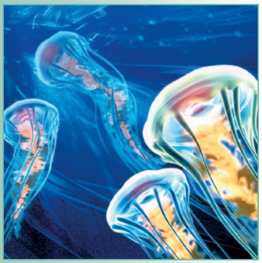

Using Egocentric Networks to Understand Communication

Social-network analysis generally helps researchers understand how groups of people interact. The author uses small-scale egocentric social networks, based on volitional, explicit connections, to understand how people manage their personal and group communications. Two research projects using this approach show that such networks can give researchers important insight into the people who communicate online. Soylent, a project based on email, shows several common patterns in social interaction. The Roles project, based on Usenet newsgroups, suggests that various online social spaces can behave very differently from each other.

Danyel Fisher
Microsoft Research

We can learn a lot about people from who they talk to — whether they talk to sociable, well-connected people or to unconnected individuals, for example. Is the person part of a large group whose members all talk to each other, or does he or she bridge social worlds? Examining such connections between people can teach us how they operate socially.

Online communication, in particular, is an increasingly important way to get information from, and keep in contact with, each other. Communication through a computer can be recorded and then analyzed; we can later use that analysis to develop systems that are more aware of how people interact online.

Although much social-network analysis attempts to examine complete, large-scale networks — ones in which every connection between network members is visible — my work takes a more restrictive view. Using *egocentric network-analysis* techniques — which examine only people's immediate neighbors and associated interconnections — helps us learn about how individuals correspond with their social networks.

I've used this approach in two different research projects: the Soylent project, which I conducted as a graduate student at the University of California, Irvine, explores and describes interaction patterns in email.[1] The Roles project, conducted at Microsoft Research, applies social networks to Usenet messages.[2]

In both cases, egocentric networks of

 Published by the IEEE Computer Society 1089-7801/05/$20.00

immediate connections provide network analysts and researchers with interesting and useful data about the roles individuals adopt and the types of interactions that develop among their social partners.

A Perspective on Social Networks

As in all network analyses, the networks in this study comprise interconnected nodes and edges — an edge connected to a pair of nodes represents a pair of people and the relationship that links them. We can measure these networks quantitatively, using techniques from both sociology and graph theory to describe people's interactions as recorded within the network. Network analysis then gives us a structural description of the group's relationships.

The networks in both the Soylent and Roles studies are somewhat different from many network analyses studied in other contexts. First, the edges in these networks are volitional — that is, every edge is motivated by a specific social choice. Such choices are linked to the people with whom individuals interact, and network ties are positive evidence of meaningful relationships. Although social constraints and local norms might drive these relationships, they are intentional. Conversely, the nonvolitional networks that many network analyses study might measure more unintentional effects — examining edges such as "has sat next to on an airplane," for instance — and can collect information about mobility and proximity (which makes them very useful for studying disease transmission, media exposure, and so on). However, they have a limited ability to measure social choice and preference.

The particular edges that Soylent and Roles examine are based on personal communication. Each edge is based on at least one message, sent from a single person and directed to another. Additionally, each message represents time and effort — that is, a decision that the message is worth sending and that it's worth sending to a specific person.

Unlike with many network analyses, I don't assume that networks in the Soylent and Roles projects are transitive in any meaningful way. Many studies identify "gatekeepers"[3] who control flows such that information can move only from A to C via B. For example, a salesman (B) might carefully guard his contacts (C) from his colleagues (A). In the systems I discuss, this doesn't necessarily occur. That A connects with B and B connects with C doesn't mean that B is a middleman between the two. Indeed, in these networks, A could talk directly with C as desired: the lack of a direct connection implies that A chooses not to contact C. In the case of Usenet newsgroups, for example, even the most casual reader can access all posters' names; anyone can reply to a previous message, thus forging the ties that the study observes.

Seeing from One Viewpoint

In addition to these constraints, the Soylent and Roles projects share two other important themes. First, they use only information that's available to the system and to the user. Many social-network analyses present a view of networks based on data that no single person could know. Frequently, they're based on extensive data collection or privileged access to servers: they list all the connections within a single organization or computer network. Such projects can discover previously unknown information that no single person can find out. In one research project from Hewlett-Packard laboratories, for example, the authors identified some of the informal communities within their organization.[4]

Soylent and Roles aggregate information that's not necessarily novel — users can already access the information, although they can't ordinarily view it from a network-based perspective. Aggregating this known data is useful initially for developing information-visualization designs that can present users with information in new ways (see the "Network Visualizations" sidebar for more information). The insights we can draw from these network analyses can also lead to the development of new user-interface tools.

