

## E-MAIL THREADS, GENRES & NETWORKS IN A PROJECT MAILING LIST

Moses A. Boudourides\*, Manolis Mavrikakis<sup>#</sup>, Eleftheria Vasileiadou<sup>&</sup>

Department of Mathematics  
University of Patras  
265 00 Rio-Patras  
Greece

\* [mboudour@upatras.gr](mailto:mboudour@upatras.gr)

<sup>#</sup> [manolism@master.math.upatras.gr](mailto:manolism@master.math.upatras.gr)

<sup>&</sup> [Eleftheria@nicomedia.math.upatras.gr](mailto:Eleftheria@nicomedia.math.upatras.gr)

We are studying here the use of e-mail by a group of individuals in a mailing list. This is a group of partners (from four European countries) working in a EU-funded RTD project, COMB (a pseudonym). The members of the COMB project have been communicating through the COMB-L mailing list. Originally this mailing list had started in August 2000 when the funding of the project was approved. Subsequently the discussions in the mailing list were concerning the negotiations with the EC for the formulation of the project contract. Eventually, the project started on May 1, 2001, and it is on-going at the moment; totally, it will last 30 months, i.e., it is expected to end up by November 2003.

The primary data for this study consists of the transcripts of 1104 e-mails distributed in the COMB-L mailing list from August 22, 2000, until December 31, 2001. The data include the following information for each e-mail:

- (i) the date of distribution,
- (ii) the sender,
- (iii) the subject of the e-mail and
- (iv) the contents of the message body.

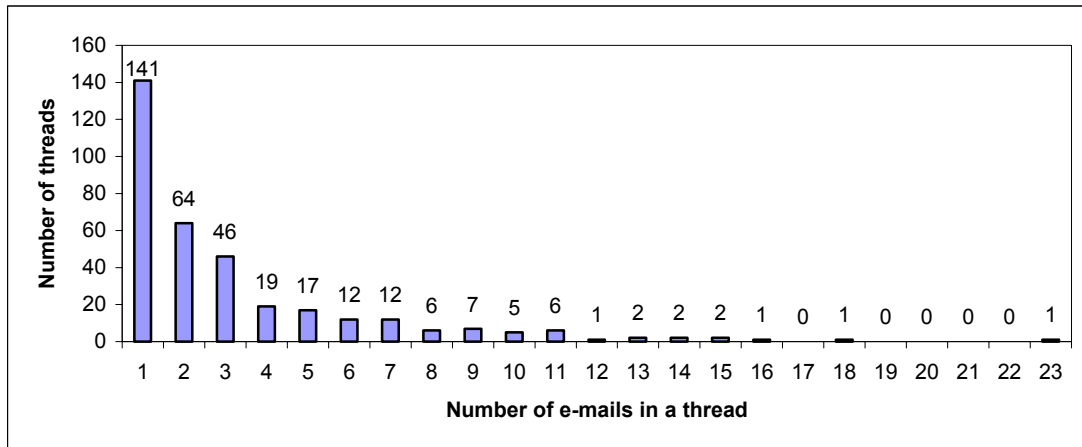
From the above information, two principal variables are derived:

- The *e-mail threads*, i.e., groups of replies to distinct initial e-mails, which are extracted from the subjects of the e-mails.
- The *e-mail genres*, which are defined through a number of different categories, coded as present or absent, by reading the contents of the e-mails.

Accordingly, our purpose is to undertake two analyses of the mailing list: a *thread analysis* and a *genres analysis*. In particular, in each one of these analyses, after giving some general statistics, we intend to generate the corresponding networks (for each month and for the whole period) and to compute certain important network parameters.

