

Presented at KM2003 Challenge - April 2003

**Does your Community Leave a Digital Footprint?
Measures and Metrics for Communities of Practice**

**Laurie Lock Lee
Computer Sciences Corporation**

Abstract

Methods for measuring the health and performance of Communities of Practice (CoP) have traditionally been limited to surveys and interviews of the membership. Practically, this means that assessments can only be done infrequently, having to be mindful of “survey fatigue” as the enthusiasm for participation in surveys drops off with each subsequent application. As communities begin to make greater use of electronic support tools, the potential for using digitally tracked activities i.e. digital footprints, as proxy measures for actual CoP activity, is increasing.

Two experiments with digital tracking of community activities are presented. The first experiment uses electronic discussion group data fed into a Social Network Analysis (SNA) visualisation tool to analyse activity patterns over a two year period. The second experiment compares a traditional SNA of a Global Network with an electronically derived sociogram using usage data collected from the web based CoP support tool used by the Network. The results were encouraging, confirming the potential for digital tracking. In particular, active discussion group analysis using SNA tools to help visualise the interactions, told quite plausible stories of what was actually happening in the networks.

Digital tracking of community activity is seen as a useful adjunct to the more traditional CoP metrics. As we move more actively into the electronic age, the value and validity of non-intrusive methods like digital tracking, will become much more widespread.

Introduction

The growing adoption of Communities of Practice (CoPs) by the business world has inevitably created a strong interest in performance measurement. Performance measures for CoPs have proven problematical, with many different views of the best way to do this. One way is to provide ROI measures for “projects” conducted by CoPs. Another is to use anecdotal evidence in the form of stories and narratives. A third method is to make use of “in-process” measures in the spirit of Total Quality Management [1] i.e. CoP activity measures. Best practice is likely to be a combination of all three approaches. It is the process metrics that are best served by digital tracking techniques.

Social Network Analysis (SNA) is arguably the leading method for understanding knowledge flows within CoPs. SNA is able to help visualise the knowledge sharing connections within CoPs. SNA statistics provide a means for measuring characteristics like “Centrality” (Core Members) and “Structural Holes” (Points of Weakness) [2]. These measures can be used to infer the strengths and weaknesses of a CoP. Traditionally, data for SNA is achieved through surveys or face-to-face interviews. This paper looks at digital tracking data as an additional source for SNA measures.

The paper takes an initial look at the applications where “digital tracking” has been used to characterise CoP activity. The results of two experiments, conducted by the author, in tracking CoPs’ electronic activities (their digital footprints), are presented. The aim of these experiments was to assess the effectiveness of digital proxies for actual CoP activity. The first experiment looks at CoP activities mapped over time, providing a temporal perspective of CoP performance. The second experiment looks at mapping (electronic) memberships of CoPs within a large Network, comparing and contrasting this with results from a traditional survey based SNA analysis of the Network. The results of these experiments are then discussed in terms of current value, shortcomings and opportunities for using electronic proxies for measuring CoP performance. Digital based SNA analysis is then placed within an overall measurement framework for CoPs.

Motivation for this Research

Knowledge Management has traditionally evoked a dual focus, variously described as tacit and explicit knowledge, personalisation and codification, knowledge and information, connections and collections etc.. While the information focus is relatively mature in terms of methods and technologies, the knowledge focus has proved much more elusive. The use of SNA to analyse people connections is considered a mature social science tool. Its major shortcoming however is that the data is invariably collected via surveys and interviews. Generally this means that SNA studies are mostly one-off studies, as few organisations would tolerate re-surveying on say a monthly basis, to track changes over time, not to mention the effort that this would involve. Other suggested shortcomings relate to the truthfulness of individual responses to the surveys i.e. will someone be honest when asked questions about personal relationships?