Additionally, Viegas and colleagues argue that showing personal information to users lets them navigate their social histories, examining who they've encountered and what experiences they've had. Sharing a social network, therefore, can be like sharing a photo album.[5] Indeed, both Viegas and Nardi and colleagues[6] note that, for many users, visualizations of personal networks become a locus of storytelling.

Finally, using only information that is available to users has beneficial privacy trade-offs. In a traditional whole-network study, researchers hold valuable and potentially dangerous information that an individual might be afraid to expose to researchers or other participants. When using only local data, we lose a global scope, but need

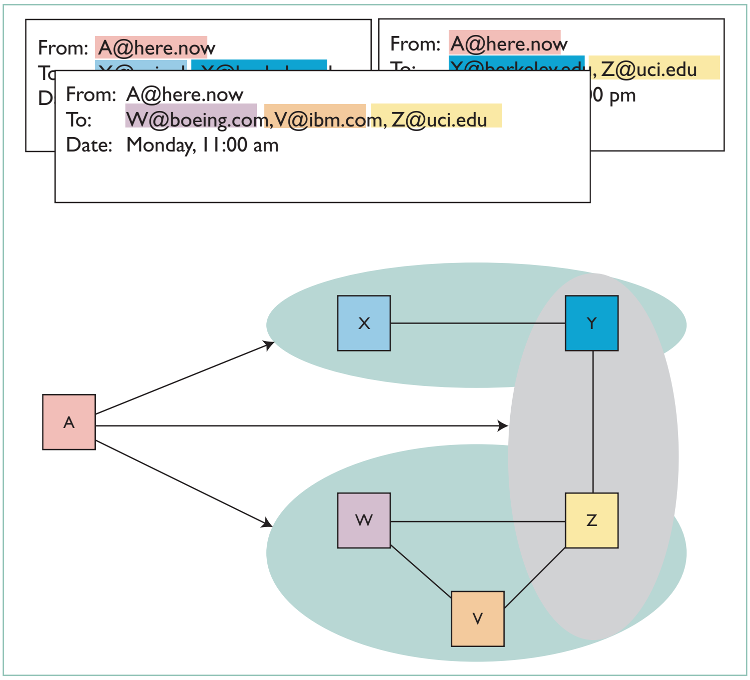

Figure 1. A schematic view of ego-centric networks, based on outgoing mail. Soylent assembles three email messages from A into a network that shows the connections between A's correspondents.

not worry about the potential harm of exposing private data. Soylent presents personal email to users via a tool that lets them review only their own data; no one else needs access to it. The Roles project examines groups that are publicly archived by Google. By presenting only information that users can control, we make it more comfortable for them to participate in the study, and make the data potentially less harmful — they need not fear exposing private information.

The second important aspect of my approach is that limiting the examination to a single-person perspective generates several salient questions for researchers — who is this person connected to, for example, and how are his or her neighbors interconnected? These, in turn, have design implications: a person connected to a sparse network might be particularly interested in a tool that helps juggle different contexts; a person linked into a dense network might prefer a tool that puts their information in one place.

In the Soylent project, for example, we start with the system's user (called Ego), who has sent email to a number of different people. Although Ego presumably knows something about who these people are, his email client doesn't necessarily make this context accessible. Even if it stores the fact that Ego has communicated with them in the past, it's unlikely to have stored information with any depth, such as who initiated the last conversation, when it happened, or who else was involved. We focus, then, on Ego's neighbors, one at a time, trying to understand their connections to each other and to the entire network. In the Roles project, we focus on one member of each newsgroup at a time, trying to understand their particular roles.