## 1. Thread Analysis

Threading of e-mail is an important tool for processing mass electronic interaction (Lewis & Knowles, 1997; Whittaker *et al.*, 1998). In our mailing list data (1104 e-mails), we have found 345 threads and among them 204 threads (i.e., 59.13% of the total threads) are containing more than one e-mail (totally all these threads are containing 964 e-mails). In fact, the distribution of the number of e-mails among threads is shown in the following chart:



### 1a. The Network of Threaded Senders

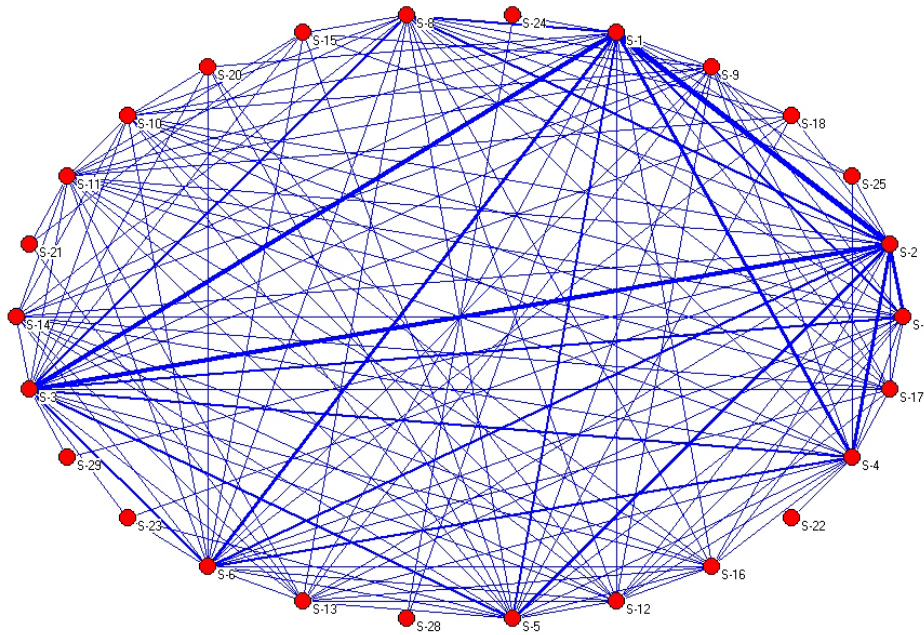
We have constructed the *network of threaded senders*, in which the actors are the senders of the e-mails: Two actors are said to be linked in this network, whenever they both have sent e-mails in the same thread and the weight of their link equals the number of different threads where these senders have participated. Apparently, as far as two linked senders are contributing in the same thread of e-mails, this might be interpreted as a somehow strong communicative interaction between these senders, which would justify the existence of this network of threaded senders.

In the following animation, we are showing successively the networks of threaded senders for each of the 17 months (from August 2000 until December 2001):

[Link to an Animation of the Network of Threaded Senders from August 2000 to December 2001](#)

Over the whole period from August 2000 until December 2001, the network of threaded senders appears in the following figure:

## August 2000 - December 2001



### 1b. Network Properties of Threaded Senders

In each month, three important properties of the network of threaded senders (density, degree centralization and betweenness centralization, all computed for the corresponding binary networks) are given in the following table:

	Density	Degree Centralization (%)	Betweenness Centralization (%)
Aug2000	0.0277	18.67	1.65
Sep2000	0.0985	28.33	3.58
Oct2000	0.1262	29.67	3.21
Nov2000	0.0000	0.00	0.00
Dec2000	0.0708	22.67	0.42
Jan2001	0.0154	11.33	0.16
Feb2001	0.0462	16.67	0.00
Mar2001	0.1169	26.33	0.44
Apr2001	0.0615	23.67	2.42
May2001	0.2123	33.33	4.34
Jun2001	0.1354	37.33	5.70
Jul2001	0.1569	39.33	7.19
Aug2001	0.0769	30.67	8.19
Sep2001	0.0985	28.33	5.38
Oct2001	0.1015	28.00	2.75
Nov2001	0.2000	34.67	5.08
Dec2001	0.0615	19.33	3.58
All	0.4677	36.00	16.39