So how can we measure community relationships over time without interviewing people? One possible answer is to collect data from the electronic tools that the CoP members use to collaborate. As geographically distributed CoPs begin to rely on electronic tools to maintain relationships, they will begin to leave their “digital footprints” which in turn will provide a rich suite of data to assess CoP performance over time. The use of electronic surveillance data for input into SNA is not new. Due to the expense required in collecting this data from disparate sources, its use has been limited to applications like fraud detection or criminal investigations. Terrorist network mapping [3] and the Australian backpacker murderer case [4] are two high profile examples.

The clear advantage of using the digital footprint approach is that community connections can be monitored non-intrusively over any time increment that suits. It also tracks actual connections made, avoiding the issue of inaccurate or untruthful interview responses.

The big question however is clearly “Does the digital footprint reflect reality or not?”; the basis for this research.

Examples of Digital Tracking of Social Networks

The stunning growth of the internet and the massive information repository which underpins it has provided a rich source of data for the data mining enthusiast. A number of researchers [5] [6] have become interested in identifying and characterising “cyber-communities”. In fact “Web Mining” is now considered a fully-fledged discipline in its own right. A branch of social science focussing on the cyber world is now emerging as both public and private organisations become interested in better understanding the behaviours of on-line communities. Identifying interest profiles for marketing purposes is one obvious application. Another is using public information from the internet to understand industry relationships.

Laurent [7] has used a combination of link analysis and content analysis to characterise the network of relationships that exist amongst the leading computer services organisations in North America. Figure 1 shows that IBM and HP are the most connected in the industry. Thickness of the links represents the strength of the relationship inferred from the link and content analysis. The sizes of the arrowheads indicate who is driving the relationship.

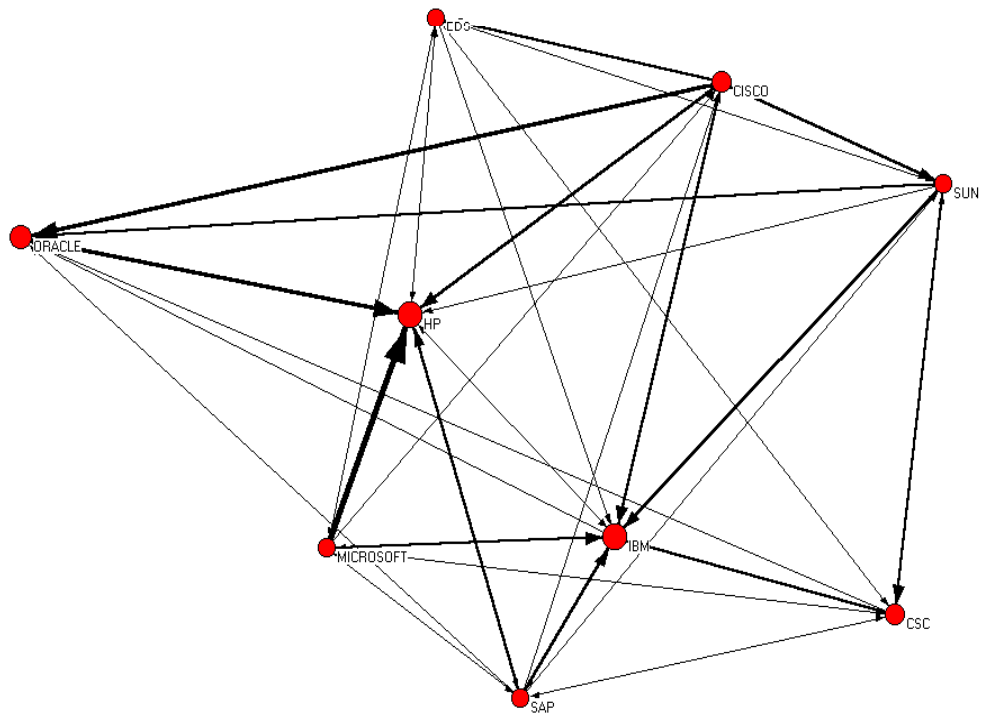


Figure 1 – Inferred relationships between Computer Services Companies

Another interesting application of digital tracking for characterising communities is Isaac Oates and Omar Ashrafi’s GradeAIM program. This system provides a visual map of communities compiled from the AOL Instant Messenger “Buddy Lists” of students at the University of Illinois [8]. The Amazon.com agent that recommends books by inferring an interest profile based on previous purchases is another example of the effective use of digital tracking. It would only be a short extension for Amazon to provide tailored invitations for clients to join “communities” around particular interest topics.

