

# INSTANT MESSAGING

## *Using log files to identify patterns of interaction in a group of teenage users*

Brian Davison and Kathy Buckner

*Napier University, Edinburgh, U.K.*

**Abstract:** This paper presents preliminary findings on teenagers' use of instant messaging (IM) as an example of an Internet technology. The results are quantitative, being drawn from the information held in users' IM log files. The current purpose is to validate the proposed data collection method, and to explore the analytical possibilities it presents. This is done through the discussion of a number of graphical representations generated from the data.

**Key words:** Instant messaging, emoticons, social networks.

## 1. INTRODUCTION

Today's children are often considered the 'net generation' who will be the inheritors of the true 'information society' (Williams, 1999). However, despite their growing use of the Internet - around half of 9-19 year olds spend between half an hour and an hour online every day - this use is narrow and skills are likewise limited (Livingstone and Bober, 2003; Schiano *et al.*, 2002). Young people's experience of Internet technologies, and the skills they bring with them when they go on to the workplace or to higher education need to be understood in order to be further developed and exploited (Smith, 2002; Farmer, 2003).

When teenagers use online communication tools in their own time and within their own peer groups, they do so out of choice. They are not constrained by externally-imposed goals or rules as are the adults in existing workplace studies (Farmer, 2003; Herbsleb *et al.*, 2002; Isaacs *et al.*, 2002; Nardi *et al.*, 2000). Observation of this natural phenomenon may provide

insights into system design, communication behaviour and online learning that are not subject to the preconceptions that may hold true in other circumstances.

Teenagers' use of online communications comes at a time when they are developing their own sense of group and personal identity. IM has characteristics (discussed below) that make it more suitable for this kind of exploration than other kinds of online communication, and this may go some way towards explaining its popularity. If we accept that online relationships are just as real and valid as in-person ones (Peris *et al.*, 2002; Suler, 2004), the use of IM by teenagers provides an opportunity to study their development in a non-intrusive way.

The first step towards either of the two general goals outlined above is the development of methodologies for capturing teenagers' IM behaviour. The next few sections provide the context for this study into the use of data from IM log files.

## 2. INSTANT MESSAGING













Interpersonal communication has been a key ingredient of the Internet since email was first invented in the early 1970s. Within one year of its introduction onto ARPANET, the forerunner of the Internet, email accounted for 75% of all traffic (Zakon, 2004). Email, however, is *asynchronous* - it does not require both participants to be online communicating in real time. This may have its advantages, but it does not foster interactive conversation. This problem was resolved in 1988 with the creation of Internet Relay Chat (IRC) (Oikarinen, n.d.). IRC allowed one user to type a message which would appear *synchronously* on the screen of the recipient. The development of IRC followed the model of existing asynchronous newsgroups, and led to the concept of the *chatroom* - a virtual 'place' where users would go in the hope of finding others to engage in conversation. Chatrooms are typically organised around topics of interest, an arrangement which enables users to find like-minded conversational partners. Like early email, these chatrooms relied solely on plain text, and in order to convey tone or emotion, users had to use uppercase letters, repetition or symbols such as punctuation. Some conventions evolved, such as the use of uppercase to indicate a raised voice, or emoticons - arrangements of text characters representing facial expressions - to indicate tone. More recently, with improvements to graphical user interfaces, Internet bandwidth and network-oriented programming techniques, variations on the chatroom concept have been developed which allow users to present a visual representation of themselves to others through the use of computer-

generated characters, or *avatars* (Talamo and Ligorio, 2001; Suler, 1997; Sulake Labs, n.d.). Such systems are tightly associated with a geographic metaphor - users 'go to' particular rooms to see 'who is there'. In the more sophisticated systems, users can even take ownership of a piece of this virtual space, decorate it as a further act of self-expression and invite other users in.

In contrast to these virtual spaces, IM prioritises the ability to locate and communicate with a set of known contacts, referred to as the user's *buddy list*. This subtle shift in emphasis leads to IM being used quite differently from chatrooms (Livingstone and Bober, 2003) and more like other interpersonal media such as mobile text messaging.

Current IM systems, typified by Microsoft's MSN Messenger used by the teenagers in this study, allow a greater degree of visual expression than the original text-only chatrooms, but fall short of the possibilities afforded by multimedia environments. The avatar is reduced to a single image that represents the user on a conversational partner's screen, and the personalisation of a piece of virtual space is similarly reduced to a choice of background pattern or colour of the conversation window. Enhanced versions of early chat conventions such as emoticons persist in IM systems as shown in Table 1.