The Soylent Project

In today's computer systems, "people" appear only incidentally, as disconnected names in address books, contacts in mailing lists, or document authors. They exist out of context with one another. The Soylent project, which I conducted from 2002 to 2004 as part of my dissertation work at UC Irvine, aims to rethink the way we currently use both email and desktop space. My goal with the project was to connect people to each other and to their documents — placing people and projects within a shared social context. I believe this is a valuable direction for future user interfaces. Today, most systems separate documents from their authors: documents are stored in file systems, whereas information about people is stored in email and contact-management tools, requiring users to manage the two separately. By storing both in a shared space, it's possible to manage and simplify this complexity.

Soylent analyzes personal social networks from email archives, looking only at outgoing email from a single user at a time. Basing the network entirely on outgoing mail lets us view users' perceptions of their social worlds. Although incoming mail might be spam, mass emails, or irrelevant messages, we can assume that each message that users send is relevant to them — at least when they send it. Based on a user's activity history, Soylent ultimately assembles a network of connected people that closely resembles that of Social Network Fragments[5] and Personal Map,[7] both of which plot the interconnections between a user's contacts and present this information to the user.

By examining the network developed by messages co-addressed to two or more contacts, we can determine what social networks the user considers relevant. Many email messages from me to both Gina and Phil, for example, suggests that I believe Gina and Phil have common interests. That I infrequently (if ever) send email addressed to both my brother and my officemate might suggest

