When we think of public space we often think of the agora in ancient Athens, or architecture of more recent origins like streets, churches, and schools. In other words, we think of the “bricks and boards” type of public space. But, in addition to architecture, we also know that media can function as public spaces. In the nineteenth century, media, such as novels and newspapers, were essential to the creation of then-new “imaginary communities” that we now know as nations, states, and cities. Of course, nations, states and cities existed before the nineteenth century, but the high-speed printing press and other media innovations of the time transformed these existing social formations into new forms that were often radically different than their pre-nineteenth century instances. Similarly, the twentieth century introduced new forms of media and public space (e.g., television) which – in turn – transformed the social formations of the nineteenth century (e.g., political parties). Contemporary media -- like the Internet -- are creating new social formations that some sociologists call the “network society.” So, the point is this, new media -- new public spaces -- facilitate or engender new social formations.

The new public spaces that I am interested are very large-scale conversations. On the one hand, very large-scale conversation is a medium that we -- who inhabit the Internet -- know very well. Usenet newsgroups, large email listservs, and other places on the Internet where large volumes of email are exchanged are good examples of very large-scale conversation. On the other hand -- from the perspective of the history of media and public space -- very large-scale conversation is an entirely new and mostly unexplored phenomenon. At no other point in history have we had a medium that supports many-to-many communications between hundreds or thousands of people. VLSC takes place across international borders, often on a daily or hourly basis. Unlike with older media -- for instance, telephones -- participants in these very large-scale conversations usually do not know the addresses of the others before the start of a conversation. VLSC on the Internet is a new space where people who might never have known of one another’s existence are now forging bonds. A space has opened in which, with some good luck and hard work, new forms of relations might be forged. Naturally, this vision about the future of a networked society -- involving a proliferation of global conversation -- implies an optimism and perhaps, as some critics would have it, a naïveté.

The vision outlined above is a goal, not a fantasy. As a goal it implies a very large amount of work that must be done in order to address it. The work is of two sorts. Firstly, while it is true that small-scale conversations -- e.g., interactions between thirty or fewer people -- can be understood, no one really has any idea (or theories) about how current understandings of conversation might be scaled up to handle the huge,
many-to-many interactions that have now become commonplace on the Internet. Secondly, although there do now exist many computer-based, information technologies, it is well-known – even with current theories of conversation and language – that conversation is about a lot more than simply the exchange of information. Do you always make conversation to exchange information? Of course not. Sometimes we converse to introduce ourselves to someone else, to acknowledge another’s presence, to have some fun, to simply cut the silence. Consequently, we cannot assume that machines – i.e., computers and computer networks – designed for the exchange of information will be wonderful media for all kinds of conversational exchange. In other words, a large amount of work must be done to transform existing information technologies into conversation technologies.

In this essay I present some first small steps towards the design work I envision above. The design challenge is this: how can computational tools be built to help us understand and participate in these very large-scale conversations? The challenge can be posed as a problem from the perspective of a participant in one of these very large-scale conversations. If I want to participate in one of these huge discussions my problem is this: How can I listen to thousands of others; and, conversely how can my words be heard by the thousands of others who might be participating in the same conversation? Phrased as a design problem, the question becomes the following: What software can be designed to help participants navigate these new public spaces? Towards this goal I have designed the Conversation Map system.
In principle, one can use the Conversation Map like Netscape Messenger, RN, Eudora, or any other conventional news or email reader. The figure below is a picture of the main screen of the Conversation Map system. Like a conventional email reader, the Conversation Map loads in a number of email messages and presents them in an interface that allows one to read the messages one-by-one and/or see some of the connections (e.g., the “threading”) between messages. Unlike a conventional email reader, the Conversation Map allows one to see quite a few new relationships between the messages (and message authors) that might otherwise have been invisible or very difficult to see.

