working paper for:

Stanford **Online Deliberation Conference, 2005**

Debate, Division, and Diversity: Political Discourse Networks in USENET Newsgroups

John Kelly, Danyel Fisher, Marc Smith

Research for this paper was supported by a grant from the
National Science Foundation, #IIS- 0441999.

Do online political discussions tend to aggregate diverse voices in cross-cutting debate and deliberation? Or do "audiences" for online discussion tend to fragment into ideological echo chambers? In the wilds of threaded discussion on the internet (as opposed to deliberative polls, moderated discussions, and other designed venues of deliberation), networks of political discourse emerge from billions of individual choices by millions of individual citizens about what to discuss online, where to discuss it, and with whom. Do these choices lead individuals to interact across ideological divides, or to cluster within them? How we understand the value of internet-based forms of political discourse is in part dependent on this question.

The danger to a healthy public sphere, well articulated by Sunstein, is that individual-level selectivity biases, enabled by the internet's facilitation of choice in information consumption ("consumer sovereignty" in Sunstein's words), will precipitate a balkanization of political discourse (Sunstein 2001). In this view, individual preferences for reinforcement in political information and dialog are held in check by public institutions—information "commons," like public parks or the mainstream mass media—in which citizens are exposed to a range of viewpoints they would not otherwise encounter. The internet represents not such much another public commons, but rather a kind of anti-commons which allows citizens to consume information and affiliate with others on the basis of shared values and interests only. The danger to a unified, balanced public sphere is compounded by the tendency of groups even slightly biased to one side of an issue to move further from the center in the direction of the bias. The implication is

that internet modes of political discourse are likely to promote the dissolution of the public sphere in to so many self-radicalizing "public sphericules," to use Todd Gitlin's phrase (Gitlin 1998).

Sunstein's dystopian fears are not without empirical merit. Network analyses of Amazon.com "also-bought" lists show an American public neatly bifurcated into conservative and liberal audiences for politically oriented books (Krebs 2004). Link analysis of the "blogosphere" likewise demonstrates the coalescence of blogging networks into distinct liberal and conservative clusters (Adamic 2005). The tendency of individuals to consume (and in the case of blogging, produce) media resonant with their pre-existing political perspectives is mirrored in the dynamics of political discussion within social networks. Face-to-face, individuals tend to discuss politics when they feel they essentially agree with their interlocutors (Mutz 2002), or when their own view seems dominant within the group, and avoid discussing politics when they believe others to hold opposing views (Noelle-Neumann 1984). In addition to succumbing to pressures for conformity within existing social networks, individuals tend to form new social networks with similar others (McPherson 2001). Within the sphere of personal relations, it is likely that most citizens will rarely encounter views opposed to their own, and even more rarely will they be required to defend those views in active debate. Indeed, as Sunstein proposes, the mass media appear to be the principal source of cross-cutting political views available to citizens (Mutz 2001).

Though the tendency toward homophily in social networks is well established, we should not assume it necessarily operates in online discussion environments. Some online discussants seek reinforcement, but others go online to encounter differing points of view. Individual motivations vary, and therefore so do individual behaviors, and ultimately the structures of discussion networks that emerge from them. The ability to discuss politics online without risking personal relationships may in fact liberate many who would otherwise avoid active participation in political discourse. The participation of even a small number of oppositional debaters in a discussion environment could, provided they are not ignored, put a brake on the formation of radicalizing "enclave deliberations." And if debate is the modal activity in a discussion environment, it is conceivable that online fora are enabling exactly the kind of public commons Sunstein fears the internet to be endangering.

The goal of this analysis is to assess the amount and implications of debate in several explicitly political USENET newsgroups, by measuring interactions across and within ideological divides, identifying patterns of interaction, and illuminating characteristics of the networks that emerge from threaded discussion.

Interestingly, much of the empirical work to-date on online political discussions has looked into online venues and found vigorous debate, rather than ideological back-patting and conspiring extremists. Indeed, the main criticism of online fora (derived from coding discussion corpora according to various operationalizations of Habermasian ideals of rational-critical discourse) is that they are in a way *too* oppositional, not that they are

insufficiently so. In their vulnerability to name-calling, flame-warring, hate-talking, ideological grandstanding, and sundry other modes of incivility, online political discussions have been judged to violate just about every key value philosophers of discourse have identified: quality, equality, reciprocity, diversity, reflexivity, etc. (Dahlberg 2001; Graham 2003) However, more in the spirit of Mill than of Habermas, a small number researchers have defended online discussions as possessing the virtue of collecting dissonant voices around contested topics (Broncheck 1997), even going so far as to defend some "incivility" as evidence of healthy political contestation (Papacharissi 2004).

Though not a critique of the exercise of applying Habermasian standards to the kind of unfettered discussion found on the internet, the present study is more in sympathy with the latter sensibility. From Mill's perspective, vigorous debate, even between irreconcilable perspectives, can be a very good thing. It is probably of limited value to judge online political fora as if they were true deliberations, because, at least according to Gutmann and Thompson's definition, they are not (Gutmann 2004). No decision will be taken by the discussants at the conclusion of debate. In fact there is no conclusion of debate. Newsgroups are ongoing arenas of ideological contestation, where facts, values, philosophies, current events and historical interpretations are fought over in a never-ending tussle among authors whose contributions range from prolific to rare—played out in front of an audience of fellow authors and lurkers alike. In this environment, "quality" is as much a matter of diversity of voices, and the density of interactions among them, as of overt indicators of deliberativeness.

Content and discourse analysis, sufficient when applying interpretive methods and formal "discourse quality" schema (Steenbergen 2003), is only one step in analyzing the structure and dynamics of discourse networks. Previous analyses have struggled valiantly to investigate issues perhaps beyond tractability given available tools. The results, sampling messages (or in some case threads) and testing hypotheses using these as the unit of analysis, have forced odd conceptualizations (e.g. assigning an ideological perspective to a thread, as opposed to an author or a post; and using adherence of discussion to topic raised in the seed post as a measure of the value of the discussion and as evidence of the seed author's "leadership" (Hill 1997)), and come to conclusions quite opposite to our own findings (e.g. that USENET political newsgroups lack diversity of perspective and suffer from the hegemony of dominant voices (Wilhelm 1999)). Messages are products of authors, who are often engaged in extended chains of dialog, some posts being long and involved, many others quick and cursory. Random sampling of messages is likely to bias consideration of the latter at the expense of the former. Threads, similarly, are products of authors' interactions. The most powerful way to assess the presence and interaction of points of view in an online forum is to study authors, who generally have stable points of view, and the interactions among them. Fortunately, new tools are being created which provide the capability to do that (Sack 2004; Smith 2005).

We use Microsoft Research's Netscan tool to view interactions among authors within eight USENET political newgroups. Netscan captures the header information for all

posts within USENET newsgroups, and can extract a wide range of meta-data on posting behavior. In the threaded structure of USENET, all posts are either "seed posts" (which try to start a new thread) or "replies," which specifically answer another post (and default to quoting the text of that post in the body of the new message). Thus, unless attempting to start a new conversation, every post in USENET is at least nominally (and usually substantively) answering, or commenting upon, a specific preceding post. Authors of posts therefore typically behave as though addressing the poster they are replying to, or less frequently, as though delivering an aside to the presumed audience also reading the same sequence. Netscan can track statistics for who is replying to whom. These patterns of response, or when reciprocal, conversation, form a discourse network—the network of who, over time, talks to whom. Our analysis relies on measuring this type of social structure and making it visible.

Previous work, using Netscan to study a wide range of newsgroups, has found that network structure varies with the type of group, i.e. the purpose of participation (Smith 2005). Consider a technical newsgroup, such a one for programming software. It is likely that a large number of people will show up there with specific questions. There will likely be a very few authors, in fact usually one or two, who have a great deal of technical expertise and can provide specific answers. This results in a kind of "star" structure—with myriad posters unconnected directly to each other, but all connected to the central "answer people" who give them solutions to their problems. Another pattern would be found in a fan newsgroup, where people gather to trade information and chat about their favorite movie star, musician, etc. Here also, and as in all newsgroups, we

would find variations in levels of participation (usually following a power curve distribution), but most authors would be answering and replying to multiple others, yielding a denser network of interaction than that found in a technical group.