Table 1. Some emoticon examples from MSN Messenger

	Smile	:-) or :)		Open-mouthed	:-D or :d
	Hot	(H) or (h)		Sad	:-( or :(
	Left hug	({)		Right hug	{)}
	Vampire bat	:-[ or :[		Red rose	(F) or (f)
	Red heart	(L) or (l)		Broken heart	(U) or (u)
	Red lips	(K) or (k)		Star	(*)

### 3. COMMUNICATION NEEDS

Workplace IM studies tend to take a goal-oriented approach to communication, emphasising information exchange and coordination of activity over social conversations (Herbsleb *et al.*, 2002; Isaacs *et al.*, 2002; Nardi *et al.*, 2000). Even where the informal nature of IM conversations is highlighted, it is in relation to the accomplishment of work-related activities (Nardi *et al.*, 2000). Adults in the workplace have a set of role-related priorities, and their communication activity reflects these. The comparatively large amount of time that teenagers spend using communication tools such as IM simply cannot be justified in the same way. For teenagers, contact is more important than content (Livingstone and Bober, 2003), and their intense communication activity is a means of developing a sense of personal and group identity (Boneva *et al.*, in press). The exploration of personal identity comes through for example in the study by Talamo and Ligorio (2001) who report on a transnational educational project centred around a multimedia environment where participants are represented by avatars. Students in the project were much more likely to try out the different avatars available than were the teachers. Suler (1998) suggests that the characteristics of IM environments are especially conducive to teenage identity experimentation. The rarefied nature of personal representation and communication in IM means that it is possible to 'deconstruct' one's personality, and present only one facet at a time. Changing avatars can be a way of changing focus from one facet to another, as can changing one's IM name or image. At the same time, teenagers are experimenting with autonomy and strengthening the bonds with their peer groups as a way of learning about their place in the world (Suler, 1998; Boneva *et al.*, in press). As older teenagers become more confident with their own personal identity, their need for online experimentation diminishes (Schiano *et al.*, 2002).

#### 3.1 IM behaviour

A feature of IM that users consistently refer to as attractive is the ability to hold multiple simultaneous conversations (Lenhart *et al.*, 2001; Nardi *et al.*, 2000; Isaacs *et al.*, 2002; Boneva *et al.*, in press). IM allows teenagers to indulge their need for group belonging by connecting with a number of their peers simultaneously, but each of those conversations is also *private* being visible only to the participants (Boneva *et al.*, in press). Self-reported multitasking behaviour differs from one study to another. Schiano *et al.* (2002) report a maximum of five concurrent IM conversations, while Boneva *et al.* (In press) report as many as 16. Multitasking in IM environments allows teenagers the peculiar mix of group belonging and

private interaction in a way that chatrooms do not, and several studies have noted a switch in teenage allegiance from chatrooms to IM with the development of improved tools. While the adults in Grinter and Palen's study (2002) characterise public chatrooms as "a waste of time", teenagers' interest seems to wane with experience (Livingstone and Bober, 2003, p.17). Studies in the UK and America have shown that teenagers are generally aware of the dangers of sharing confidences with strangers online. The anonymity of the chatroom provides ample opportunity for deception, and many teenagers avoid them for this reason (Lenhart *et al.*, 2001, pp. 23-4; Livingstone and Bober, 2003, pp. 21-27).

Because IM conversations typically involve two participants, some of the problems encountered in chatroom conversations are less significant. In particular, the two main requirements for coherence noted by Greenfield and Subrahmanyam (2003), namely establishing who is participating in a particular thread and what constitutes a relevant response, are clearly less relevant with only two interlocutors. While their study describes a number of techniques for exploiting the graphical and textual characteristics of the chatroom to maintain coherence, the same functionality available in IM environments is more likely to be adapted for other purposes.

This study aims to explore the potential analytical uses of IM log file data for the capture of interpersonal relationships and communication activity among teenagers, and to find ways of representing that information in order to facilitate more detailed work.

#### 4. METHOD

MSN Messenger provides the facility to log all conversations to a series of files in XML format, making them amenable to manipulation. The data stored includes the content of the conversation, time and date values for all messages as well as the names of the participants. Within the log files, messages are grouped into sessions, each of which represents the activity carried out in a single MSN window from the moment it is opened until it is closed.