Given a few hundred, or even a few thousand, email messages, the Conversation Map system analyzes those messages using a set of computational linguistics and sociology techniques. This automatic analysis yields three sorts of interrelated summaries. One summary it generates – the social network – reveals who is talking with whom. The second summary lists the themes of discussion that are important to the conversation embodied in the messages. The third summary is perhaps the most interesting of all. The Conversation Map system generates a sort of rough-draft thesaurus. This thesaurus can be understand as a network of definitions or metaphors of the discussion that have emerged over the course of the conversation. It might reveal that the group is inventing new words or new ways of talking about known subjects. To do this, all of the participants’ language -- i.e., the content -- of the email messages is analyzed and “summed together” using several techniques of computational linguistics.

The picture shown below was generated by the Conversation Map system after it analyzed about 1300 messages posted in the spring of 1999 to the Usenet newsgroup soc.culture.albanian. It is a graphical summary of a large argument that engaged Albanians, Serbs, and others.
A screen-shot of a Conversation Map automatically generated by the system

**Social Networks:** The upper left quadrant of the Conversation Map shows a set of social networks that record who -- within the group of conversational participants -- is reciprocating with whom. By “reciprocating” I mean who is mutually responding to and/or quoting from whom. Thus, if I were to post a message to the newsgroup and then you were to respond to it and then, later in the discussion, you were to post to the newsgroup and I replied to -- or quoted from -- your message, then the two of us would be -- according to my definition -- said to be reciprocating with one another. Reciprocation is an important property of smaller scale discussions that has been explored by, for instance, researchers in the field of conversation analysis and so I believe this may also be an important property of large-scale discussions. As reciprocating participants of the discussion, you and I will appear as nodes in the graphs representing social networks and a line will be drawn between us. If we reciprocate many times over the course of the conversation we will be plotted close together. In contrast, those pairs of participants who reciprocate only once will be plotted relatively far apart. Note that posters who spam the group with many messages, but who receive no replies do not even show up on the graph. Those participants who show up closely connected are pushed to the middle of the graph and can be understood as virtual mediators of the newsgroup. I say participants in central positions of the social network are “virtual moderators” because most of the analyses I have done have been of unmoderated, public discussion spaces on the Net. To end up in such a position one needs -- not only to post many messages – but also to have others in the group reply to
or quote from many of one's messages. So, the social network display acts both as a filter for spammers and a means to identify some of the main players in a discussion.

“Themes” of Conversation: The menu in the upper-middle of the interface lists the “themes” of the conversation. (I put “themes” in quotes because, according to the terms of linguistics, what is being calculated by the system is – strictly speaking – not the themes, but the lexical ties between messages; a necessary, but not sufficient, step in the determination of the themes of discussion.) Let's say I post a message about football, and then you respond with a message that includes some reference to baseball. Then, perhaps later in the discussion, you post a message about skiing and I respond with one concerning skating. Our reciprocation will be represented in the social network, but some approximation to the theme of our exchange will also be listed in the menu of “themes.” In this case, since football, baseball, skiing, and skating are all sports, the term “sports” might be listed on the menu of themes. Calculating that these four terms are all sports requires, of course, a machine-readable thesaurus. The thesaurus employed in the Conversation Map system is WordNet, a lexical resource created by George Miller, his colleagues, and students at Princeton University.

Semantic Network: One way to understand the difference between the menu of themes and the graph depicted in the upper right-hand corner of the interface is this: while construction of the menu of themes requires the use of a predefined thesaurus, like WordNet, the calculations performed to create the semantic network in the upper right-hand corner do not use a thesaurus, but, rather, automatically generate a rough-draft thesaurus. To create a rough-draft thesaurus the Conversation Map system does the following: First, the content of all of the messages exchanged during the conversation is parsed -- i.e., subjects, verbs, objects and some other modifying relations are identified between the words of each sentence in the texts of the messages. Next, for each unique noun mentioned in the corpus of messages a profile is built. By “profile” I mean that, for each noun, a vector is created that records all of the verbs for which the subject functioned as a subject; all of the verbs for which the noun functioned as an object; all of the adjectives which modified the noun; etc. Once a profile has been calculated for each noun, the nouns’ profiles are compared to one another and each noun’s nearest neighbor is identified. If two nouns are nearest neighbors then, according to this calculation, they appear in similar contexts. Or, to put it more plainly, if two nouns have similar profiles, then they can be said to have been “talked about” in similar ways by the participants in the discussion. On the semantic network, if two nouns are nearest neighbors, then they are plotted as two nodes connected to one another.

