investigating whether there is any underlying pattern of use of emails require knowledge of social network analysis.[7] The method has been used relatively rarely in health services research, mainly in studies of networks of support in mental health services. The fundamental assumption underpinning social network analysis is that there are regularities in the behaviour of groups of people, whether one is concerned with the routines involved in getting to work or who emails whom in the course of a week or month. These regularities can be so embedded in the way that activities are undertaken that they are stable enough to study in their own right. Equally, behaviour can and does change over time: we change jobs, or firms move or go out of business, or we find ourselves emailing different people as our job roles change over months and years.

Recent research has revealed striking mathematical similarities in the underlying structures of many diverse types of real world network. Until recently it has been necessary to assume that networks, of any kind, approximated either to a random graph or to a regular lattice. This has allowed researchers to develop useful tools for understanding social and other networks, even though it has been clear that these networks are not strictly ordered, regular lattices nor entirely random: the assumption has proved to be a useful simplification, if used with care. The 'small world' model proposed by Watts and Strogatz in 1998 demonstrated for the first time that it was possible to describe and understand networks that exhibited mixed characteristics between order and randomness (see Figure 1).[8] Their work opened up new ways of studying social networks, which earlier research had already shown often had the characteristics of large and highly locally clustered graphs with short characteristic path lengths (that is, the number of steps

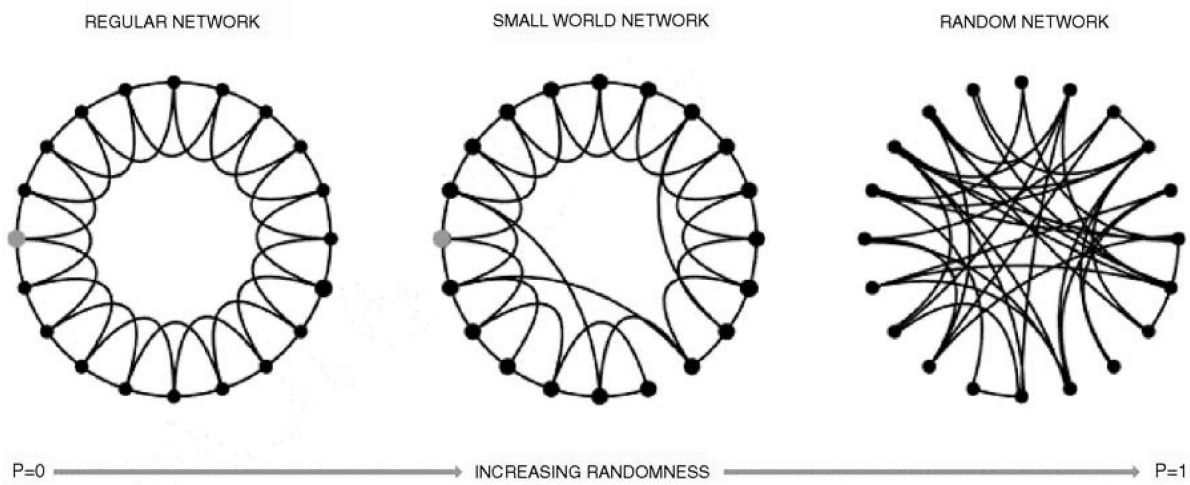

Figure 1 The 'small world' model of Watts and Strogatz is in the middle, between the regular lattice and the random network

required to travel between any two nodes). Social networks, such as networks of friendships, can exhibit this 'small world' quality.[9]

Investigation of these underlying structures involves generating a number of statistics. One approach is to plot the degree distribution. A node's degree is the number of links that attach to it. The easiest way to display the degree distribution of a network is to make a histogram of the number of nodes possessing a certain degree (or number of links). Random graphs exhibit a Poisson degree distribution, in which most nodes have a degree equal or close to the mean degree. In contrast, real world networks often display exponential or power law distributions.

A second important statistic is the clustering coefficient. This reflects the extent to which a network is clustered around a small number of nodes, as is the case in a hub-and-spoke network. This contrasts with a random network where there are few or no hubs. In this study it was possible that PCTs would serve as hubs. The clustering coefficient is an adequate single number measure for non-directed networks, that is, networks where the 'direction of travel' between any two nodes is not important, as is the case in power transmission lines where electricity can travel in either direction. When the networks of interest are directed, as they were in this study – an email is sent in a particular direction between one organisation and another – the concept of transitivity comes into play.[10] For the purposes of this study the following definition of transitivity was used:

The triad involving nodes i, j and k is transitive if whenever $i \rightarrow j$ and $i \rightarrow k$ then $j \rightarrow k$ or $k \rightarrow j$