Several teenaged MSN users who were known to one of the researchers were invited to take part in the study by activating their logging facility and providing the accumulated data for analysis. Some were understandably wary and declined, but four informants were eventually found. Having obtained signed consent forms from their parents, all IM activity of these individuals was logged for a period of four weeks.

As a trust building measure, and as a protection of privacy, the content of individual messages was not used in this study. It was also felt that the

knowledge that their words would be later read and analysed might inhibit the natural behaviour of the informants. At the point the logs were harvested, the message content was hidden by applying a text filter that replaced all alphanumeric characters with 'x'. This preserved the word count of each message, any linguistic devices such as punctuation and emoticons. The XML document structure made it possible to do this for message content while leaving all other details intact including the names used by the participants. Because the content of conversations was removed, and because of the difficulties inherent in trying to identify MSN users from their session names, it was not deemed necessary to seek wider consent from the informants' contacts.

The result of the first data extraction phase was the series of spreadsheets shown in Table 2. The second phase of the analysis was performed within the spreadsheets shown in the table, or in modified copies. A further phase of analysis was the creation of a matrix of connections between IM participants that was used as input into the network visualisation tool, UCInet (Borgatti et al, 2002). This facilitated the production of a single network diagram representing the connections between individuals - the social network underlying the IM activity itself. Social network analysis is most frequently used to analyse data that has been gathered by questionnaires, from observation or from documents. Questions such as "Who are your three best friends?" or "Who do you ask for help with homework?" are often used to elicit information. This type of question can be problematic as accuracy in response depends on an individual 'choice on the day' and willingness to cooperate and reveal personal information (Scott, 2000). Using log file data avoids this limitation since it relates to actual (measurable) interactions between individuals.

*Table 2. Main spreadsheet files*

Name	msn_names.xls	msn_sessions.xls	msn_messages.xls
Contents	Participant names Identification numbers	Source (informant and file) Session number, start date/time End date/time Number of messages Mean message length Standard deviation Number of participants List of participant identification numbers	Source (informant and file), Session number Date/time Participant identification number Number of words Mean word length Standard deviation

## 4.1 Limitations

Four main problems were encountered while processing the log file data. The first of these was the lack of a reliable method for relating the variety of names used by an individual over the course of several sessions. This facility had been present in a previous plain-text version of the log file that had been used when planning this study, but disappeared with the introduction of the XML format. To overcome this problem, an additional step was included in the analysis whereby the informants were asked to verify the attribution of names to particular individuals.

Another issue with MSN Messenger was the duplication of log file data. Under certain circumstances, messages from sessions with more than two participants would be logged to more than one file. This issue is discussed below, but was not explicitly dealt with at this stage.

The third problem related to the processing of some of the exotic characters used by participants as part of their names. The *awk* text processing tool (familiar from Unix) was not always reliable in its ability to handle Unicode characters, and a small number were therefore lost and replaced by question marks in the filtered files. This did not affect the main analysis, since the majority of names were legible and identifiable.

Finally, a large amount of data was produced, and this created issue for some of the tools being used in the analysis. In particular, both Microsoft Excel, and the spreadsheet embedded in UCINET are limited to 255 columns. This was therefore the maximum number of participants that could be accommodated in the network diagram. Some of the participants with very little activity in terms of the number and size of messages were therefore omitted.

## 5. RESULTS AND DISCUSSION

### 5.1 Overall characteristics

In total approximately 30,000 messages were identified, spanning 883 sessions. The earliest logged session was on 4/10/2004 and the latest on 10/11/2004. The exact number of sessions was difficult to determine because of the log duplication mentioned above. The difficulty in attributing names to individuals also made it impossible to give an exact total number of participants, but the approximate figure is 180. This includes the informants themselves, their immediate interlocutors, and a number of other individuals involved in IM sessions into which the informants were invited.

In general, the results obtained confirmed those reported by other authors regarding the number of participants in a typical IM session, and the nature of IM messages. The average number of participants in an IM session was 2.11 with a standard deviation of 0.55. Of the nine sessions with more than three participants, eight were between five and seven, with a single outlier of ten. Clearly, the one-to-one conversation is the norm in this set of data.