Furthermore, the senders (as actors in this network defined by threadings) possess the following network properties during the whole period (August 2000 to December 2001):

	No of Mails	% Mails	No of Threads	% Threads	Degree	Closeness	Betweenness
S-1	159	14,40	99,00	46,26	80.00	83.333	18.233
S-2	155	14,04	89,00	41,59	80.00	83.333	8.333
S-3	132	11,96	87,00	40,65	80.00	83.333	7.632
S-4	114	10,33	66,00	30,84	68.00	75.758	9.206
S-5	77	6,97	55,00	25,70	72.00	78.125	3.422
S-6	76	6,88	49,00	22,90	76.00	80.645	3.799
S-7	68	6,16	48,00	22,43	60.00	71.429	0.140
S-8	56	5,07	35,00	16,36	72.00	78.125	2.638
S-9	41	3,71	27,00	12,62	60.00	71.429	0.610
S-10	38	3,44	26,00	12,15	60.00	71.429	0.674
S-11	36	3,26	24,00	11,21	72.00	78.125	4.926
S-12	31	2,81	18,00	8,41	64.00	73.529	0.821
S-13	24	2,17	14,00	6,54	56.00	69.444	2.667
S-14	23	2,08	14,00	6,54	60.00	71.429	0.140
S-15	17	1,54	14,00	6,54	32.00	55.556	1.027
S-16	16	1,45	14,00	6,54	52.00	67.568	0.026
S-17	14	1,27	8,00	3,74	48.00	65.789	0.000
S-18	5	0,45	3,00	1,40	28.00	56.818	0.000
S-19	5	0,45	1,00	0,47	0.00	00.000	0.000
S-20	4	0,36	4,00	1,87	24.00	54.348	0.000
S-21	3	0,27	2,00	0,93	12.00	46.296	0.042
S-22	2	0,18	1,00	0,47	4.00	43.860	0.000
S-23	2	0,18	1,00	0,47	16.00	50.000	0.000
S-24	1	0,09	1,00	0,47	8.00	47.170	0.000
S-25	1	0,09	1,00	0,47	16.00	53.191	0.000
S-26	1	0,09	0,00	0,00	0.00	00.00	0.000
S-27	1	0,09	0,00	0,00	0.00	00.00	0.000
S-28	1	0,09	1,00	0,47	8.00	47.170	0.000
S-29	1	0,09	1,00	0,47	8.00	48.077	0.000

## 2. Genre Analysis

Since the discussions taking place in the COMB-L mailing list were intended to advance the work of the project, these exchanges of mailings can be considered to constitute a computer-mediated organizational communication. As such it is very interesting to investigate the genres of communicative discourse developed in the e-mails of the mailing list. For this purpose, we are going to follow the work of Joanne Yates and Wanda Orlikowski who have extensively and analytically studied the genres of organizational communication. The remarkable idea developed by these scholars is that genres, through which information is communicated, shaped and shared for particular purposes, are not just an aspect of organizational work; these genres in a sense do constitute the organizational work in itself. Because in organizations, groups and professional communities, each genre repertoire “defines a different set of interaction norms and work practices, and each serves to define a different kind of community” (Orlikowski & Yates, 1994b, p. 5). In other words, it is a genre repertoire that defines an organization’s nature, its communicative interactions, its rules and work practices. Thus, undertaking the genres analysis of the project mailing list would help us understand better the organizational structures and dynamics developed and communicated through the collaborative work by the members of the project team.