This definition was formulated to give an approximation to the similar 'clustering coefficient' statistic applied to non-directed networks. Further useful statistics are the path length and the 'harmonic mean' average path length. The path length is a measure of the number of nodes one has to pass across to get from one node to another. Thus the average path length reflects the average number of steps between any two nodes. If a network is highly clustered then the densely connected nodes tend to reduce the number of steps between any two nodes in the network: densely connected nodes reduce the average path length in a network. Newman shows how the harmonic mean distance between pairs of nodes can be calculated.[11]

It should be stressed that currently available methods do not allow investigation of the dynamic development of networks directly. The collection of two cross-sectional datasets from the same location allowed us to observe the differences in the way email was used in two time periods, 13 months apart.

Results

Table 1 lists basic properties of the two email datasets. Table 2 shows that while the minimum and maximum size of emails did not change much between the two periods, the mean increased by a factor of approximately 2.5 and the median doubled. The distribution of email sizes is also wider, indicated by the larger standard deviation. The larger increase in mean than in median indicates that a few very large emails have 'pushed up' the mean, altering the distribution.

Table 1 Volume of traffic

Statistic	2002	2003	2003 (adjusted)*
Total number of recorded email events	58 655	588 724	294 362
Amount of 'in' traffic (traffic from external addresses)	10 509 (26.4%)	218 477 (43.5%)	109 238 (43.5%)
Amount of 'out' traffic (traffic to external addresses)	8336 (20.9%)	196 407 (39.1%)	98 203 (39.1%)
Amount of internal traffic recorded twice	18 810	86 920	43 460
Amount of 'true' internal traffic	21 000 (52.7%)	86 920 (17.4%)	43 460 (17.4%)

* Numbers adjusted to equivalent days for comparison with 2002 data.

Table 2 Size of emails 2002 and 2003

Year	Email size statistics (bytes)				
	Min	Max	Mean	Median	Standard deviation
2002	177	6 751 428	49 997	2228	194 912
2003	172	9 989 173	128 276	4435	549 506

Figure 2 shows the geographical pattern of email use in Airedale PCT in 2002. The other PCTs showed similar patterns in the same period. The thickness of the lines reflects the volume of email exchanged between any two sites during the two weeks in October 2002. The PCT is clearly at the centre of email traffic. The numbers of emails sent in each direction between any two sites in this period was roughly equal.

In Table 3 the following points can be noted:

1 The mean degree for the organisational network is low. Typically, each organisation sends email to only three others.
2 The network displays a high clustering coefficient in comparison to the random net. It also has a significantly lower average path length.
3 The average path length is short but harmonic mean average path length is higher. This indicates that the contribution of 'non-existent' paths is significant for both networks.
4 In- and out-degree are heavily positively correlated. This indicates that those organisations that send larger volumes of email also receive a larger volume. (In contrast the in–out degree correlation of a random net tends to zero as network size increases.)
5 The clustering coefficient is significantly higher than that of a random network.
6 The majority of traffic was generated during normal working hours.

In the graph in Figure 3 the three hubs of the star-like configuration represent three PCT headquarters (Airedale, Bradford City and Bradford South and West). This graph indicates that in September 2002 the vast majority of organisations sent and received email to and from one or more of the PCTs, but not one another.

Similar points can be made about the 2003 data shown in Table 4:

1 The mean degree is higher than 2002, indicating that the networks are now much more connected (that is, they contain more links).
2 The network displays a high clustering coefficient in comparison with the random net. However, in this case they both have similar average path lengths to