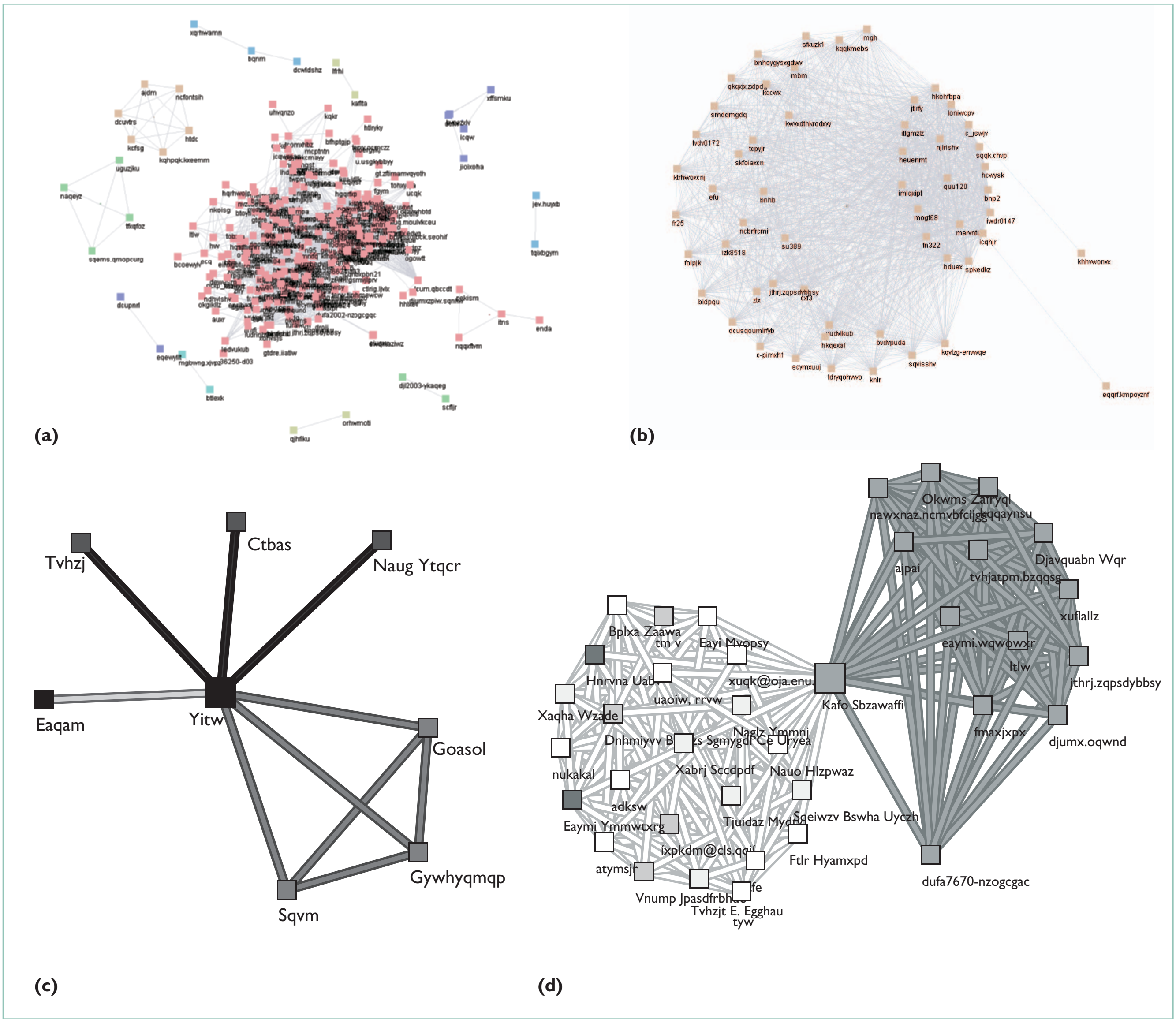

Figure 2. Four network patterns. (a) Disconnected clusters occur when groups of people aren't at all connected. (b) The onion pattern shows a tightly connected group at the center of a loosely connected group. (c) In the nexus pattern, a single person shares multiple contexts with Ego. (d) The butterfly pattern occurs when two different groups are connected by one person.

that they inhabit very different social worlds. This network — the connections between my alters (or neighbors) — represents an ego-centered perspective on how email contacts are interrelated.

I ran the tool on 15 users' machines, generating and exploring their networks, and then asked the users to describe what sorts of connections and networks they observed. Who was connected, for example, and where did those connections come from? I observed commonalities in the networks that came out of these users' mail. The users explained the networks, telling stories about the connections — they identified the cluster representing their soccer team, for example, or the ties between marketers and designers at their companies. Figure 2 shows the four core patterns that I interpreted as structural patterns from certain recurrent roles that emerged from the project. We can view these images from Ego's perspective — Ego does not appear in the images, but each pattern shows how his neighbors are connected to him.

The *disconnected clusters* pattern occurs when groups of people aren't connected to each other at all. In Figure 2a, small social clusters are visible around the outside of the display, whereas a large central cluster shows a tighter set of interactions. Ego connects many disconnected groups that have nothing to do with each other. Although this image suffers from the crowding discussed in the sidebar, it's meant to illustrate only one feature: the clusters' disconnected nature.

In the *onion* pattern, a tightly connected group emerges at the core of a loosely connected group. Ego is a participant in the core group. This pattern was common in many circumstances, including those in which we found a small project team at the core of a larger set of interested parties — contractors, administrators, and so on. In Figure 2b, for example, we see the presenters at a conference surrounded by the larger network of attendees.

In the *nexus* pattern, a single person shares multiple contexts with Ego. In Figure 2c, we see a coworker who cooperated with Ego on various projects. The nexus was also common among managers and administrators — people repeatedly in touch with Ego even as Ego shifted among different contexts.

The *butterfly* approach occurs when two different groups of Ego's neighbors overlap via a single person. In Figure 2d, for instance, the one person was involved in Ego's research group and also in a workshop that Ego arranged.

These network patterns suggest that different contexts differ visibly as well. The patterns are detectable, recognizable, and indicative of the different sorts of online relationships that people share. In other words, the social contexts in which people are embedded are, in fact, ones that are worth recognizing.

But what can we do with these patterns, once we've identified them? We might be tempted to straightforwardly cluster the nodes from the network, breaking them into lists based on which people are most connected.[8] Each person might be placed in one list (such as "coworkers" or "friends") based on their nearest neighbors. Yet, the patterns' complexity argues against such a simplistic approach. The onion pattern, for example, demonstrates that both inner and outer circles of participation often exist; thus, any possible solution must be able to adjust its lists to different levels of intimacy. The butterfly pattern shows that we can't simply characterize a person into a single context at a time; rather, our relationships are multifaceted and might be tied to various concerns. The nexus pattern similarly suggests that even strong collaborators operate in a wide range of contexts. An email from a collaborator might be associated with one of many different projects; knowing something about the range of projects that the collaborator is involved in might help an individual sort incoming email into the correct category.