Orlikowski and Yates define genres “as socially recognized types of communicative actions – such as memos, meetings, expense forms, and training seminars – that are habitually enacted

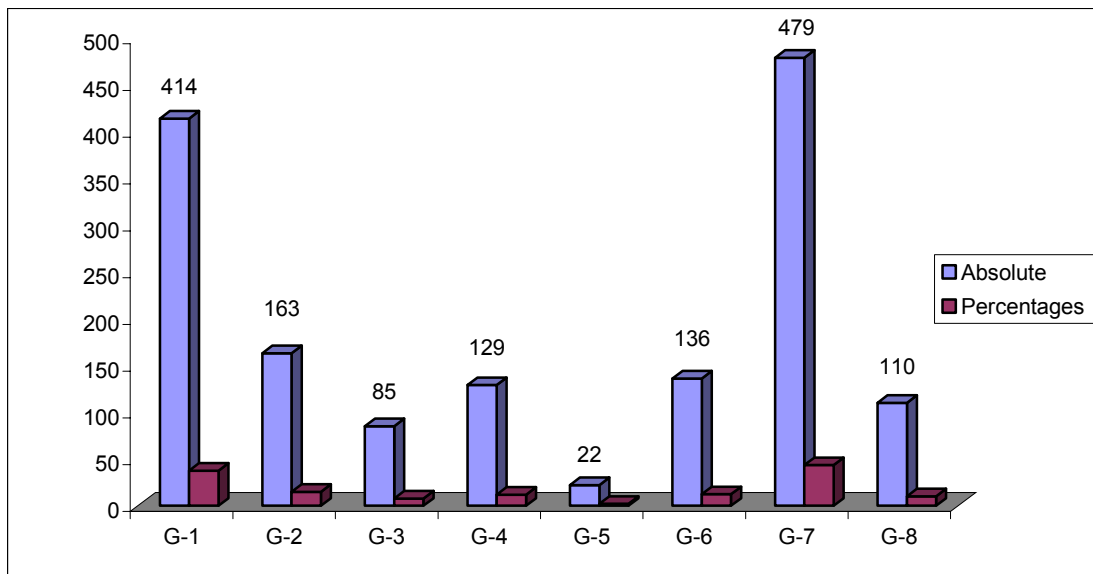
by members of a community to realize particular social purposes” (1994a, p. 542). Furthermore, they require that a genre might be identified by its socially recognized purpose and shared characteristics of form (Yates & Orlikowski, 1992, pp. 301-2). The purpose of a genre is not the individual’s private motive for communicating, but a purpose constructed and recognized by the relevant organizational community, whether small or large. Form refers to the observable aspects of the communication, such as communication medium (e.g., pen and paper, telephone, or face to face), structural features (e.g., text formatting devices such as lists and structured fields) and linguistic features (e.g., level of formality, specialized vocabulary, or graphic devices).

The genre repertoire that we have employed in this paper consists of the following eight genres defined in terms of their specific form and purpose:

<p><b>G-1: Dialogue</b>  <i>Form</i> = embedded message, reply subject.  <i>Purpose</i> = response.</p>
<p><b>G-2: Team Announcement</b>  <i>Form</i> = subject relevance, professional/neutral language, general opening.  <i>Purpose</i> = informative.</p>
<p><b>G-3: Socializing</b>  <i>Form</i> = informal language.  <i>Purpose</i> = socializing.</p>
<p><b>G-4: Distribution of Completed Work</b>  <i>Form</i> = attachment, subject relevance.  <i>Purpose</i> = completed work.</p>
<p><b>G-5: Reminder</b>  <i>Form</i> = professional/neutral language, general opening.  <i>Purpose</i> = reminder.</p>
<p><b>G-6: Group Decision</b>  <i>Form</i> = reply subject, embedded message.  <i>Purpose</i> = decision related.</p>
<p><b>G-7: Distribution of Project Work</b>  <i>Form</i> = attachment, embedded message.  <i>Purpose</i> = document preparation.</p>
<p><b>G-8: Criticisms</b>  <i>Form</i> = informal/colloquial or professional/neutral language.  <i>Purpose</i> = criticisms.</p>