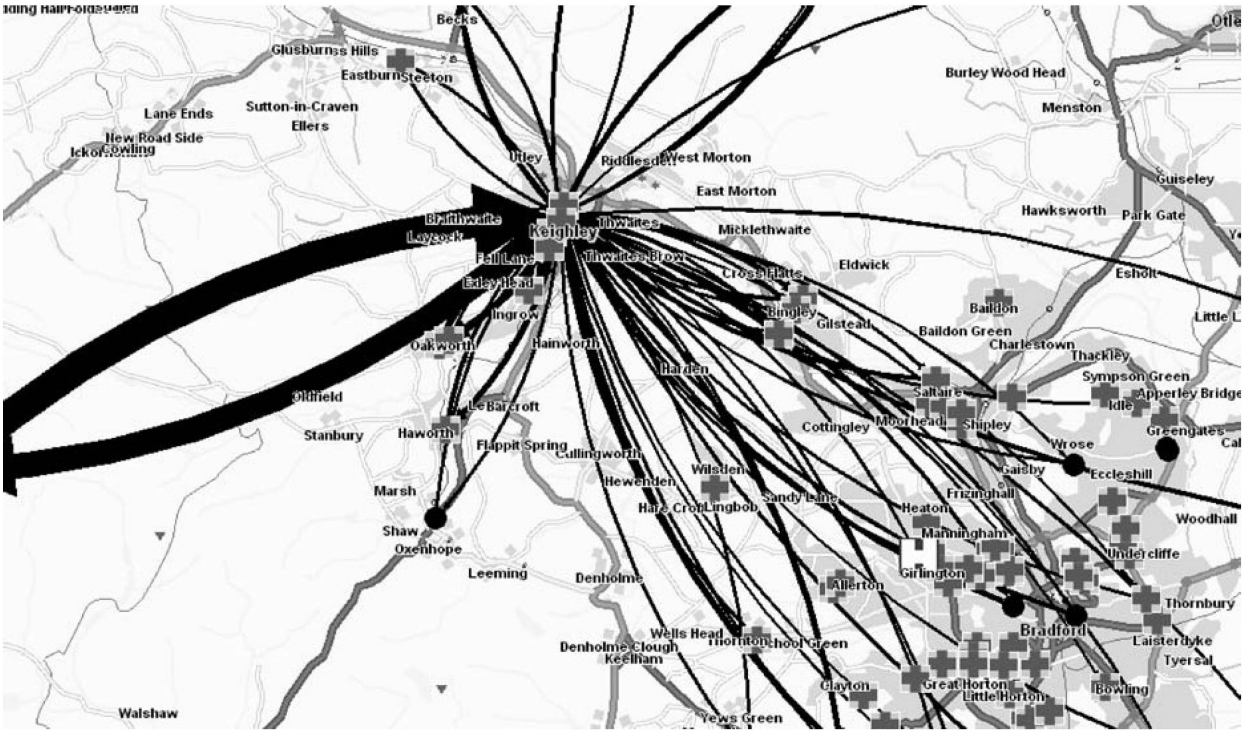

Figure 2 The geographical pattern of email use in Airedale PCT in 2002

Table 3 2002 data compared to a random net with same number of nodes and links

Network property	Bradford NHS organisations	Random net
Number of nodes	116	116
Number of links	323	323
Mean degree (in and out are equal)	2.78	2.78
Average path length	2.35	4.55
Harmonic mean average path length	34.2	41.6
In–out degree correlation	0.994	–0.109
Clustering coefficient (links assumed non-directed)	0.280	0.056

the equivalent random nets. Harmonic mean average path length is again much larger than standard average path length.

3 In- and out-degree are again positively correlated.

4 The clustering coefficient is higher than that of an equivalent random network.

In Figure 4, the threshold value is set to 30, so only instances where more than 30 emails were sent between pairs of nodes are represented by an arrow. The circles indicate where at least 30 mails were sent within an organisation. A diagram showing every link would not be intelligible, the lines being too dense to represent. Figure 4 shows a substantially different topology to

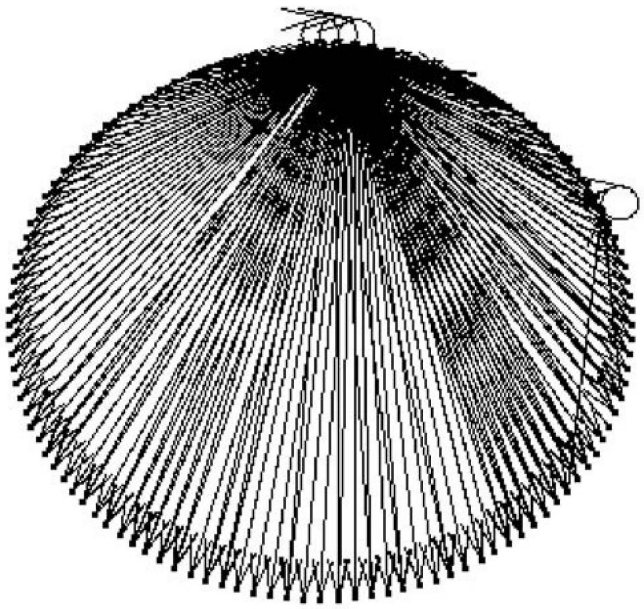

Figure 3 Visualisation of 2002 data

that of the 2002 network. Here it is apparent that many organisations are communicating directly with each other, rather than only with the PCT headquarters. There are also many more 'self edges' (the loops attached to some nodes), indicating internal traffic within individual organisations. Elements of the 'star' type graph remain; one node near the top (A) links to many in the lower right portion of the circle (B).