Interfaces for handling contacts and communication should be lightweight and supportive of dynamic groups. We can more easily build interfaces that support a notion of "relevant people" if they maintain that notion dynamically, based on social structure. We might build systems that focus on not just the person to whom a particular email is addressed but also the several sets of people to whom the recipient is most closely associated.

Roles in Conversation

I conducted the Roles project with Microsoft Research's Community Technologies Group in 2005, looking at archived data from 2001 and 2004. Microsoft Research has had a continuing interest in social behavior in online spaces.[2] One persistent question has involved identifying the good answers in question-and-answer oriented newsgroups. To do this, we must know something about who asks and who answers questions within these newsgroups. The Roles project looks at the social roles that emerge among Usenet newsgroup members and uses structural data around these groups to help us understand how contributors to online social spaces differ from each other.[9]

My own contribution was to bring a social network approach to the Usenet data. I based this research on a replied-to network that specifies which authors have replied to which others within a given newsgroup. I then examined these networks to see whether authors preferred to reply to some people rather than others, and whether different sorts of authors or different newsgroups exhibited characteristic patterns.

Usenet is an interesting testbed for this project. A 25-year-old online distributed discussion and posting space, it predates blogs, Web-based discussion boards, and even the Web itself. Although less prominent today than it once was, Usenet continues to have substantial traffic: hundreds of millions of messages are posted, read, and sorted into tens of thousands of topical newsgroups annually. One particularly interesting aspect is that all groups receive the same interface. Whether a newsgroup dedicated to technical questions and answers or one devoted to religious discussion, users have a similar experience — they can read messages in threads, post new messages in response, or start new threads.

This commonality allows for a broad comparative analysis.

Within this uniform space, different behavior and roles emerge. A politically oriented newsgroup will exhibit very different behavior than a social-support newsgroup, for example. The Roles project attempts to distinguish some of these roles using social-network tools.

For given newsgroups over a given month, I extracted social-network information from Microsoft's Netscan archive[10] by collecting all occasions in which one person replied to another. I then created an ego-centric network around participants in these newsgroups. For each of these hundreds of networks per newsgroup, a single node represented an individual author, whereas directed edges represented the number of times that one author replied to another. Thus, an edge from *A* to *B* with weight 3 indicates that *A* replied to a message from *B* three times. The network was restricted to two degrees out from the center: we saw the participant, the people who the participant interacted with, and the people who they, in turn, were connected to.

My research group extracted networks representing a month's worth of replies from nine different newsgroups on five different topics. (We sampled eight of the nine groups through January 2001, and `microsoft.public.windows.server.general` during November 2004). Table 1 lists the newsgroups we sampled and their general topics. To demonstrate how large the newsgroups were, it also lists the number of replies (that is, messages involved in conversation) and distinct authors during that time period.

Table 1. Newsgroups sampled during the Roles study.

Topic	Name	Replies	Authors
Technical support	comp.soft-sys.matlab	712	437
	microsoft.public.windows.server.general	1,489	855
Discussion	rec.kites	924	276
	rec.music.phish	6,105	1,263
Political discussion	alt.politics	16,059	2,187
Flame	alt.flame	3,802	618
	alt.alien-vampire.flonk.flonk.flonk	6,494	755
Social support	alt.support.divorce	2,937	339
	alt.support.autism	3,095	188

We wanted to identify "answer people" — those who are instrumental in answering technical questions posed in technical newsgroups. Although few newsgroups overall are technical, such groups are particularly interesting for understanding the community that grows around software, and for observing a highly active user population. Answer people are rare, but distinctive, and we can recognize them using local social-network data. To do so, we examine both an individual's *out-degree* — the number of different people to whom they've responded — and those of their immediate neighbors. Answer people tend to respond to those who don't respond to many others — that is, they have a high out-degree, but are largely connected to those with low out-degrees.

We can summarize this behavior with *out-degree histograms*, which show the distribution of the degrees for any given person's neighbors. This measure lets us infer how a person interacts with others. For example, a heavily left-weighted histogram is common to people offering technical support. Figure 3a (next page) shows an answer person from the group `microsoft.public.windows.server.general`. The answer person, at center, replies to several people who are themselves largely disconnected from anyone else — that is, those who are just starting conversations or who have responded to only a small group of people. In this example, the answer person also replies to one person with a high out-degree, illustrating that answer people occasionally clarify answers for each other or disagree with another's answers.