We might expect political newsgroups to have their own characteristic patterns of organization. A group focused on debate, assuming that authors belong to meaningful ideological (like "liberal" or "conservative") or issue-position (like "pro-choice" vs. "pro-life") clusters, would feature thick conversational links across the political divide(s) and fewer links within clusters of the like-minded. The kind of group Sunstein fears, in which the like-minded reinforce one another and move the spectrum of discourse in a more radical direction, would feature stronger linkages within clusters than between them—if indeed different clusters of opinion were present. It is also possible that authors from only one side of an issue, or one ideological position, are present in the group. This intuition provides the research agenda for this study. We choose a selection of newsgroups, code the ideological or issue positions of the key authors, and examine the patterns of replies among them. We are looking for: a.) meaningful opinion clusters, i.e. is there opinion heterogeneity; and b.) differences between the density of connections across vs. within clusters.

This approach is similar in conception to one taken by Agrawal, *et al.*(Agrawal 2003), in which quotation links (we use replies, essentially the same thing) among posters in several newsgroups discussions were used to build a social network graph. On the assumption that "antagonism" (i.e. disagreement) was the principal motivation for

quoting another author, these graphs were used to predict the issue position of unknown authors based on the position of several known ones (who were coded by content analysis). Using selected messages from three newsgroups (talk.abortion, talk.politics.guns, and alt.politics.immigration) they test various algorithms based on this graph partitioning approach and achieve predictive accuracy significantly superior to chance. Though it relies on "antagonism" (i.e. debate), the method has some tolerance for "reinforcement." Accuracy, however, relies on as many authors as possible having more antagonistic links than reinforcement links. Our approach, which measures these links, forms a nice complement to theirs.

Methodology

This study uses Netscan to look at interactions among the most active posters in eight USENET political newsgroups during November, 2003. A corpus was assembled from samples of threads in which selected posters were most active, and qualitative content analysis was used to code these posters' basic political or issue positions. Netscan was used to track posting metrics for all coded posters, including the numbers of outgoing and incoming replies to and from posters in the various political categories. Advanced Netscan features were used to create network maps, including ego networks for coded posters and group-wide maps of interactions among posters. The resulting data on interactions among discussants forms the basis of our analysis.

Basic newsgroup statistics:

type	newsgroups	Posts	Replies	XPosts	Posters	Repliers	coded	codemsg	%coded
ideology	1 alt.politics.**bush**	56216	48923	51222	3873	3384	40	9448	19.3%
	2 alt.politics.**democrats**.d	20314	18256	19951	1590	1469	40	4874	26.7%
	3 alt.fan.noam-**chomsky**	2777	2523	2633	463	425	29	1107	43.9%
issue	4 talk.politics.**mideast**	10678	9222	10370	1101	992	25	2521	27.3%
	5 talk.**abortion**	9825	9603	5758	340	301	26	5186	54.0%
	6 alt.politics.**immigration**	3504	2984	1833	443	394	21	1209	40.5%
small	7 soc.politics.**marxism**	223	189	0	45	30	27	156	82.5%
	8 **hsv**.politics	218	174	1	32	25	21	150	86.2%

posts: how many posts during November, 2003
replies: how many posts were replies, rather than seed posts
Xposts: how many posts were also cross-posted to other newsgroups
posters: how many unique authors
repliers: how many posters issued replies
coded: how many authors were coded in content analysis
codemsg: how many messages were sent by coded authors
%coded: what percentage of all messages were sent by coded authors

Newsgroup selection:

From hundreds of USENET newsgroups with explicitly political orientations, we selected eight for analysis. In terms of content and focus, political newsgroups come in two main varieties, *issue* oriented (e.g., abortion, immigration, gun control) and what we will call *ideology* oriented. In *ideology* newsgroups a range of topics are discussed, and authors usually address any number of these from their respective ideological positions. *Ideology* groups can include such nominally "fan" groups as alt.fan.rush-limbaugh and alt.fan.noam-chomsky, in addition to groups with explicit ideological or party-affiliations tags, such as libertarian, republican, democrat. *Issue* groups feature discussion principally about one issue of importance to the public agenda.

Theoretically, it is important to look at both types of groups because we might expect differences to exist based on the kinds of "audiences" they collect, which may differ psychologically in their commitments to the discussion. *Ideology* group authors may be expected to represent members of the "attentive public," citizens who follow a range of issues on the public agenda, pay attention to the news and to political leaders, and bring coherent political philosophies to bear on different problems. Authors in *issue* groups, on the other hand, may be drawn from "issue publics," and have strong commitments to a single issue about which they have very strong opinions, often based on personal experiences, interests, or issues of identity. It is conceivable that these two types of author will behave very differently.

Newsgroups also come in a wide range of sizes, from a few very large ones that are active for many years, down to a great many that have only a few messages before they die out. It's also important to note that, since most threads are posted to more than one newsgroup, newsgroup boundaries vary in their permeability. We chose groups to represent positions along the range of these dimensions. We have three *ideology* groups, and three *issue* groups, of various sizes and levels of cross-posting, and two additional newsgroups chosen because they are very small and have no cross-posts.

Ideology: **alt.politics.bush, alt.politics.democrats.d**, and **alt.fan.noam-chomsky**
Issue: **talk.politics.mideat, talk.abortion**, and **alt.politics.immigration**
Small: **soc.culture.marxism** (almost a kind of hybrid ideology/issue group) and **hsv.politics** (which is like an ideology group, but for Huntsville, Alabama. Note: there

are no formal mechanisms for limiting participation to Huntsville residents, but reading of posts gives some evidence for thinking participants are local.).

Time period: We chose to analyze all activity for the month of November, 2003, a year before the U.S. Presidential elections. During this time, candidates for the Democratic nomination were campaigning, and President Bush was staying in the public eye as well. The public was being activated by a flood of political communication from both parties, and yet the election was not so close that campaign agendas were fixed and loyalties to particular candidates crystallized.

Author selection: Top posters in each newsgroup were selected on the basis of number of days active, rather than alternative measures such as number of messages posted during the time frame. This has the virtue of favoring those who participate most regularly over those who show up infrequently but post large numbers of messages. Because of the temporal, threaded nature of USENET discussions, consistency of presence is more critical than sheer numbers of messages to the control of ongoing discourse. Top posters were selected (as opposed to a random sample of posters) because they form the core of the network of interactions, accounting for a disproportionate number of the messages and guiding direction of threads though consistent participation.

For the two largest groups, **bush** and **democrats.d**, the top 40 authors were coded. For **hsv** and **marxism**, all authors who contributed two or more messages were coded (21 for **hsv**, 27 for **marxism**). For the other groups in the middle range of size, the approach was

to code the top 20 authors, adding those over 20 who were active the same number of days as the 20th author (**mideast,** 25; **chomsky,** 29; **immigration**, 21; and **abortion,** 26). The percentage of all messages in the newsgroups generated by the coded authors is shown below (fig. 1).