The number of words per message was very low – 28% of all messages consisted of a single word, and 52% were three words or less. This again confirms the expectation that IM messages are typically very short. Many of these very short messages could be backchannel contributions intended simply to encourage the other participant and to keep the conversation going. A further factor could be the intense use of linguistic shorthand, contractions and emoticons adapted from email, chat and text messaging (Thurlow, 2003). Some of these forms have the function of entire sentences, and therefore may give a false impression of the significance of a very short message.

Looking at the average number of words per message per participant gave a mean of 4.2 and a standard deviation of 5.11 on the current dataset. The higher mean and standard deviation suggest that there is variation between IM users, each having a personal preference for longer or shorter messages. This was not, however, controlled for the number of message generated by each participant and remains a suggestion at this point.

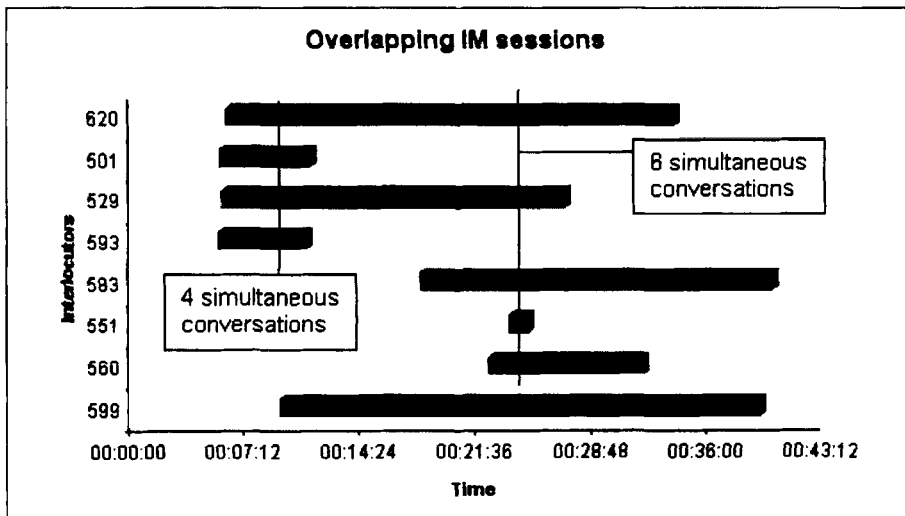


Figure 1. An example of multitasking from a single informant



## 5.2 Conversational behaviour and multitasking

Comparing the start and end times of sessions from a single informant's log files revealed those sessions that overlapped in time, ie those occasions on which the informant had more than one conversation in progress with different interlocutors. This information was used to produce timeline diagrams such as the example in Figure 1. One model of multitasking behaviour reported by Grinter and Palen (2002) is that of a central group conversation with peripheral one-to-one conversations. Evidence for this would have come from the appearance of group sessions sharing interlocutor numbers with simultaneous one-to-one sessions; however, no evidence for this was seen in this limited set of data suggesting a much less structured arrangement of conversations.

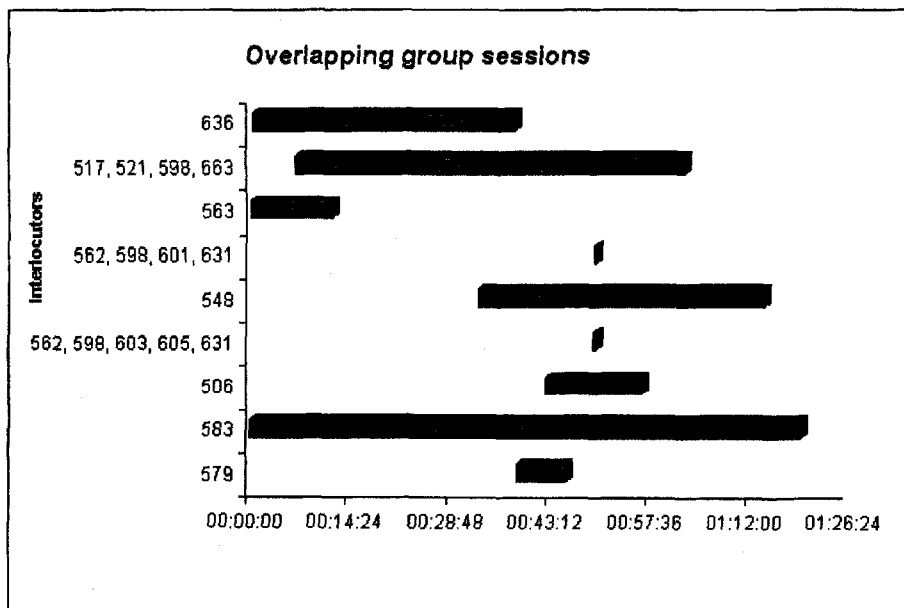


Figure 2. An example of simultaneous group and individual sessions from one informant