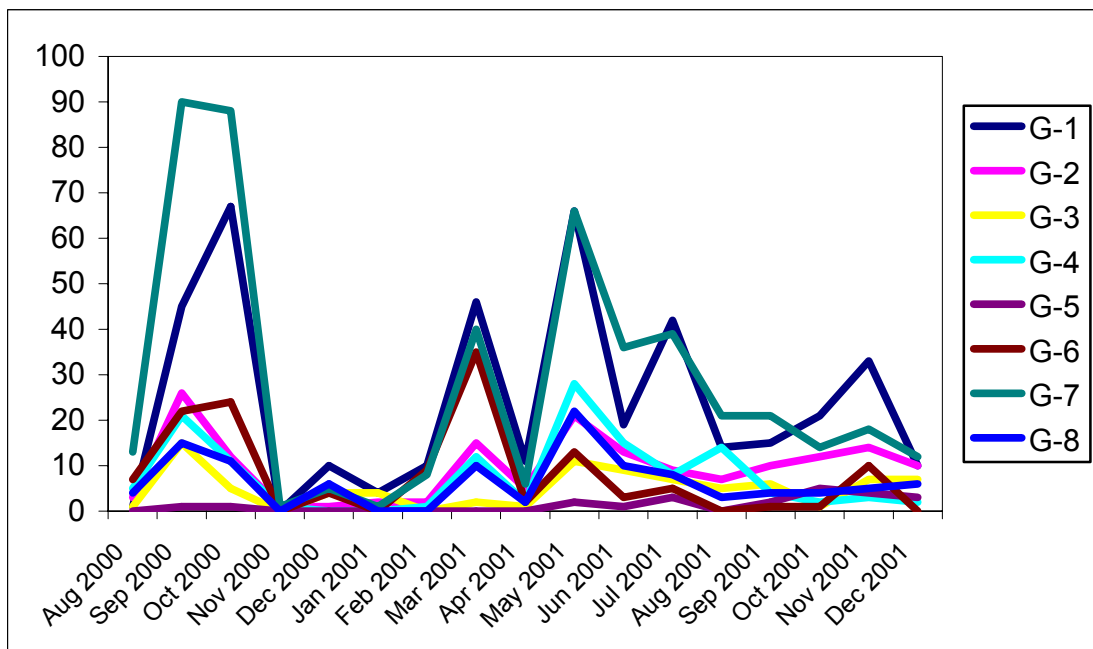
The above genres have been constructed as follows: We have used a coding scheme based on the two dimensions constituting the definition of genre: form and purpose (Orlikowski & Yates, 1994a, p. 552). Our coding scheme consisted of 36 categories (12 of form and 24 of purpose). After reading all the e-mails of the mailing list, these categories were coded simply as present or absent. Certain of these categories were included in the definition of the used genres as shown in the above table. In general, the used genres in our project mailing list were more or less similar to the ones employed by Orlikowski and Yates in their study of genres emerging from organizational communication in mailing lists (cf., Orlikowski & Yates, 1994a; Yates, Orlikowski & Okamura, 1999a & 1999b).

Thus, we found that the above eight genres were distributed among the 1104 e-mails of the COMB-L mailing list as follows (note that more than one or none genre were found in some e-mails):