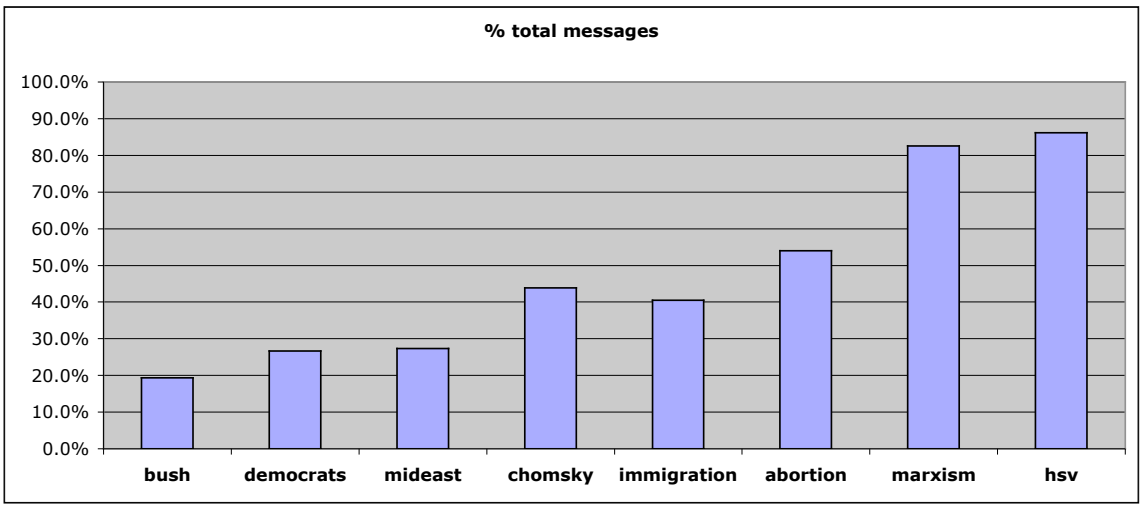


fig. 1

Message sampling: Sufficient messages were sampled to give a good sense of the authors' positions as taken in a number of conversations. To begin, a 25-message block (sequential, chosen to be dense with the author's contributions) from each of the top five threads for each author were extracted (using Google Groups) and assembled into a corpus. Since these blocks contained messages from other authors in the sample, redundancies were avoided by selecting different blocks on the same thread for each author, and sometimes (once a thread was completely extracted) additional threads were chosen. During content analysis, when the number or quality of selected messages for a given author was insufficient to determine their ideological leanings or issue position, additional threads were added to the corpus. The final corpus contains over 16,000 messages.

Content analysis/author coding: Using NVivo, messages were read for each selected author and passages indicating ideological perspective and/or issue position were coded. For this study, *ideology* groups were coded into simplified *right* (code 1) and *left* (code 2) categories. While most authors spoke from perspectives within what could be considered the American political mainstream, that is identifiably "liberal" or "conservative," there were also other perspectives that were possible to code either *left* (including progressive, socialist, communist, and anarchist perspectives) or *right* (various strains of libertarian). A third code was reserved for "other," and included perspectives that did not fit the conventional left/right schema (e.g. racist nationalists who admire Karl Marx).

Issue group authors were coded according to their issue positions, as determined by stated policy preferences and associated rhetoric. *Issue* group authors were also coded for *left*/*right* ideological categories, but under the condition that issue-position alone was not sufficient evidence of an ideological commitment, this was surprisingly hard to do. There are, for instance, liberal democrats who take a "pro-life" stance on abortion, and committed republicans who are firmly "pro-choice." The difficulty of coding ideology in *issue* groups is good evidence for an "issue public" conception of authors attracted to issue newsgroups. These authors are generally passionate about one issue, this view is not explicitly coordinated with a higher-level political philosophy, and it is difficult to find them discussing anything else online.

Authors whose posts gave insufficient evidence for determining their political beliefs or policy preferences were coded "0." Despite the caveats required for this kind of

interpretive process, particularly regarding the artificiality of the conventional *left-right* dichotomy, most authors were very easy to code according to the selected schema. The exception was **immigration,** in which a revised coding scheme was suggested following early data analysis and a re-reading of the discussions.

code chart

	code 0	code 1	code 2	code 3
bush	unknown	right	left	other
democrats.d	unknown	right	left	other
chomsky	unknown	right	left	other
mideast	unknown	pro-Israel	anti-Israel	other
abortion	unknown	pro-choice	pro-life	other
immigration	unknown	immig. is ok	anti: econ	anti: racist
hsv	unknown	right	left	other
marxism	unknown	capitalist	marxist	other

issue code notes:

- **mideast** is almost entirely about the Israeli/Palestinian conflict. Code 1 authors support Israel without question. Code 2 authors oppose Israel completely. Code 3 (other) authors take positions in the middle, e.g. supporting Israel's right to existence and peace, but criticizing policies, or pushing for understanding between groups.
- **abortion** features only pro-choice and pro-life authors, nobody was coded "other."
- **immigration** is almost entirely about illegal immigrants from Mexico, and features only two authors defending the *status quo*. Those opposed were recoded into authors focused on economic and legal implications of illegals, and those who dislike Mexican immigrants for racial and cultural reasons. The latter often used very offensive terminology. The single "other" author was a middle/upper class Mexican citizen (in Mexico) who understood the desire to keep illegal immigrants out of the US, but spent a lot of time countering cultural, historical and economic arguments of code 2 and 3 authors.

Netscan analysis:

The Netscan system collects information about who wrote each message, and assembles the threads of messages. In this way, it has data on which authors replied to other authors.

This data can be interpreted as a directed, weighted social network, with authors of posts as nodes, and with edges representing "has replied to". Thus, an edge from A to B with weight 3 indicates that A has posted three messages that are direct replies to B three times.

We collected this data for each of the groups. We then used a custom social-network visualization tool based on the JUNG toolkit (REF) to both visualize this data and to collect network-based statistics on it. Per-author network statistics included the number of replies and number of messages within- and without-group for each coded author. We also generated visualizations of reply structure within the networks for coded authors.

Findings

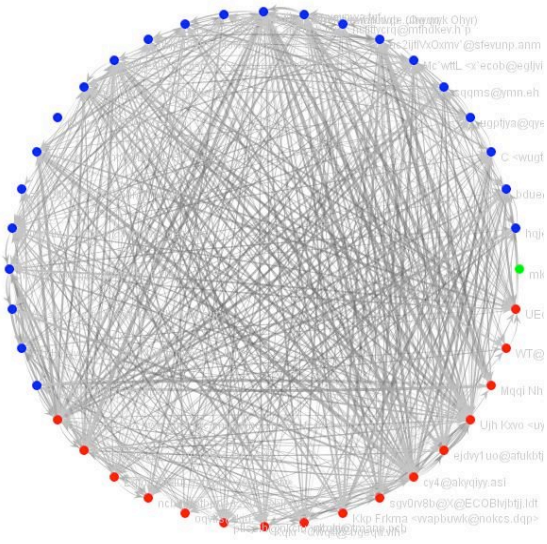


fig. 1

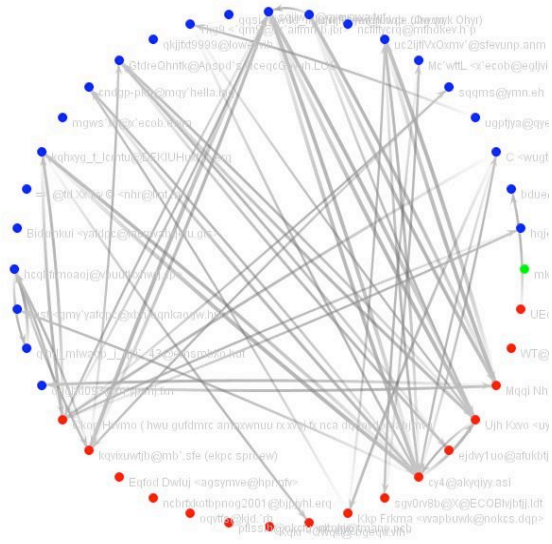


fig.2

Above we see two network visualizations of reply patterns among the top 40 authors (by days present) *alt.politics.bush*, the largest political newsgroup in November, 2003. *Left* authors are blue, *right* authors are red, green is *other* (in this case a paranoid author who thinks the Freemasons control the world). The circular layout is designed to make the relative strengths of network connections (based on direct replies) visible, but in the diagram on the left, in which every line represents a connection between two authors based on a single reply, the density of connections is so thick it is hard to immediately ascertain anything. But, looking closer at the left diagram [fig. 1], a few conclusions are apparent: a.) over the course of a month everybody spoke with (or at least to) a large portion of the rest of the group; b.) crosstalk between ideological clusters is very strong, at least on a par with replies within clusters; c.) there are more *left* authors in **bush** than *right* ones, but not overwhelmingly so; and d.) at two degrees, i.e. counting all the people

"who talk to the people I talk to," everybody is linked together. This is a tight discussion core.