In contrast, a more heavily right-weighted histogram occurs with people who carry on prolonged and frequent discussions, such as members of a political newsgroup. Figure 3b shows a network diagram and a histogram for the group `alt.politics`, which is based on brisk discussion of current political issues. Although this stripped-down network diagram is crowded, we can see several important differences between it and the technical support group's diagram. Most obviously, the central core is tightly interconnected. Unlike in the technical support group, tightly connected people reply to other connected people.

Figure 3c shows a characteristic member of one flame group we observed. Flaming involves vocal, often insulting, conversations within groups.[11] This diagram looks somewhat similar to the political discussion group's, with one exception — conversation in flame groups has a very strong reciprocal aspect. Whereas we traditionally see reciprocity as a valuable aspect of conversation, in online groups, it frequently indicates flame wars. Our data shows flame wars to be concentrated around pairs of people — although many people might participate at the beginning, most of the conversation

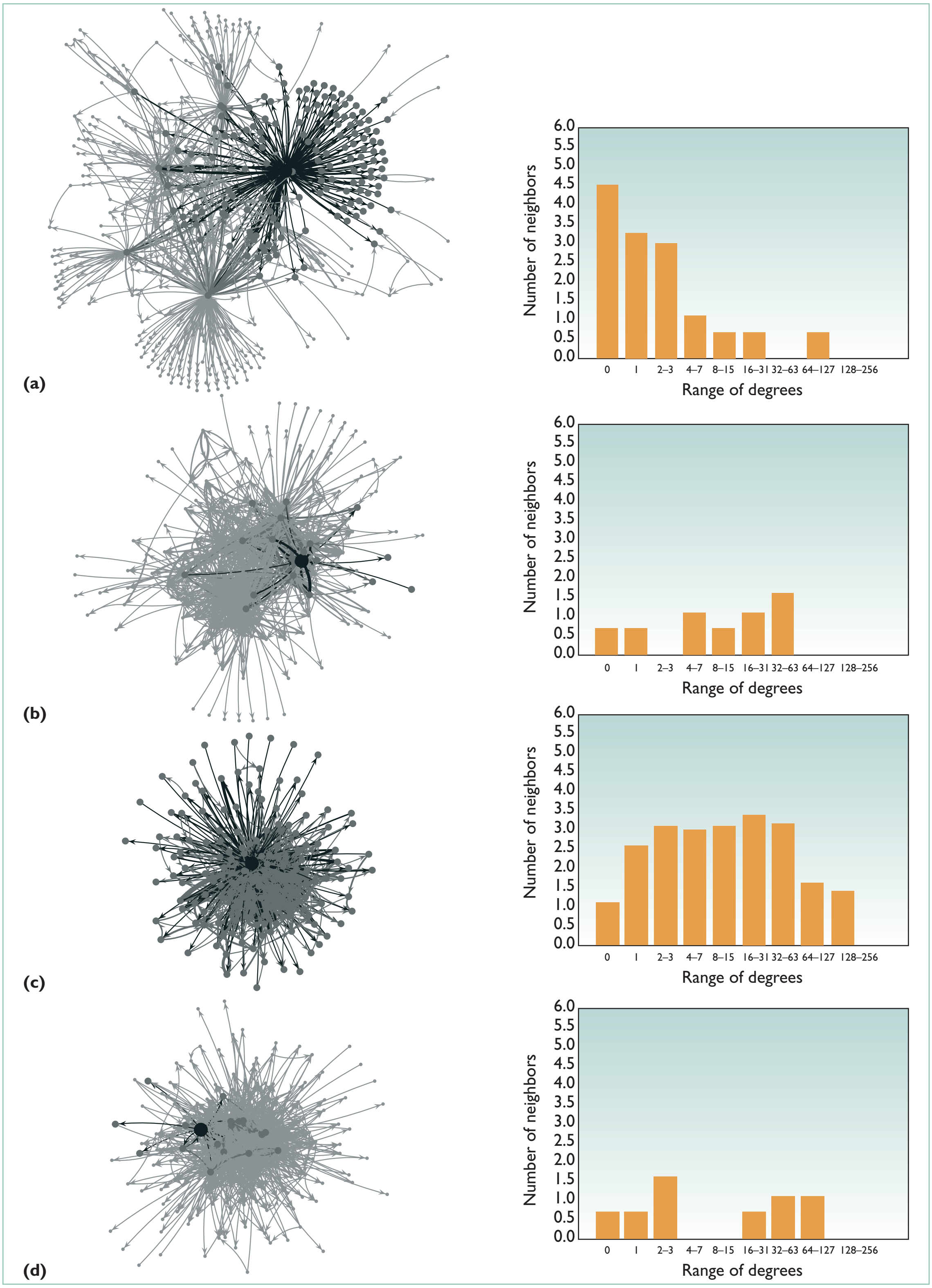

Figure 3. Typical roles in four different newsgroups. We examined groups for (a) technical support, (b) political discussion, (c) flaming, and (d) social support. The left side shows a given contributor's immediate neighbor networks; the right side shows the out-degree distributions for each group. Out-degree distributions are on both a logarithmic x-axis, so that each bucket is twice as large as the previous one, and a y-axis, so that a column one unit higher represents roughly twice as many neighbors.

occurs between two discussants by the end. Indeed, the figure shows several thick lines around the core member, indicating his flame-war history.

Finally, a substantially different characteristic pattern exists in groups dedicated to social support. Figure 3d shows the typical out-degree distribution within `alt.support.divorce` — the most active people have a bump at both the low and high ends. This suggests that they both respond to the requests of new participants looking for support and continue to talk with the most active members.

Interestingly, in none of these groups can we explain the distribution merely using the participants' raw in- or out-degrees. That is, their activity levels don't correlate with their roles. All of the groups share the common power-law-like distribution of in- and out-degrees: a large number of alters to whom very few messages were sent, with a long tail of few alters to whom far more messages were sent.