Discussion

The results show that there were substantial changes in email use between 2002 and 2003, reflected in the proportions of internal and external messages, the average sizes of emails and the patterns of use within Bradford. The GIS results allow us to visualise these patterns, and are useful for communicating general results. They do not, however, reveal any deeper systematic features in the data. The social network analysis results are more technically demanding to generate, but offer insights into the dynamic development of email use in Bradford. The results of these analyses suggest that both the 2002 and 2003 patterns of email use were more clustered than random networks with the same numbers of nodes and links.

This study did not address questions of the value of email use, and we do not know how far email is substituting for existing communications or adding a new service. Neither can we judge how far the growth of email use has influenced the volumes of use of other communications media such as faxes and telephone

Table 4 2003 data compared to a random net with the same number of nodes and links

Network property	Organisations	Random net
Number of nodes	132	132
Number of links	1964	1964
Mean degree (both in and out are equal)	14.9	14.9
Average path length	2.08	2.09
Harmonic mean average path length	8.90	8.86
In–out degree correlation	0.835	0.072
Clustering coefficient (links assumed non-directed)	0.665	0.218

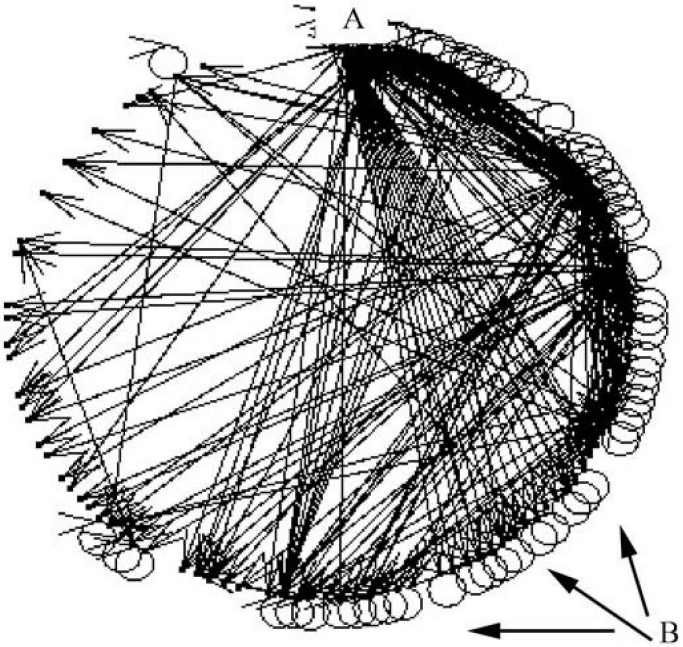

Figure 4 Visualisation of 2003 data

conversations. However, we believe that the results offer support for the view that email use has grown through a mixture of an increase in the size of the network and embedding of email use in people's work. Organised activities such as the phased introduction of new organisations onto the network are to some extent driving these changes, but the 'organic' growth we have observed cannot, by its nature, be planned.

REFERENCES

1 Burns T. *The Management of Innovation.* Oxford: Oxford University Press, 1994.
2 National Audit Office. *Better Public Services Through e-Government.* HC 704-III, Session 2001–02. London: The Stationery Office, 2002.
3 Delbanco T and Sands DZ. Electrons in flight – email between doctors and patients. *New England Journal of Medicine* 2004;350:1705–7.
4 White CB, Moyer CA, Stern DT and Katz SJ. A content analysis of email communication between patients and their providers: patients get the message. *Journal of the American Medical Informatics Association* 2004;11:260–7.
5 Newman M, Forrest S and Balthrop J. Email networks and the spread of computer viruses. *Physical Review E* 2002;66:Article 035101.
6 Ebel H, Mielsch L-I and Bornholdt S. Scale-free topology of email networks. *Physical Review E* 2002;66:Article 035103.
7 Watts D. *Six Degrees: the science of a connected age.* London: Vintage, 2004.
8 Watts D and Strogatz S. Collective dynamics of small world networks. *Nature* 1998;393:440–2.
9 Newman M. Fast algorithm for detecting community structure in networks. *Physical Review E* 2004;69:Article 066133.
10 Wasserman S and Faust K. *Social Network Analysis: methods and applications.* Cambridge: Cambridge University Press, 1994.
11 Newman M. The structure and function of complex networks. *SIAM Review* 2003;45:167–256.

CONFLICTS OF INTEREST

None.

SOURCE OF FUNDING

The work reported in this paper was partly funded by the Department of Health.

ADDRESS FOR CORRESPONDENCE

Professor Justin Keen
Institute for Health Sciences and Public Health Research
University of Leeds
71–75 Clarendon Road
Leeds LS2 9PL
UK
Email: j.keen@leeds.ac.uk

Accepted July 2005