Looking at the diagram on the right [fig. 2], which raises the threshold for drawing a line to nine replies linking authors, sharpens conclusion **b** dramatically. Of all the linked pairs in the network, only three are between *left* authors, and one is between *right* authors. Looking at the many links across the ideological divide, it is also clear that these authors tend to converse with multiple authors on the other side, not just one or two. And to iterate, rarely do they converse at such length with authors on their own "side."

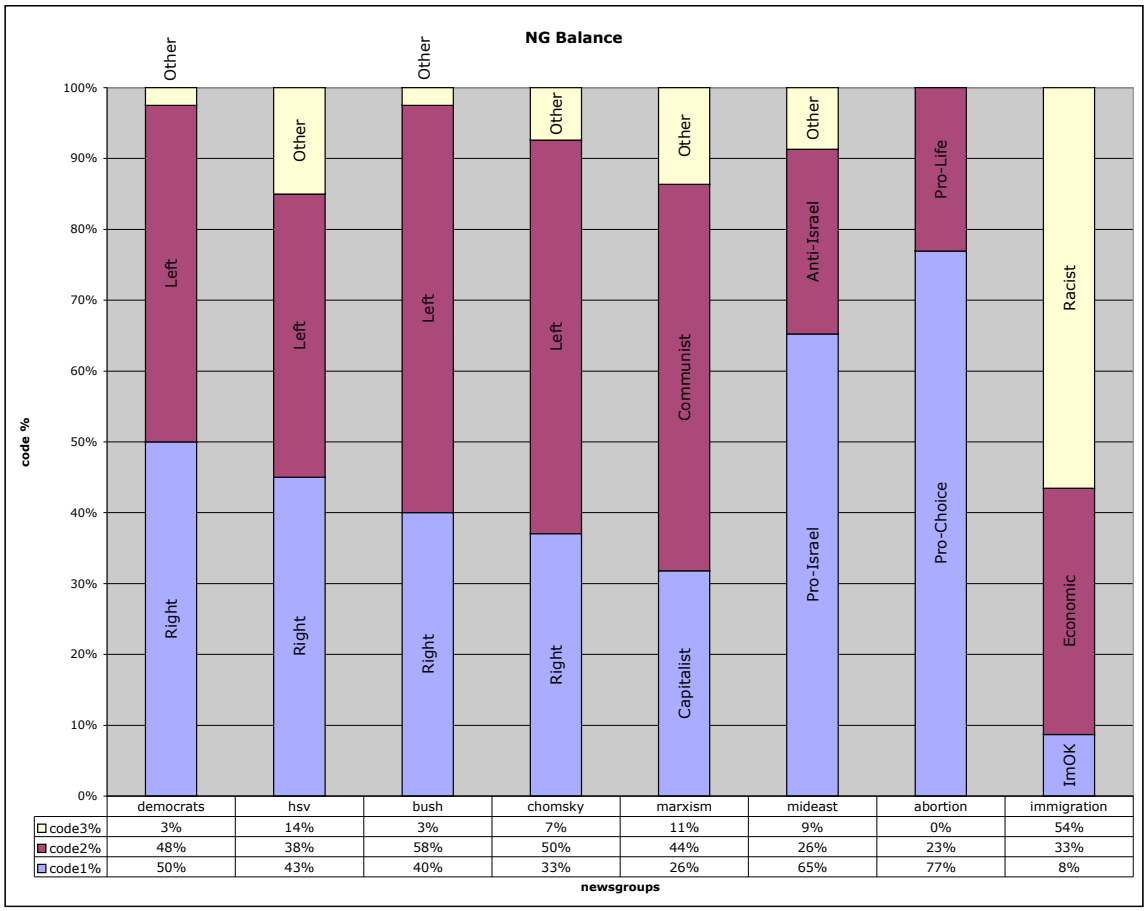
These simple patterns, intuitive if one assumes that going online to talk politics (at least in USENET) means going online to *argue* politics, generally hold for the other newsgroups as well, although with some interesting exceptions and variations that we will discuss. In depth analysis will follow, but to deliver the punchlines in advance:

- Discussion in political newsgroups is overwhelmingly across clusters of the like-minded, not within them. This is, with the partial exception of **immigration**, as true of the unbalanced *issue* groups as the relatively balanced *ideology* groups.
- There is a modicum of talk within clusters, particularly within the majority cluster of the more unbalanced groups.
- With some exceptions (principally **immigration**), groups are dominated by two main clusters. This is obviously related to the assumptions of the qualitative

analysis by which categories were assigned, but the quantitative results bolster the validity of the clusters.

- *Ideology* groups are generally balanced in the population of top posters belonging to the two dominant clusters, *issue* groups on the other hand are highly unbalanced, with the preponderance of authors advocating one side of the issue.
- Authors from the minority clusters often post more replies to majority cluster authors than vice-versa, so that message traffic between clusters is generally more balanced than the population of authors. In the case of **bush**, the message output of the minority cluster exceeds that of the majority cluster. In the case of **democrat,** the message output of the *right* magnifies its slight advantage in number of posters.
- Patterns that are obvious in the six larger *ideology* and *issue* groups are present, but harder to discern in the two small groups, mostly due to the smaller number of authors and replies to measure. Therefore, some these groups will be discussed only when there are relevant and clear observations to make.

The Balance of Perspectives: *Ideology* vs. *Issue* Groups:



	democrats	hsv	bush	chomsky	marxism	mideast	abortion	immigration
code3%	3%	14%	3%	7%	11%	9%	0%	54%
code2%	48%	38%	58%	50%	44%	26%	23%	33%
code1%	50%	43%	40%	33%	26%	65%	77%	8%

fig. 2

In the above diagram (fig. 2), newsgroups are sorted horizontally, in descending order, according to a calculation of "balance" based on the ratio of authors from the two dominant clusters. Immediately striking is that the *ideology* groups, those with dominant clusters definable as "right" and "left," comprise the four most balanced, and the *issue* groups are the most unbalanced. There is, however, an important distinction between the balance in the population (how many authors are in which category), and the balance in the message traffic they generate. The differences between these two measures of balance are in some cases striking, and in others small. But the differences occur in two different patterns:

***ideology* groups**

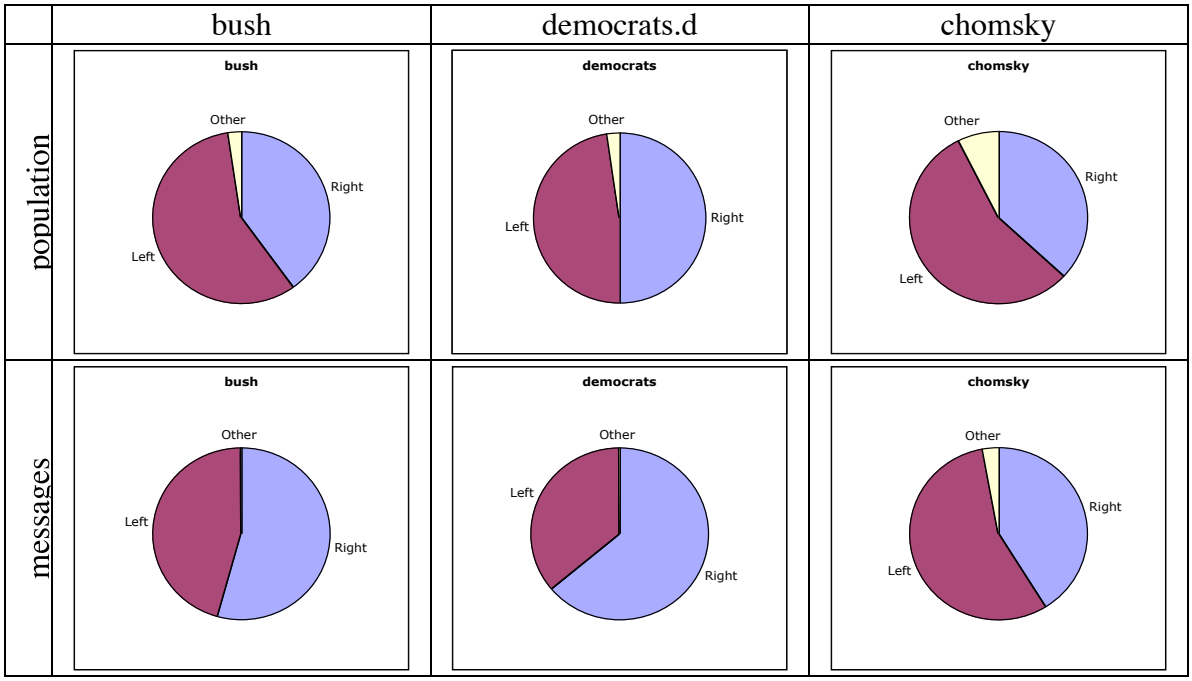


Fig. 3

- In *ideology* groups, the *right* "speaks more" than the *left*, meaning that the *right* generates more messages per author that the *left*. This has the effect of making **democrats.d** more unbalanced in favor of the *right*, making **chomsky** more balanced by lessening the slight majority the *left* enjoys in population, and actually reversing the dominance in **bush** from *left* to *right*. (see **fig. 3**)