Why, one might ask, is this sort of analysis of interest for the navigation of very large-scale conversations? To answer this question, I compare this sort of analysis with some work done by the cognitive scientists George Lakoff and Mark Johnson. Lakoff and Johnson wrote a book entitled *Metaphors We Live By*. The book is filled with a set of metaphors that Lakoff and Johnson claim are central to “our” (presumably, middle-class, white, American) culture. In this book, for instance, they claim that one emergent metaphor of our culture is that arguments are buildings. As part of their method to
argue for the validity of insights like this, they show how two nouns, which might \textit{a priori} be considered to be completely unalike one another -- like the two terms "argument" and "building" -- show up in very similar contexts. For example, one can say "The building is shaky" but one can also say "The argument is shaky." One can say "The building collapsed" and also "The argument collapsed." Similarly, both buildings and arguments can be said to have "foundations," "to stand," and "to fall," "to be constructed," "to be supported," "to be buttressed," etc. A set of similar sentences of this sort provides an empirical means for thinking about and discovering how definitions and metaphors are produced over the course of a large amount of discussion. Thus, this tool for automatic, rough-draft thesaurus generation can be seen as "training wheels" to allow us, within the context of a specific conversation, to begin to generate the sorts of hypotheses that Lakoff and Johnson explore in their book. So, the Conversation Map gives one some data exploration/navigation tools to start to understand how different conversations differ from one another according to the metaphors and definitions that are produced by the collective efforts of their participants.

\textbf{Message Archive:} The lower half of the interface is a graphical representation of all of the messages that have been parsed and analyzed by the Conversation Map system. Messages are organized into threads where a thread is simply defined as an initial post, all of the responses to the initial post, all of the responses to responses, etc. The threads are organized in chronological order from upper left to lower right. The lower half of the screen is divided into a grid and so the first thread posted to the newsgroup appears in the upper left-hand corner and the last thread posted appears in the lower right-hand corner. If a thread contains many messages it shows up as an almost completely green square on this display. If a thread contains few messages, then it shows up as an almost completely black square. Thus, by reading across and down -- from upper left to lower right -- the lower-half of the screen provides a rough guide to the posting activity in the newsgroup over the period of time spanned by the messages.

Having briefly described the Conversation Map, I would like to return to the points about public space and new social formations with which I began this paper. If, in fact, very large-scale conversation is a new kind of public space, then what are the new types of social formations that are facilitated or engendered by it? I am exploring four of these new sorts of social formation by myself and also in collaboration with several social scientists.

One area of online conversation that I am interested in might be characterized as national and international discourse about mass media, especially television. Certain television shows that are broadcast today are watched by audiences all over the world. For instance, the television show \textit{Baywatch} has over one billion viewers. Regardless of your opinion of the content of this show, the sheer size of its audience makes it worthy of serious consideration. Internet discussions about television shows make it clear that audiences are not simply "masses," but, in fact, highly interconnected groups of people negotiating the meaning and significance of shows, characters, and studio politics in daily, international, online forums.
With Joseph Dumit, an anthropologist and professor in the Science, Technology, and Society Program at MIT, I am exploring very large-scale conversations that focus on medicine and health issues. Specifically, we are trying to understand illness-based social movements that use online organizing as an essential means with which to establish themselves as a united group. Dumit calls some of these illnesses, “illnesses you have to fight to get.” Gulf War syndrome, multiple chemical sensitivity, chronic fatigue syndrome, and attention deficit disorder are all examples of “illnesses you have to fight to get.” With these it is often the case that a set of people claim to be suffering from a sickness that the medical and insurance industries refuse to recognize as existent. Consequently, the sufferers mobilize amongst themselves to get the status of their illness changed and thus recognized so that they will be treated by medical practitioners and reimbursed for medical costs by their insurance companies. Since it is rarely the case that sufferers are geographically localized or members of some pre-established lobbying group, the Internet is an essential forum for their self-organization as cohesive groups.