In Figure 2, group sessions are indicated by a series of interlocutor numbers on the vertical axis. No group participants are simultaneously involved in one-to-one conversations with the informant. An interesting feature though is the lower pair of group sessions, (562, 598, 601, 631) and (562, 598, 603, 605, 631). These sessions share three participants, and clearly relate to the same message data as indicated by the identical bars on the chart. The reason that they appear as separate sessions is an artefact of

the way in which messages may be logged to more than one file by MSN Messenger.

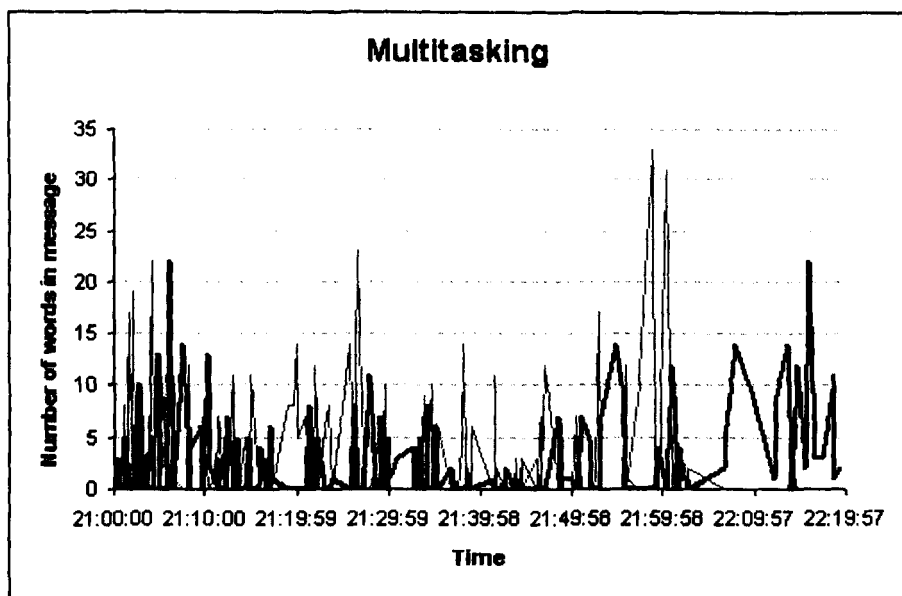














Figure 3. Message activity within simultaneous sessions.

Figure 3 represents the activity of the informant in the sessions shown in Figure 2 in terms of the number, size and timing of messages sent. The heavy line corresponds to the conversation with interlocutor 583 in Figure 2, while activity in all other sessions is shown in grey. Although the highlighted session is the longest, there is no other sense in which it appears to be dominant. Activity is relatively constant all the way through, with some variations where the informant is composing messages to other conversations. This is noticeably the case at the point leading up to the highest peaks on the graph. These long messages – over 30 words each – obviously took longer to compose than the usual rapid-fire contributions, and it is understandable that this should interrupt flow in the other sessions. This suggests, however, that an additional pressure to keep message length short may come from the need to maintain all active conversations. Long messages to one session may cause unwanted delays in others, and a hypothesis for future investigation might therefore be that average message size decreases with the number of simultaneous conversations.