***issue* groups**

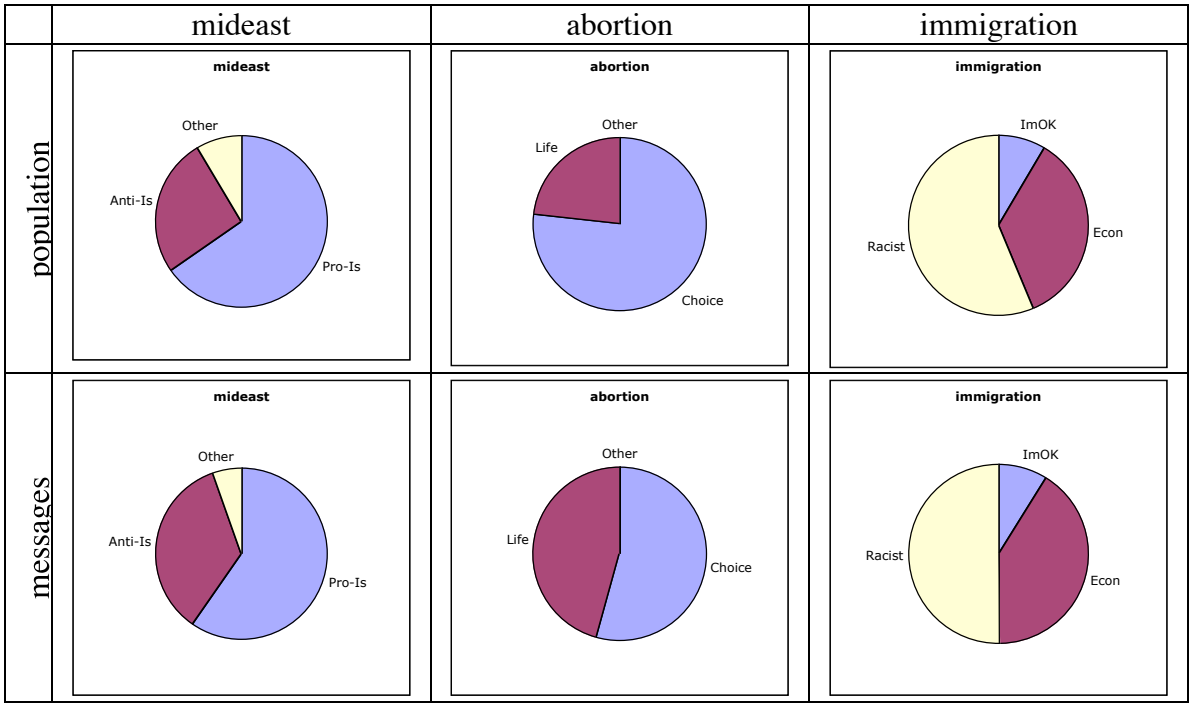
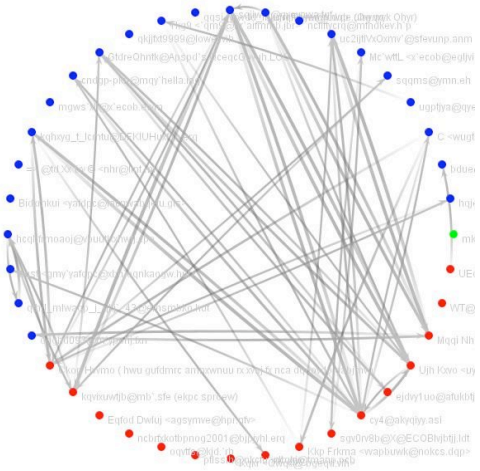
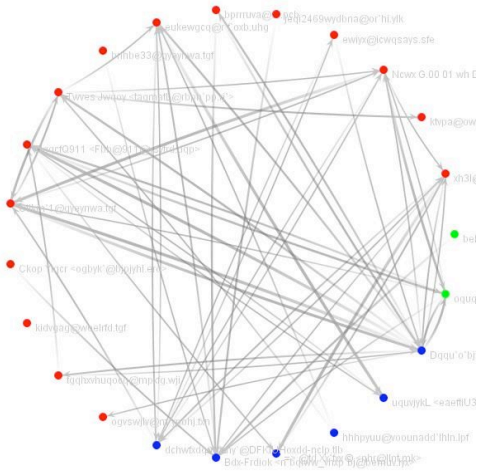
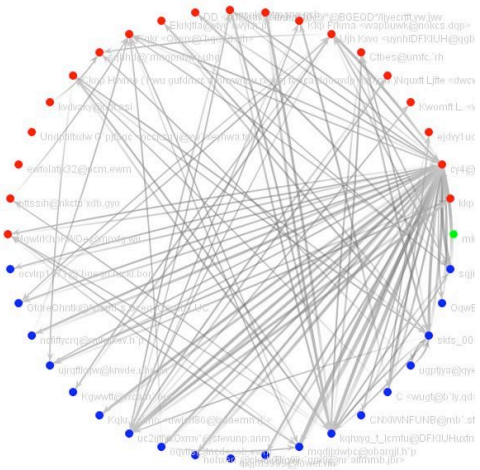
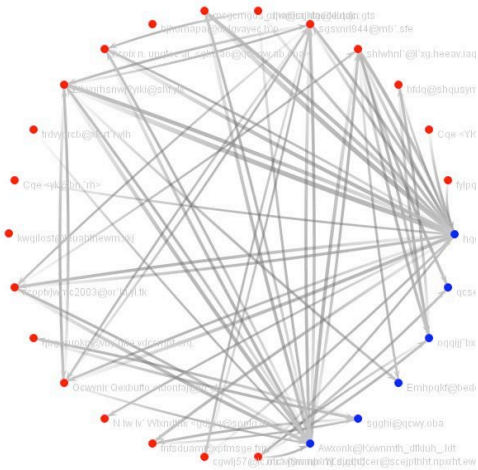
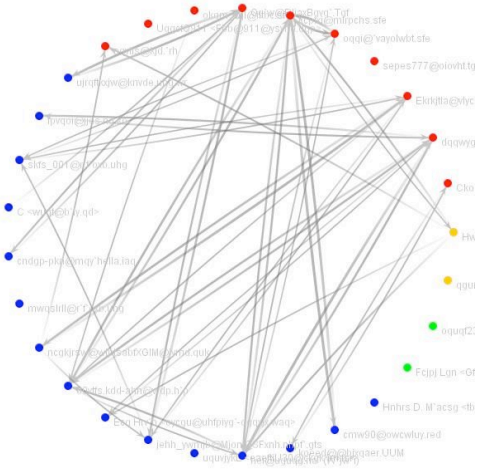
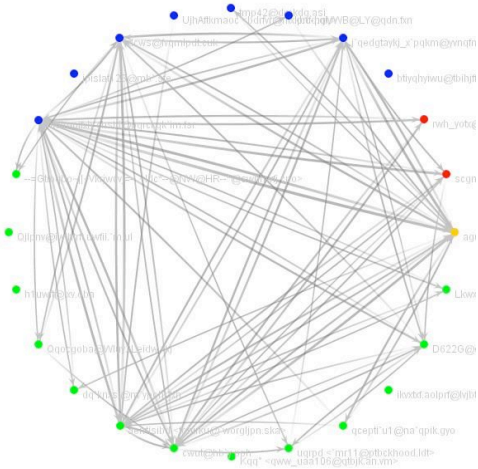