Although our work is still in progress, it leads us in some intriguing directions. Distinct behaviors emerge in different newsgroups: `alt.flame` has no answer people, for example, whereas `alt.politics` appears to have no social-support people. On one level, this is unsurprising — the groups have, after all, different goals and needs. However, it's extremely surprising when we consider how few formal constraints exist to regulate how actors behave in these groups (posters can't be kicked off or fined for misbehavior, for example). Although all the newsgroups' interfaces look the same, radically different roles emerge in the spaces.

The perspectives of both the Soylent and Roles projects are intentionally limited. In many cases, examining a whole group or full network can lead to interesting and useful results. However, each of those approaches runs into issues of both complexity and privacy. Our combination of building on public data and maintaining an immediate perspective seems like a natural and valuable way to study online communication.

Given that the Solvent project studies only outgoing email, it fails to include other resources through which people collaborate, including documents, presentations, other email accounts, and face-to-face contact. To broaden this project's scope, we could annotate all documents within a computer system with social data, such as where (and whom) it came from and who has edited it, sent it, or received it. Then we could broaden the patterns I've just discussed to include artifacts as well as people.

In the Roles project, we might ultimately be able to automatically detect and classify the distinct behaviors that emerge among users. Users could thus discover whether the person they're reading tends to take the role of an authority or a flamer. An enhanced search engine could provide contextual information — are the results from someone who provides social support or discusses issues? Different audiences might find this information valuable in different contexts. The techniques just suggested — using social network data — could expand the ways in which we use social metadata to understand and reflect group behavior.

It would also be valuable to track the stability of both groups and the roles played by individuals. Several questions emerge once we think of a newsgroup as inhabited (and, in part, defined) by a set of roles within it. How rapidly do individuals within a single group change? Are their roles consistent between different newsgroups (suggesting that people search for newsgroups that match their strengths), or do they adapt to the environments in which they find themselves? How firmly do newsgroups hold on to these roles, and under what circumstances do they change? Our work at Microsoft continues to develop and explore these questions further.