5.3 Identity

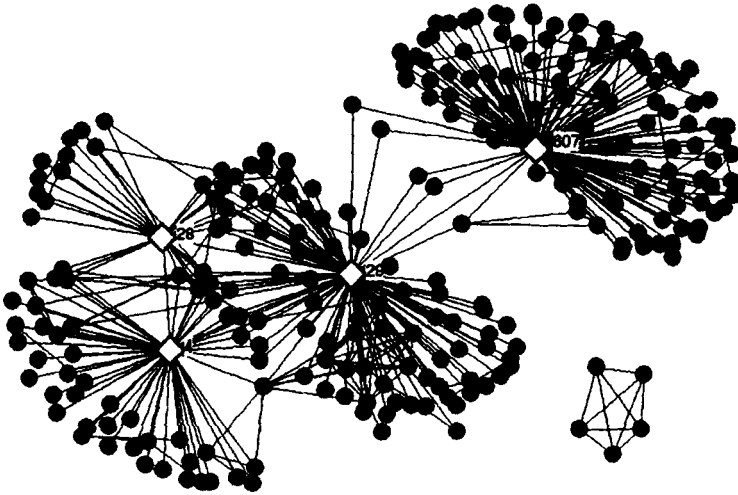
Before starting the analysis of the data, it was expected that individuals would use a number of different names. What was unexpected was the scale of this phenomenon. One participant was found to have changed names 119 times during the logging period. This individual was a regular session partner with one of the informants, and was therefore responsible for a large number of messages in the logs. When the number of messages per user was taken into account, it was found that on average a participant changes name after every 22 messages. This, of course, is a very rough figure and does not take into account messages that participants may have exchanged with interlocutors other than the four informants, i.e. messages not captured in the dataset. Nevertheless, it is a good indication of the dynamic nature of identity in IM. There were numerous examples of the participants taking advantage of the visual qualities of the medium, including using emoticons like those in Table 1 to enhance their names. Other typographical creativity involved the use of Unicode characters as in the examples in Table 3.

Table 3. Example MSN names

1	  \$tæc€*  []æ@r€ñ   vi@ {i  (3iLi*   €mmæ  Jøhñ  
	(Text parts read as: STACEY, DARREN, VICKI, BILLY, EMMA, JOHN )
2	I'm šupë@gi@l änd I'm hë@ë tő šävë thë wö@ld
3	(\_/ (=':'= (.(''').)  ('')_('') B†t†h†y bŭññ† catching things and eating their insides 

6. NETWORK ANALYSIS

The representation of the interpersonal links generated using UCINet exhibited the expected explosion of connections surrounding the four informants, shown as light square nodes in Figure 4. Perhaps more interesting though is the information captured about the wider network in which they are located. A number of links between other individuals are identified such as those within circle A, as well as a small group of five in circle B who do not appear related to the main population at all.



*Figure 4. Social network*

The nodes in group B are in fact the result of two sessions into which one of the informants was invited, but did not contribute any messages. There is therefore an implied passive link, but only active contributions are represented in Figure 4. A possible line of future enquiry might be to investigate the strength of the links in graphs like the one above, and to test whether passive links are a true indication of peripheral group membership.

Figure 4 also shows that certain individuals such as those in areas C and D hold pivotal positions in the overall structure. The network diagram is therefore an excellent way to identify those people who might profitably be included in a future large-scale study.

## 7. CONCLUSION

The work presented here has identified a number of opportunities for quantitative analysis presented by IM log file data. It has also identified some limitations inherent in the various technologies used that will need to be overcome in future stages of this study. Future work would need to be based on a larger sample of selected IM users to allow conclusions to be drawn with any confidence, but the value of this kind of information for triangulation with qualitative data is clear.