Fig. 4

- In *issue* groups and *small* groups, authors in the second (minority) clusters "speak more" than those in the majority. This makes all *issue* groups more balanced in message traffic than they are in the population of top authors. In the case of **immigration**, the clusters that become more balanced are the two varieties of anti-immigration voices. (**fig. 4**)

Fig. 5

Ideology Groups	Issue Groups
 alt.politics.bush (9)	 talk.politics.mideast (3)
 alt.politics.democrats.d (3)	 talk.abortion (12)
 alt.fan.noam-chomsky (3)	 alt.politics.immigration (3)

The Dominance of Debate and Cross-cutting Ties:

Spend a few minutes looking at Figure 5, and becomes clear that each diagram tells a similar story. Regardless of how balanced or unbalanced the population of authors in a political newsgroup, the strongest conversational links are across political divides. Here and there one finds an author speaking with someone on their own side, but this pales in comparison with how often we find authors in conversation with multiple "opponents." There are some very active authors who seem to take on the nearly everyone on the other side. It is also striking how interconnected these groups are through this pattern of opposition. And keep in mind that, to enhance the signal in the noise, the diagrams are only showing connections above a certain threshold of message traffic (# of messages in parenthesis after newsgroup name). Political newsgroups are very densely connected venues of discussion. This judgment is based on direct replies, a very high standard for establishing a link between authors in newsgroups where a great many people are reading conversations without necessarily joining in at all times. Looser standards for defining a conversational link, such as co-threadedness, would be vastly denser still.

That newsgroups are, overwhelmingly, arenas of debate is obvious from this analysis, but the simplicity of this conclusion hides some complexities that are worth looking at more closely. Debate is not the only interaction in newsgroups. People often chat with folks on their own side of the issues: sharing information, being social, chiding them for going easy on the competition, etc. And sometimes authors compliment one another across

political divides. Some of these newsgroup gladiators show obvious respect for the best of their opponents and disdain for the worst of their co-ideologues, while others believe their "enemies" to be truly evil people. Some people refuse to "speak" to people on the other side of the issue at all, directing all replies within their own cluster. The larger patterns we have observed emerge from the collective behavior of a range of individuals, among whom there are dominant tendencies, but nonetheless a range of behaviors.

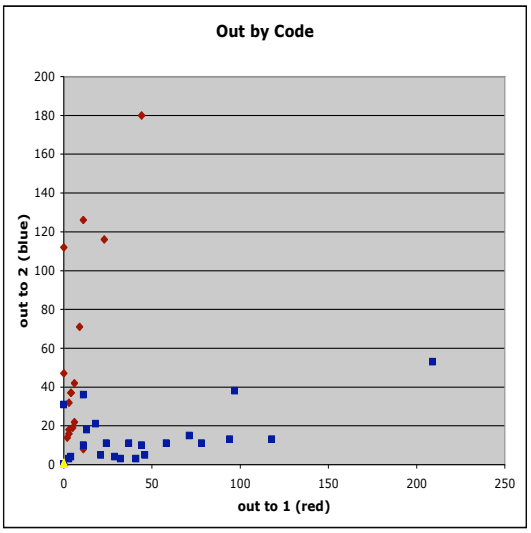


fig. 6

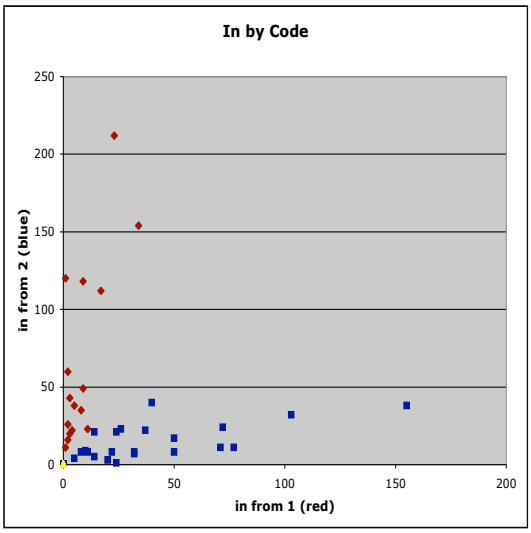


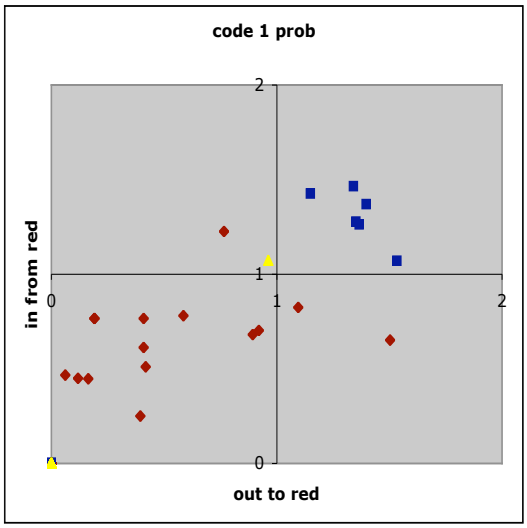
fig.7

The diagrams above plot individual authors in **bush** according to the number of replies sent to (fig. 6), and received from (fig. 7), all other coded authors, according to ideological category. **Out** links represent authors' choices about whom to talk to, whereas **in** links represent others wanting to talk to an author. Put another way, **out** is the author's *intention*, and **in** is the author's *market*. *Right* (code 1) authors are red, and *left* (code 2) authors are blue. The pattern is clear, authors overwhelmingly reserve the bulk of their replies for members of the opposing cluster. Interestingly, even those rare

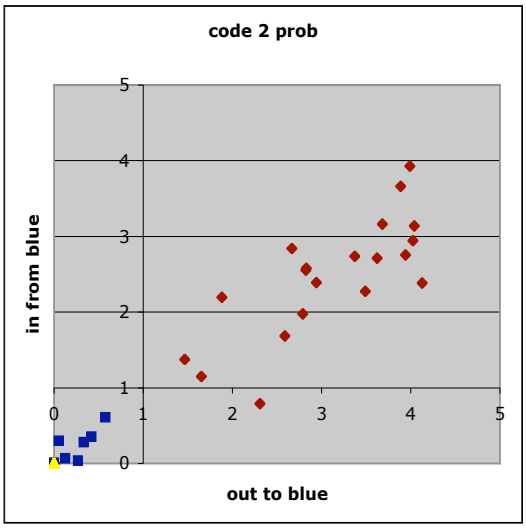
authors who target their own kind more often (notice the blue authors located near the y-axis in fig.6) nevertheless still receive more replies from the other camp. This "market correction" phenomenon suggests that graph-partitioning approaches to predicting unknown individuals' positions would benefit from using directed graphs that distinguish between in- and out-links in the algorithm. This general pattern found in **bush** holds in the other newsgroups as well, with the exception of **immigration**.

Because **bush** is a population-balanced newsgroup, it is easy to understand how the propensity of most authors to debate gives rise to the distribution seen above, and to the "market correction" mechanism. But what of the dynamics in unbalanced groups? If you are in the minority of a very unbalanced group, and reply randomly to other authors (and receive replies randomly from them), chance alone will probably link you more closely with the opposition. It is useful, especially for analyzing unbalanced groups, to consider authors' deviations from chance based on population distribution.