Acknowledgments

I thank Paul Dourish, my graduate advisor and collaborator on the Soylent work, and my collaborators on the Roles project, Marc Smith, Ted Welser, and Tammara Turner, as well as Paul Johns, Eric Gleave and AJ Brush. IBM Research, Intel Research, and the US National Science Foundation partially supported the Soylent project. I also thank the anonymous reviewers for useful insights and clarification.

References

1. D. Fisher and P. Dourish, "Social and Temporal Structures in Everyday Collaboration," *Proc. Conf. Human Factors in Computing Systems* (CHI 2004), ACM Press, 2004, pp. 551–558.
2. T. Turner et al., "Picturing Usenet: Mapping Computer-Mediated Collective Action," *J. Computer-Mediated Comm.*, vol. 10, no. 4, art. 7, 2005; http://jcmc.indiana.edu/vol10/issue4/turner.html.
3. R.S. Burt, *Structural Holes: The Social Structure of Competition,* Harvard Univ. Press, 1995.
4. J. Tyler et al., "Email as Spectroscopy: Automated Discov-

Network Visualizations

Perhaps the most recognizable feature of many social-network studies are the ubiquitous network-wide visualizations, or *sociograms*.[1] In general, they try to represent how "far apart" nodes are by placing them at distances proportionate to their connectivity with minimal distortion. These visualizations can be confusing and crowded when the network has many nodes or interconnections; any distortion can make the diagram deceptive.

Despite these concerns, the work I present has leaned heavily on network diagrams as an interpretive tool for understanding social networks. With the Soylent and Roles projects, I try to simplify the graphs to reasonable sizes by focusing the graph only on single individuals and the persons immediately around them. These graphs are far simpler, and less prone to distortion, than graphs that try to provide overviews of broader spaces.

Additionally, network diagrams aren't end-user tools in my projects. Because they present information nonlinearly, readers can find information only by searching entire graphics rather than scanning simple lists. Instead, I use the graphs as intermediate, interpretive visualizations. Knowing the networks' general shapes and properties — the commonalities and differences that occur between different networks — lets developers design interfaces that reflect this underlying data. I generated all visualizations with the Java Universal Network/Graph (JUNG) toolkit.[2]

References

1. L.C. Freeman, "Visualizing Social Networks," *J. Social Structure*, vol. 1, no. 1, 2000; www.cmu.edu/joss/.
2. J. O'Madadhain et al., "Analysis and Visualization of Network Data Using JUNG," to be published in *J. Statistical Software*, 2005.

ery of Community Structure within Organizations," *Proc. Communities and Technologies*, Kluwer, 2003, pp. 81–96; http://portal.acm.org/citation.cfm?id=966268.

5. F. Viegas et al., "Digital Artifacts for Remembering and Storytelling: Posthistory and Social Network Fragments," *Proc. 37th Hawaii Int'l Conf. System Sciences*, IEEE Press, 2004, p. 10,409a.
6. B. Nardi et al., "ContactMap: Integrating Communication and Information Through Visualizing Personal Social Networks," *Comm. ACM*, vol. 45, no. 4, 2002, pp. 89–95.
7. S. Farnham et al., "Personal Map: Automatically Modeling the User's Online Social Network," *Proc. Interact 2003*, IOS Press, 2003; www.interact2003.org/docs/INTERACT03flyer1.pdf.
8. M. Girvan and M. Newman, "Community Structure in Social and Biological Networks, *Proc. Nat'l Academy of Science USA*, vol. 98, US Nat'l Academy of Science, 2002, pp. 7821–7826.
9. M. Smith et al., "Social Network Visualizations Of Usenet Newsgroups," *Proc. Int'l Network Social Network Analysts (INSNA) Conf. Social Networks* (Sunbelt 2005), to be published in 2005.
10. M. Smith, "Tools for Navigating Large Social Cyberspaces," *Comm. ACM*, vol. 45, no. 4, 2002, pp. 51–55.
11. L. Sproull and S. Kiesler, *Connections: New Ways of Working in the Networked Organization*, MIT Press, 1991.

Danyel Fisher is a researcher in Microsoft Research's Community and Technologies group. His work centers on applying social network analysis and information visualization to online social environments. Fisher has an MS from the University of California, Berkeley, and a PhD from the University of California, Irvine, both in computer science. He is a member of the ACM. Contact him at danyelf@acm.org.