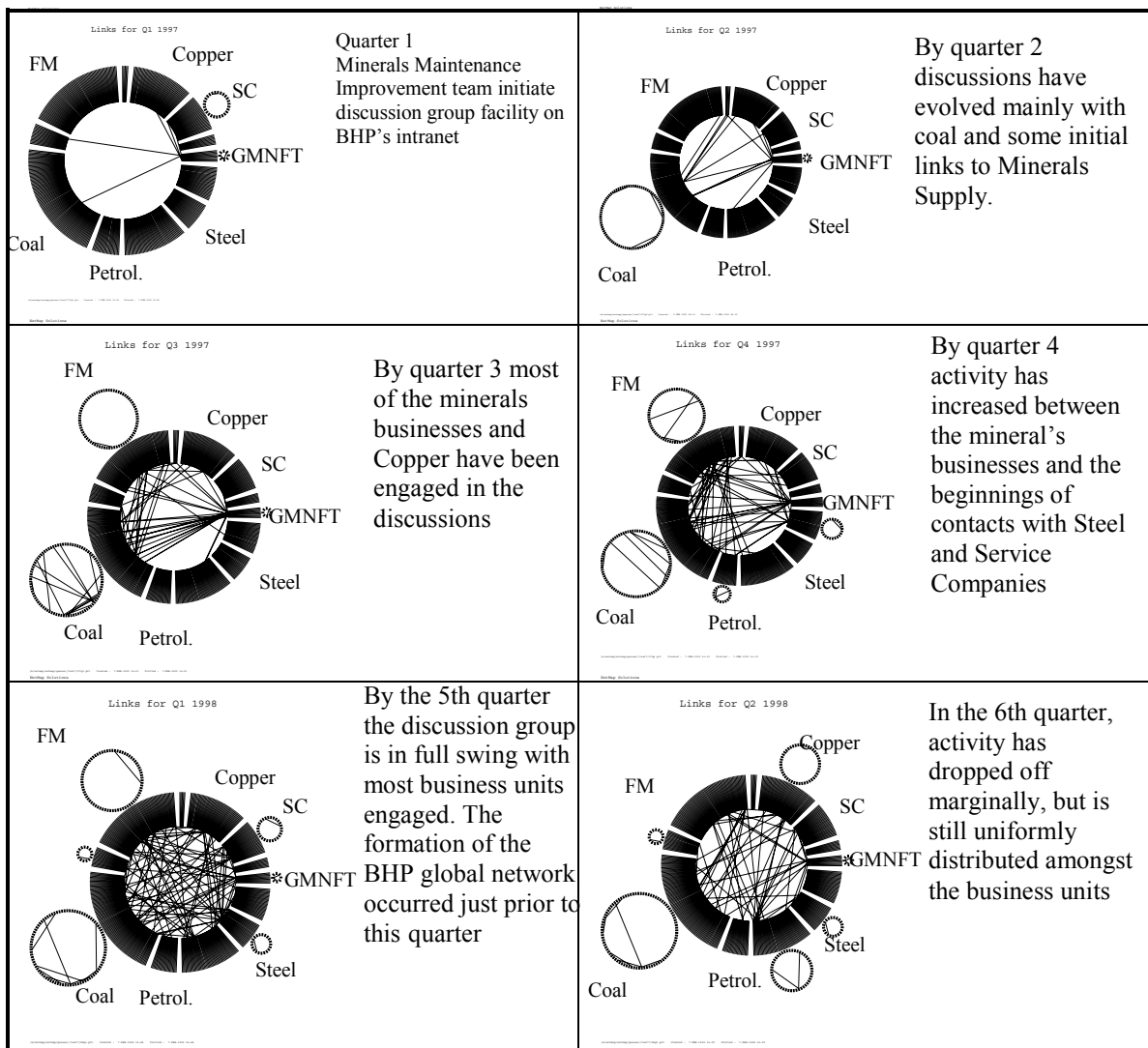
The majority of “connections” data that would be useful for tracking social networks, would be held by the infrastructure companies e.g. telecommunications and information technology data centres. The problem appears not to be a lack of data, but a lack of sophistication in mining it. A good example of this is the poor attempts many companies make at targeting prospective clients from data held in their customer relationship management (CRM) systems. We have all been victims of this. The challenge is clearly to design appropriate tools that people can use profitably yet can still leave a useful “digital trail” for subsequent analysis.