## REFERENCES

- Boneva, B., Quinn, A., Kraut, R., Kiesler, S., Shklovski, I. (In press). Teenage Communication in the Instant Messaging Era. In Kraut, R., Brynin, M., Kiesler, S. (Eds). *Domesticating Information Technology*. Oxford University Press.
- Borgatti, S.P., Everett, M.G. and Freeman, L.C. 2002. Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.
- Eldridge, M. and Grinter, B. (2001). Studying Text Messaging In Teenagers. Conference of Human Factors in Computing Systems (CHI) Workshop: Mobile Communications: Understanding Users, Adoption & Design
- Farmer, R. (2003). Instant messaging - Collaborators Tool or Educator's Nightmare! In *Proceedings of the Web-Based Learning Conference 2003 University of New Brunswick* <http://naweb.unb.ca/proceedings/2003/PaperFarmer.html>.
- Greenfield, P. M. and Subrahmanyam, K. (2003). Online discourse in a teen chatroom: New codes and new modes of coherence in a visual medium, *Applied Developmental Psychology*, volume 24, pp. 713-738.
- Grinter, R. E., and Palen, L. (2002). Instant Messaging in Teen Life. Proceedings from Conference on Computer Supported Cooperative Work 2002. New York: ACM Publishers.
- Herbsleb, J.D., et al. (2002). Introducing Instant Messaging and Chat in the Workplace. Conference of Human Factors in Computing Systems (CHI), Minneapolis, MN. pp 171-178. [http://www.crew.umich.edu/Technical%20reports/Herbsleb\\_Atkins\\_Boyer\\_Handel\\_Finholt\\_Introducing\\_instant\\_messaging\\_12\\_10\\_01.pdf](http://www.crew.umich.edu/Technical%20reports/Herbsleb_Atkins_Boyer_Handel_Finholt_Introducing_instant_messaging_12_10_01.pdf)
- Isaacs, E., Walendowski, A., Whittaker, S., Schiano, D.J. and Kamm, C. (2002) *The Character, Functions, and Styles of Instant Messaging in the Workplace*. In Proc. ACM CSCW Conf., ACM, New York, 11-20. <http://www.izix.com/pro/lightweight/images/IM.pdf>
- Lenhart, A., Rainie, L. and Lewis, O. (2001). Teenage life online: The rise of the instant-message generation and the Internet's impact on friendships and family relationships, *Pew Internet & American Life Project*, Washington, D.C.
- Livingstone, S. and Bober, M. (2003). UK Children Go Online: Listening to Young People's Experience. Available online at <http://personal.lse.ac.uk/bober/UKChildrenGoOnlineReport1.pdf>
- Nardi, B. Whittaker, S. and Bradner, E (2000). Interaction and outeraction: instant messaging in action. *Proceedings of the 2000 ACM conference on Computer supported cooperative work* ACM:USA pp.79-88 [http://www.research.att.com/~stevew/outeraction\\_cscw2000.pdf](http://www.research.att.com/~stevew/outeraction_cscw2000.pdf)
- Oikarinen, J. (n.d.) IRC history. Available online at [http://www.irc.org/history\\_docs/jarkko.html](http://www.irc.org/history_docs/jarkko.html) [accessed 9 Nov 2004].
- Peris, R., Gimeno, M. A., Pinazo, D., Ortet, G., Carrero, V., Sanchiz, M. & Ibanez, I. (2002). Online chat rooms: Virtual spaces of interaction for socially oriented people. *CyberPsychology & Behavior*, 5, 43-51.
- Schiano, D. J. et al. (2002). Teen Use of Messaging Media. Human Factors in Computing Systems: CHI 2002. Extended Abstracts. NY: ACM. <http://hci.stanford.edu/cs377/nardi-schiano/CHI2002.Schiano.pdf>
- Scott, J. (2000) Social network analysis: a handbook. London: Sage
- Smith, N. (2002). Teaching as Coaching, Helping Students Learn in a Technological World, *EDUCAUSE Review*, May/June 2002. <http://www.educause.edu/ir/library/pdf/erm0233.pdf>

- Sulake Labs (n.d.) Habbo Hotel. <http://www.habbohotel.co.uk/habbo/en/>
- Suler, J. (1997). From ASCII to holodecks: Cyberpsychology of an online multimedia community. *The Psychology of Cyberspace*, <http://www.rider.edu/suler/psycyber/palsummary.html>
- Suler, J. (1998). Adolescents in Cyberspace: The Good, the Bad, and the Ugly. *The Psychology of Cyberspace*, <http://www.rider.edu/suler/psycyber/adoles.html>
- Suler, J. (2000). Bringing Online and Offline Living Together: The Integration Principle. *The Psychology of Cyberspace*, <http://www.rider.edu/suler/psycyber/integrate.html>
- Suler, J. (2004). The Final Showdown Between In-Person and Cyberspace Relationships. *The Psychology of Cyberspace*, [www.rider.edu/suler/psycyber/basicfeat.html](http://www.rider.edu/suler/psycyber/basicfeat.html) (article orig. pub. 1997).
- Talamo, A., & Ligorio, B. (2001). Strategic identities in cyberspace. *CyberPsychology & Behavior*, 4(1), 109-122.
- Thurlow, C. (2003). Generation Txt? The sociolinguistics of young people's text-messaging. *Discourse Analysis Online*. <http://www.shu.ac.uk/daol/articles/v1/n1/a3/thurlow2002003-paper.html>
- Williams, P. (1999). The net generation: The experiences, attitudes and behavior of children using the Internet for their own purposes. *Aslib Proceedings*, 51(9), 315-322.
- Zakon, R. H. (2004) Hobbes' Internet timeline v7.0. Available online at <http://www.zakon.org/robert/internet/timeline/> [accessed 9 Nov 2004]