Probably the most widely recognized of the new social formations engendered by online communications is the Open Source Movement. Technical discussions and exchanges over the Internet have resulted in new pieces of software -- e.g., Linux – that constitute a new social and economic force and challenge existing products and conventional production methodologies -- e.g., Microsoft's production of the Windows operating system. Very large-scale conversations devoted to technical issues are often some of the most tight-knit and productive of online forums.

Finally, the newsgroup analyzed for the example Conversation Map screen-shot described above (i.e., the graphical summary of soc.culture.albanian during several weeks of the spring of 1999 during the war in Kosovo) is an example of people functioning as -- what I call -- “citizen diplomats.” Ordinary citizens, without diplomatic appointments from their governments, are now conducting international relations through their discussions with other ordinary citizens in other hostile and friendly nations. These discussions are often very messy as they are untethered by the diplomatic secrecy and journalistic finesse applied to reporting the process and outcome of negotiations conducted by official diplomats. However, even in the form they take today, these very large-scale conversations point to new possibilities in the conduct of international relations.

So, I argue these four sorts of groups -- media audiences, illness-based social movements, the Open Source Movement, and citizen diplomats -- constitute examples of the new kinds of social formation that are being engendered or facilitated by the new public spaces of very large-scale conversations.

Unlike other newsgroup browsers, the Conversation Map makes some of the social relations produced through online conversation visible by diagramming them as social networks. It identifies and sorts the themes of discussion according to their importance in the social network; the more links of the social network supported by a theme -- i.e., the greater the number of pairs of people exchanging messages about a given theme --
the higher the theme is rated. Thus, the menu of themes gives one some insight into what is central to the conversation. By overlaying the results of the theme analysis on the computed social networks, I have demonstrated how a sketch can be computed of what sociologists have long called “social cohesion.” The Conversation Map shows how computational tools might be built to empirically explore social cohesion, its differences, and importance to various online communities. Finally, if one wants to explore hypotheses concerning possible emergent metaphors of discourse – like Lakoff and Johnson do -- the Conversation Map provides a tool that can help one make these sorts of hypotheses. All of these dimensions of online conversation -- social networks, themes of discussion, social cohesion, and emergent metaphors and definitions -- are indicative of the new social formations produced in the public spaces of very large-scale conversation.

There is not enough space in this essay to demonstrate differences between the Conversation Maps of VLSCs productive of the four, new, social formations identified above. Nevertheless, it should be clear that – as an email/newsgroup reader – the Conversation Map system is designed to show aspects of these public spaces that are difficult to perceive with a conventional newsreader. The interested reader can find several example Conversation Maps here: http://www.sims.berkeley.edu/~sack/CM

Participants in the emergent medium of VLSC will need new technologies to orient themselves in navigating through and interacting with these new public spaces. Moreover, the group of participants involved in a VLSC need a means to recognize themselves as socially, politically, or economically (in)cohesive entity for purposes of self-governance. The Conversation Map system is an attempt to provide a first step towards building tools for community self-recognition and self-governance.

In synoptic form, I have attempted to convince you of the following points: Firstly, public space provides a place for the creation of shared social and linguistic relations. VLSC constitutes a new public space for emergent communities, democratic politics, discussion and debate. Secondly, insights from sociology, linguistics, and graphical interface design need to be combined to understand the phenomenon of VLSC. And, thirdly, new summarization technologies and content-based browsers can be designed by combining and extending the text analysis techniques of quantitative-oriented sociology and computational linguistics. The Conversation Map is an example of this new kind of technology. I hope that this sort of technology will be appropriated by online communities for democratic purposes.

More information about this project can be found in the following publications and at the web site http://www.sims.berkeley.edu/~sack/CM