Experiment #1: Social Network Digital Evolution

This experiment was aimed at studying social network changes over time. The selected electronic data source was an active electronic discussion group of a global engineering network. The data source was selected over e-mail as it was thought that participation in an electronic discussion is principally for knowledge sharing

purposes, whereas e-mail has many other purposes. The discussion activity was reported quarterly, with a focus on business unit level connections and their changes over time.

The following figures describe the evolution of the discussion group over its first 2 years of operation. The arcs represent business units, the satellites showing intra-business contacts. Links between arcs within the main circle are inter-business contacts. Links are basically a discussion post. A discussion thread will create a number of links either linked back to the original post or to other responses to the original post. Only major business units are shown (GMNFT is the Network facilitating team):



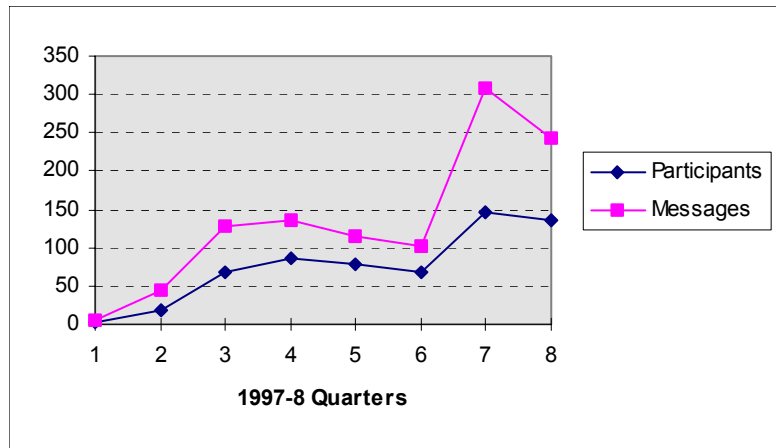
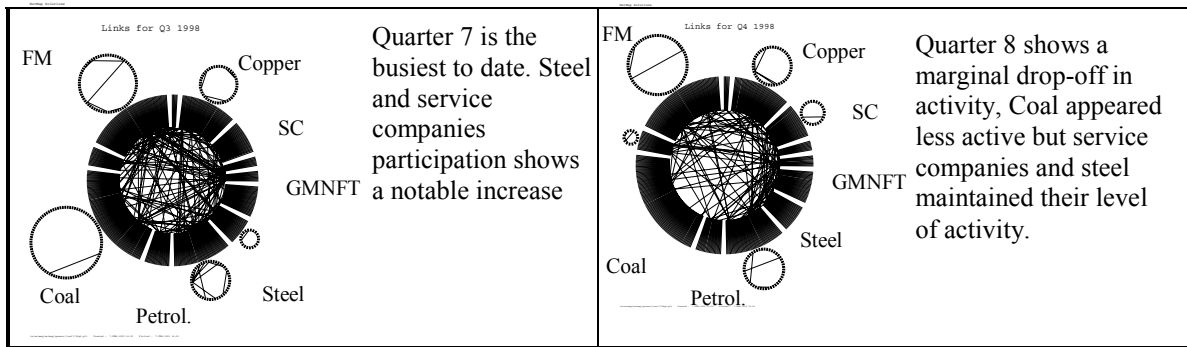