(fig.8) mideast



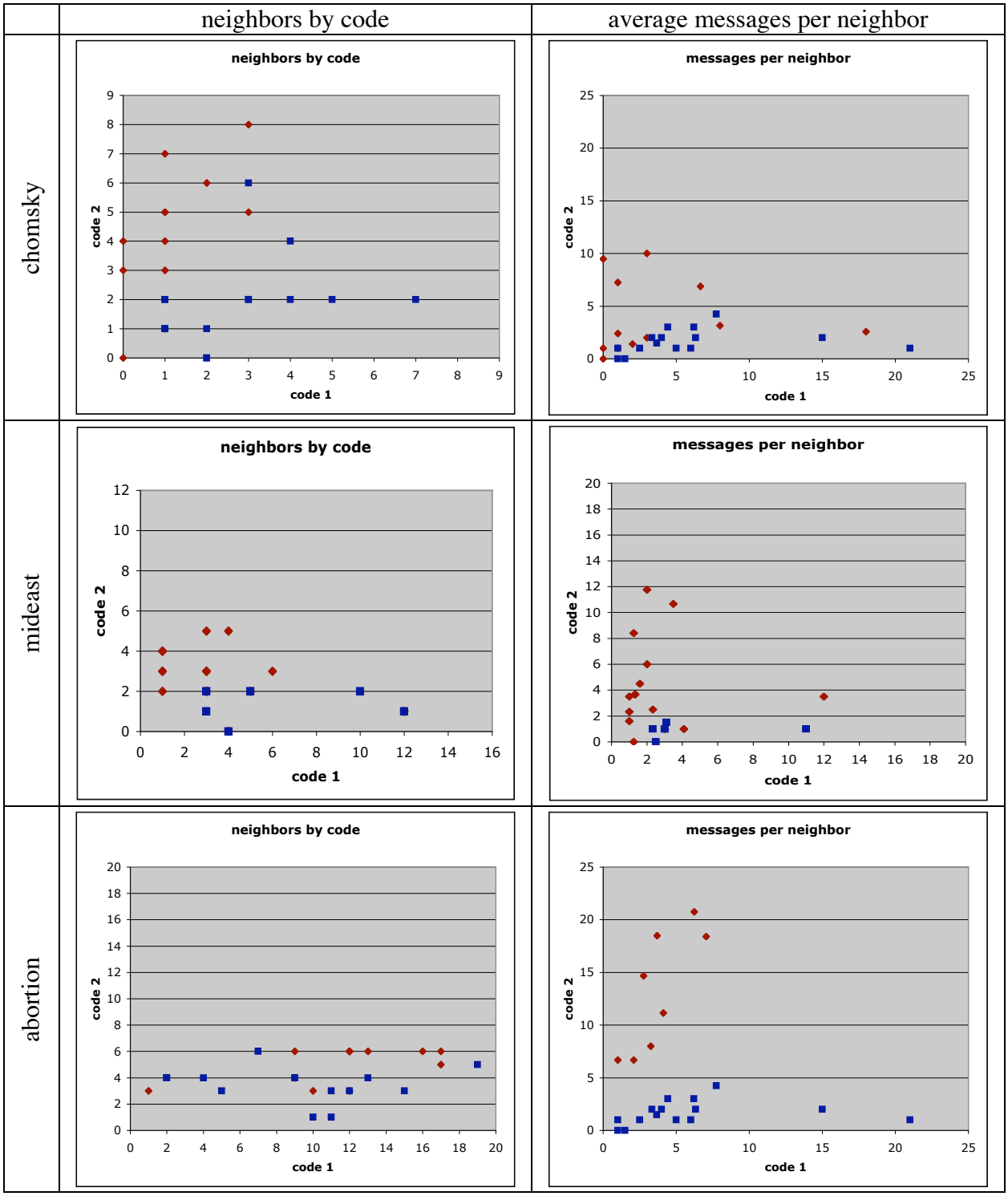
(fig.9) abortion



Correcting for probability, we still see that the preference for heterogenous interactions holds even in very unbalanced newsgroups. In **fig. 8** and **fig. 9**, the dimensions represent the proportion of actual links to links predicted by random assignment to the population of coded authors. The origin (1,1) is where you would be if you replied to, and got replies from, other authors randomly rather than according to ideological/issue criteria. An author's coordinates can be thought of as (x,y), where x=preference for replying to members of an ideology/issue cluster, y=preference of an ideology/issue cluster's members to talk to the author). In the case of **mideast** (fig. 8), code 1 (pro-Israel) is heavily in the majority. We see that anti-Israel authors (blue) uniformly target their replies to, and receive their replies from, code 1 authors (red). Conversely, two code 1 authors prefer to reply to others within code 1, yet despite these overtures are still more likely to get replies from code 2. Interestingly, one anomalous code 1 author is more popular with her own cluster. In **abortion** (fig.9) we see the same preference for debate, but since we are looking at (deviations from random likelihood of) links to and from the

minority code (code 2, pro-life), in the magnitude of deviation we can see the determination with which code 1 authors interact with code 2 authors.

Intuitively, debate is more conducive than back-patting to long chains of replies. There may in fact be something in the nature of threaded discussion that helps long arguments prosper, as anyone who spends a lot of time in newsgroups has probably contemplated. An argument could be made that more message traffic across (rather than within) ideological divides may be largely due to this fact, and not because of a strong preference for being in contact with heterogenous others. So it is useful to decompose measures of message traffic into a.) measures of in and out *degrees* (i.e. how many direct "neighbors" one has), and b.) average number of messages sent to neighbors in one's own cluster vs. sent to neighbors in other clusters. Interestingly, we see strong similarities in patterns between different groups, but also some differences, particularly between *ideology* and *issue* groups.

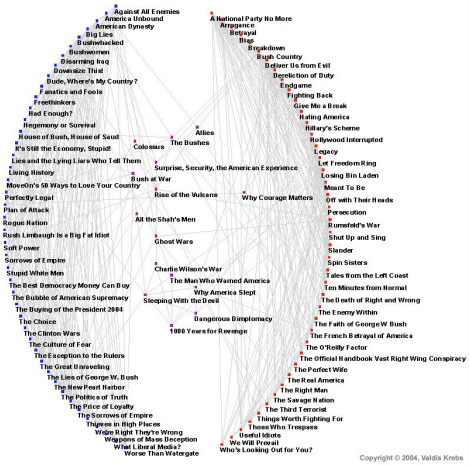

neighbors by code
average messages per neighbor
chomsky
mideast
abortion
neighbors by code
messages per neighbor
code 1
code 2

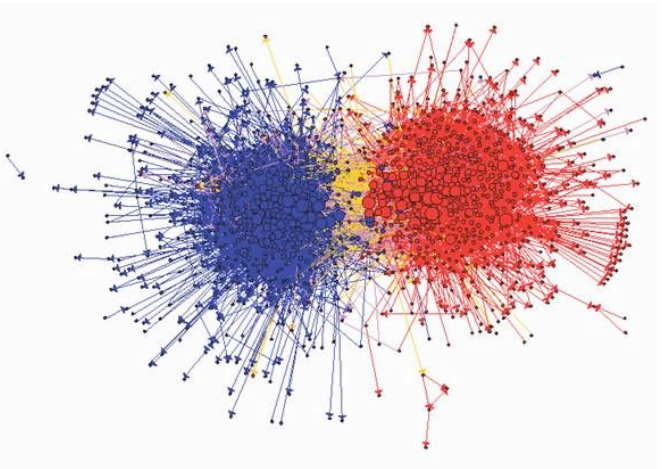
(fig. 10)

In these diagrams (fig. 10) we see relationships based on **outgoing** replies, in other words we are looking at authors decisions about whom they talk to. The patterns we see in **chomsky** are shared by the other *ideology* groups as well. In terms of neighbors (i.e. authors directly connected by incoming and outgoing replies), code 1 (*right*) authors generally have more code 2 (*left*) neighbors. Similarly, *left* authors connect with more *right* authors than with their own kind. The effect of this is then magnified by the tendency to send more messages per neighbor to authors across the ideological divide. In balanced *ideology* groups, therefore, the preponderance of heterogeneous message traffic is not just an artifact of longer debate chains, but arises from choices to connect with more people across the divide, as well as to send those people more messages each.

Interpretation is a tad more difficult for the *issue* groups, since the extreme imbalances in author ratios limit the number of minority authors available to talk to, making patterns in the distribution of neighbors less sharp. Nonetheless, it appears that, at least in the case of **mideast**, the same preference for replying to opponents is present. In terms of average messages per neighbor, we see the same basic pattern as in the *ideology* groups.