Figure 2 – Evolution of Discussion Group activity

The reports were provided to the Network’s core facilitating team (GMNFT) for evaluation against their perceptions of Network activities. Overall the facilitating team found the interpretations and results quite plausible. In several cases they could identify events or situations that had occurred within the network that could explain / justify the interpretations. The reports also showed clearly the relative participation rate of the different business units and the impact of the facilitating team, seeding discussions when activity quietyens. One common characteristic was the increase in discussion activity following a major face-to-face conference.

Experiment #2: Social Network Digital Identification

This experiment related to another global network’s use of a custom developed CoP support tool, compared against a traditional SNA assessment of the Network. This Global Network was divided into several topic specific CoPs, who each made use of the electronic space provided by the electronic CoP tool. Network members were free to volunteer to join the “electronic CoPs”, but clearly not all CoP members joined the “electronic CoPs”. It is estimated that about 60% of Network members had joined the CoP electronic space. Some are members of multiple CoPs.

The traditional SNA analysis was conducted across the Network to assess inter-office interactions and connections within particular professional disciplines. Around 40 Network members were interviewed, basically being asked to nominate their key “trusted advisors” within their specialty domain. The resulting sociograms described

Table 1 compares the top 10 (out of 33) rankings for both the off-line and on-line networks. For the traditional SNA the rankings were determined based on centrality i.e. the number of input links from colleagues. For the on-line community the relationships were inferred by the number of different colleagues a member shares more than one CoP membership with. e.g. Mark is a member of two CoPs that both Charles and Gary are also members of.

Top 10 “On-Line” participants (in rank order)	Top 10 “Off-Line” participants (in rank order)
Charles	Mark
Gary	Graham
Geoff	Geoff
Doug	Andy
Andy	Trevor
Graham	Katie
Katie	Doug
Rob	Tim
Ben	Rob
Mark	Bob

Table 1 – Off-line verses On-Line Comparison

The results show that 70% of the top 10 are common between the on-line and off-line communities, which would suggest that those well networked members who choose to join the on-line world are happy to participate in multiple CoPs. Rank orders within this grouping do not correlate that well, suggesting that there is a limit to how well networking behaviour might map from the off-line to on-line world. The top two ranking members for the on-line community do not figure in the top ten of the off-line community. Charles is the Network’s leader and clearly plays a facilitative and oversight role in the on-line world, but is perhaps regarded more as a “manager or information broker”, than discipline specialist in the off-line world.

Discussion

On the experiments:

The above ad-hoc experiments were not conducted as part of a purposeful research plan and therefore it is difficult to draw scientific conclusions from the results. However the following observations can be made:

- If we are interested in knowledge sharing, the best electronic data will come from activities that are uniquely for knowledge sharing, as opposed to say, administration tasks. For example, discussion group activity is likely to be more representative than e-mail or telephone data, as its use is almost exclusively for knowledge sharing.
- The use of sociograms for mapping discussions adds an extra dimension above simple counts of postings. By showing the source of each link, one can

reasonably infer intra-organisational linkages, especially if the discussion is active.

- It should also be noted that the discussion group analysed was mostly a “request for help” on operational issues group, meaning that responses were often drawn from a broad range of respondents – perhaps more representative than if the group was discussing more conceptual issues.
- Clearly comfort with the on-line world is going to have an impact on how representative electronic activity will be of off-line activity. In both cases under study the level of IT literacy was relatively high.
- While an electronic discussion response is clearly not as strong as an explicit link nomination in a survey, with sufficient data and looking at the data aggregated at the business unit level, it does provide some insight into inter-business unit interactions.
- In the on-line world the activities of facilitators becomes visible e.g. seeding discussions, joining electronic CoPs for oversight. In the off-line world analysed by SNA, facilitators are likely to appear as information brokers, rather than as “advisor” links.

On the effectiveness of digital tracking:

The digital tracking of CoP activities will form only one of a collective of measures available for assessing the health and value of a CoP [9]. The following figure illustrates where SNA fits within the full framework of CoP metrics.

Value Derived	Direct Output Measures	
	CoP sponsored Project ROIs	New Practices Developed
	Time/cost savings from CoP assistance	New Solutions Developed
CoP Health	In-Process Measures	
	No. of CoP meetings	Growth in membership
	<i>Quality of connections (SNA)</i>	Usage statistics on shared electronic space

Figure 5 – CoP Measurement Framework

Outcome measures typically capture benefits accruing from specific CoP projects or the development of new artefacts e.g. new processes or procedures. SNA type measures are more related to assessing structural aspects of communication patterns within CoPs i.e. in-process metrics [1]. By identifying “trustful” relationships in the interviews, SNA does more than just identify co-ordinator or broker roles, it also infers that highly connected individuals possess valuable and sought after tacit knowledge. Despite the clear differences between interview based SNAs and digitally derived SNAs, digital tracking was still seen as effective at:

- Identifying the major communication patterns between membership groups e.g. formal departments or business units and professional disciplines within the community. The author has conducted several interview based SNAs where the subjects were asked to differentiate between knowledge based contacts and information providing contacts. Largely, the patterns for knowledge contacts and information contacts were similar when viewed at the aggregate level. This also appeared to be the case with digital transactions.
- Identifying important brokers and co-ordinators in the community. These members may take the form of discussion group moderators or content managers, but in general it was observed that active members in the “off-line” community would also be active if they choose to participate on-line as well.
- In geographically dispersed communities, can identify key, remotely located contributors that would not be otherwise identified in an off-line only community.

A number of sociology inspired studies have been conducted to try and understand the level of community that can be generated through participation in on-line forums. Eminent sociologist, Barry Wellman [10] argues that on-line communities should not be studied in isolation, but as an integrated component of everyday life. Wellman suggests that on-line community participation increases an individual’s social capital, through increasing the breadth and frequency of contacts. Therefore, the effectiveness of digital tracking should be judged by the added value that it brings in understanding social constructs within CoPs, more so than its value as a substitute for interview based SNA’s.

CoP types can be placed along a continuum from tight co-located communities who meet regularly face to face, through to fully virtual communities where members may never meet face to face. The following descriptive analysis is provided to assist with understanding the different communication mechanisms and how they might contribute to CoP success.

	On-line CoP	Off-line CoP	Combined
Defining Characteristic	<i>Established as an on-line community e.g. Yahoo Groups</i>	<i>Established as a “personal contact” community.</i>	<i>Off-line community which makes use of on-line tools to assist with communications.</i>
Key Value Generated	Highly scaleable both geographically and numerically.	Highest trust levels attainable	Allows some up-scaling within the constraints of acceptable trust levels.
Key Limitations	Can take a long time and many “transactions” to achieve acceptable levels of trust within the community. Will never engage all members.	Typically requires members to be co-located and therefore limits participation.	Can develop a division between those that socialise easily in face to face situations and those that thrive in digital environments.

Table 2 – Characteristics of On-Line and Off-Line worlds

It is worth noting that very few communities would be “On-line only” or “Off-line only”. With on-line communities it is common practice for side discussions to occur either by e-mail or telephone. It is also unlikely that off-line communities will not use some level of e-mail or telephone communications (which could potentially be tracked).

On the future:

As with most data/web mining activities, the acquisition and cleansing of the source data is the most costly and time-consuming task. But as more experience is gained in understanding which forms of data are most strongly predictive of community health and performance, the source applications can be designed to provide the appropriate quality of tracking data.