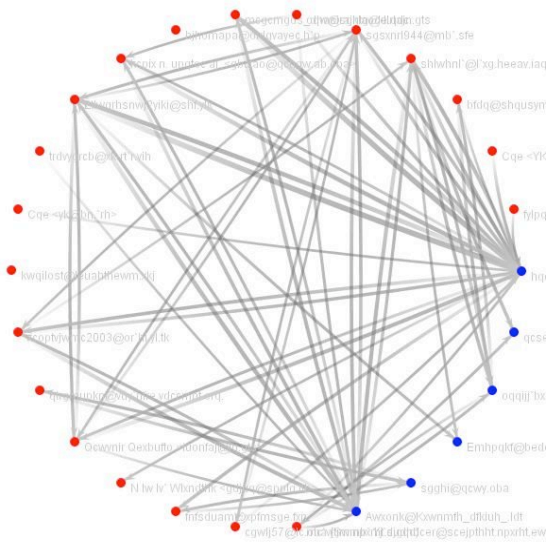
Conclusions



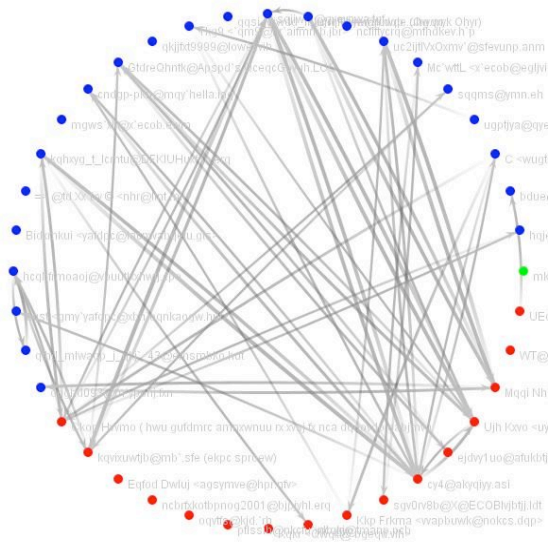


[Valdis Krebs] [Adamic and Glance]

The two images above have entered the pantheon of images often discussed at the crossroads of network analysis and political communication. On the left we see Valdis Krebs' visualization of buying patterns of politically relevant books from Amazon.com, demonstrating two distinct audiences, divided by ideology in terms of information they consume. On the right we see Adamic and Glance's visualization of links among liberal and conservative political blogs, demonstrating the same ideological divisions. A claim found in the tradition of Communications research is that, the world one lives in is created by the world of information one consumes and exchanges. Here we see division, seperate worlds, ideologically divergent public sphericules. By way of contrast, let us consider the following diagrams:



talk.abortion



alt.politics.bush

Here we have two USENET newsgroups, and the patterns of ties are exactly the opposite of those seen with book purchases and blogs. One the left, pro-choice and pro-life authors engage each other vigorously, eclipsing the time and energy spent of dialoging with people they already agree with. On the right, liberals and conservatives do the same, discussing a wide range of topics on the public agenda. Neither groups is, to be exact, *deliberating*. Most of what one reads would not rate highly on Habermasian metrics of rational-critical discourse. But here is a place where even the most extremely opposed voices are talking to one another, sometimes in name-calling flame wars, sometimes in extended rational conversations, and sometimes quoting the New York Times, Adam Smith. Karl Marx, the Federalist Papers, and the Oxford English Dictionary. The important point is that these environments are public commons that expose citizens to debate with diverse others. J.S. Mill would approve.

That this assessment contradicts much previous research on online political discussion (mainly, like this one, studies of USENET) is (we hope) partially a result of the fact that more powerful tools are now available to investigate the matter. First-generation methods employing message sampling and simple content analysis cannot capture the multidimensional dynamics of political discourse on the internet. Tools such as Microsoft Research's Netscan, Warren Sack's Conversation Map, Agrawal, *et al's* constrained graph partitioning, and the kinds of network visualizations pursued by Adamic, Glance, and others are the way forward. The trick is to leverage rich, qualitative interpretative techniques that require human analysts against the extraordinary power of computational tools that can make sense of very large scale phenomena like USENET, blogs, etc.

The point here is not that USENET political newsgroups are the end-all of the online public sphere. But "open source" spaces for dialog on the internet can, and probably do, collect minds with a range of policy preferences and ideological groundings—and they talk to each other. An online commons is likely perceived as ideational territory to fight for by those in the neighborhood. Blogs, and discussion environments accessed through gateways that filter the public by common selectivity biases (e.g., discussion fora on mass-media websites), may in fact be more likely to trigger the kind of "birds of a feather" phenomena Sunstein so rightly warns us against.

For those trying to understand online political discourse, including those designing next-generation online forums, there are clearly lessons to learn by digging deeper into the

behavior we see in political USENET groups. Clear patterns in aggregate behavior mask a wide range of individual behaviors. Are those who respond more frequently to members of the same ideological cluster "reinforcing" rather than debating? Or are the analyst's ideological categorizations too restrictive, burying a rich terrain of variation in political thought under crude conventional labels? Conversely, perhaps conventional ways of understanding political thinking are accurate, and individual variations in behavior reveal preferences for whom to engage in dialog. These and many other questions are posed by the variation in individual behavior we see in newsgroup authors.

Adamic, L. a. N. G. (2005). The Political Blogoshphere and the 2004 U.S. Election: Divided They Blog. BlogPulse, Intelliseek.

Agrawal, R. S. R. R. S. Y. X. (2003). Mining Newsgroups Using Networks Arising from Social Behavior. WWW2003, Budapest, Hungary.

Broncheck, M. S. (1997). From Broadcast to Netcast: The Internet and the Flow of Political Information. Political Economy and Government. Cambridge, Massachusetts, Harvard**:** 259.

Dahlberg, L. (2001). "Computer-Mediated Communication and The Public Sphere: A Critical Analysis." Journal of Computer Mediated Communication **7**(1).

Gitlin, T. (1998). Public Sphere or Public Sphericules? Media, ritual, and identity. T. Liebes, J. Curran and E. Katz. New York, Routledge**:** viii, 265.

Graham, T. a. T. W. (2003). "In search of online deliberation: Towards a new method for examining the quality of online discussions." Communications **28**: 173-204.

Gutmann, A. a. D. T. (2004). Why Deliberative Democracy?

Hill, K. A. a. J. E. H. (1997). "Computer-Mediated Political Communication: The USENET and Political Communities." Political Communication **14**: 3-27.

Krebs, V. (2004). orgnet.com.

McPherson, M. a. L. S.-L., and James M. Cook (2001). "Birds of a Feather: Homophily in Social Networks." Annual Review of Sociology **27**: 415-445.

Mutz, D. C. (2002). "The Consequences of Cross-Cutting Networks for Political Participation." American Journal of Political Science **Vol. 46**(No. 4): pp. 838-855.

Mutz, D. C. a. P. S. M. (2001). "Facilitating Communication across Lines of Political Difference." American Political Science Review **Vol. 95**(No. 1): pp. 97-114.

Noelle-Neumann, E. (1984). The spiral of silence : public opinion, our social skin. Chicago, University of Chicago Press.

Papacharissi, Z. (2004). "Democracy online: civility, politeness, and the democratic potential of online political discussion groups." New Media and Society **6**(2).

Sack, W. (2004). Conversation Map.

Smith, M. a. D. F. (2005). Social Network Visualizations of Usenet Newsgroups. Sunbelt XXV: International Sunbelt Social Network Conference, Redondo Beach, CA.

Steenbergen, M. R. a. A. B., Markus Sporndli, and Jurg Steiner (2003). "Measuring Political Deliberation: A Discourse Quality Index." Comparative European Politics **1**: 21-48.

Sunstein, C. R. (2001). Republic.com. Princeton N.J., Princeton University Press.

Wilhelm, A. (1999). Virtual Sounding Boards: How deliberative is online political discussion. Digital democracy : discourse and decision making in the Information Age. B. N. Hague and B. Loader. London ; New York, Routledge**:** xvi, 277.