Comfort in the on-line world is currently an issue that will undoubtedly diminish as the “generation X-ers” start to dominate the workforce. The Internet age is still relatively young, so those that have “grown up” in the Internet age are mostly in their teens. But already this group will have generated one of the richest data sets for inferring social connections i.e. the database of SMS messages!

The growing sophistication of the web mining tools and advanced content analysis engines would suggest that the availability of effective digital tracking tools for measuring community performance will not be far away. For those communities that begin on-line and largely operate on-line e.g. Yahoo Groups, one could anticipate that SNA tools could become a value added service for community leaders/facilitators.

But for those communities established off-line but with an on-line presence, the largest potential “show stopper” is a non-technical one i.e. the evolving privacy laws that could potentially outlaw the use of digital tracking data without the explicit permission of all participants. There is also the issue of social acceptance of what amounts to a “big brother” monitoring activity. Communities may indeed object to having their activities monitored, even if it is in the name of increased efficiencies or effectiveness.

Conclusions

The motivation for this research has been the need to devise a way where the health and performance of CoPs could be monitored continuously without the need to continuously survey its members. “Survey fatigue” is a common problem with social science studies, so the ability to use digital proxies is particularly attractive.

The two experiments described provided some insight into the potential for developing electronic CoP performance proxies. The most promising area for mining was seen as the on-line discussion groups, where the usage is nearly exclusively for knowledge sharing. The visualisation of discussion group patterns using SNA tools proved beneficial in helping to “tell the story” of how a CoP evolves over time. There was also some evidence that key CoP members, as determined through traditional SNA, will also be active in the supporting on-line world. Digitally derived SNA will not effectively replace the traditional interview based SNA. They will however provide a complementary value, by providing some of the same benefits in certain situations e.g. major communication patterns and some new benefits e.g. inclusion of remotely located participants.

Web mining as a discipline is rapidly maturing to the extent that one could expect that tools will become available for reliably identifying the “digital footprints” of on-line CoPs. The technology will also enable emerging communities of interest to be identified from their web based activities. The extent to which digital tracking can effectively be used to monitor CoP performance or discover new CoPs may only be limited by growing privacy legislation and/or social acceptance in general.

References

1. Lock Lee L., “Knowledge Sharing Metrics for Large Organisations” in “Advances in Knowledge Management: Classic and Contemporary Works”, MIT Press, January 2001.
2. Borgatti S., Jones C. and Everett M., “Network Measures of Social Capital”, CONNECTIONS 21(2):-36 or www.analytictech.com/borgatti/borg_social_capital_measures.htm
3. Krebs, V., “Mapping Networks of Terrorist Cells”, CONNECTIONS 24(3): 43-52.
4. Dawson, R., “Beware! Netmap may be Watching”, the Bulletin, July 22, 1997.
5. Flake, G., Lawrence, C., Giles, L. and Coetzee, F., “Self-Organisation of the Web and Identification of Communities”, IEEE Computer, 35(3), 66-71, 2002.

6. Gibson, D., Kleinberg, J. and Raghavan, P., "Inferring Web Communities from Link Topology", Proceedings of the 9th ACM Conference on Hypertext and Hypermedia, 1998.
7. Laurent, D., "Understanding External Relationships in the Knowledge Era: An Exploratory Study that Applies Social Network Analysis to the World Wide Web", unpublished Honours Thesis, School of Information Technologies, Sydney University 2002.
8. see "The "Small World" at the University of Illinois Just Got Smaller", <http://alexia.lis.uiuc.edu/gslis/school/news/gradeaim.html>.
9. Wenger, E., McDermott, R. and Snyder, W., "Cultivating Communities of Practice", Harvard Business School Press, 2002.
10. Wellman, B., "Computer Networks as Social Networks", SCIENCE, Vol 293, September 2001.