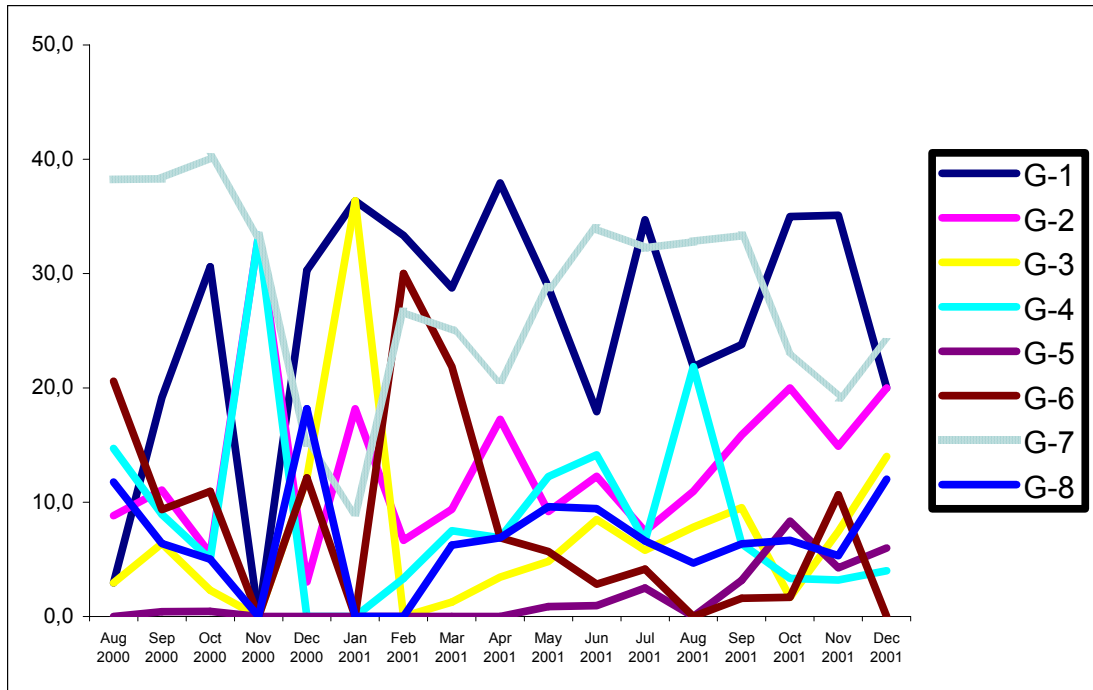


## 2a. Time Variation of Genres

We found the following two diagrams for the *time evolution of the used genres* during the whole period of 17 months during which the e-mails were distributed in the COMB-L mailing list. The first diagram shows the evolution in absolute volume of e-mails per month in which the genres are appearing:

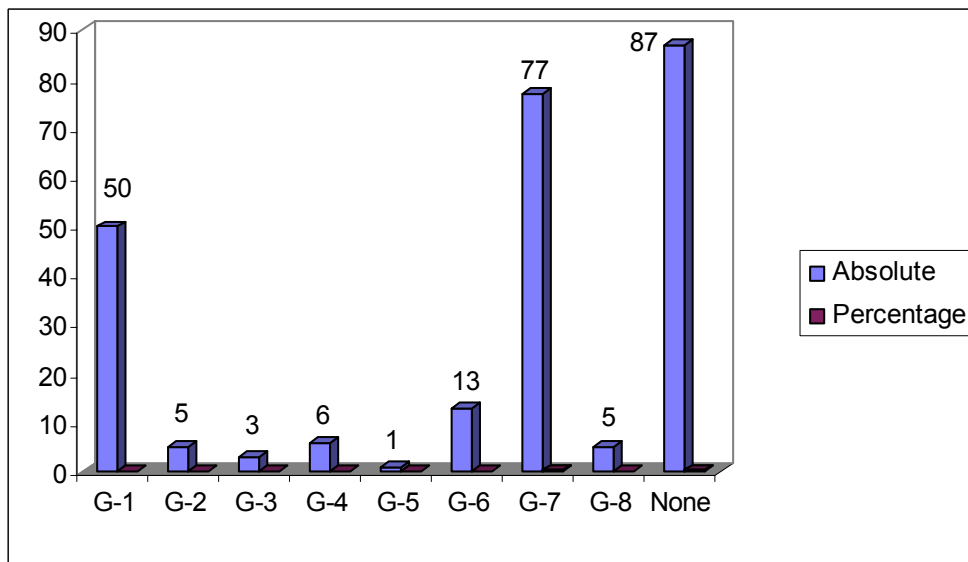


The second diagram depicts the relative time evolution of the used genres, in which the volume of e-mails for each genre per month is divided by the total number of e-mails during that month:



## 2a. Genres over Threads and Senders

Next comes the study of how the genres are distributed over the e-mail threads. In particular, our purpose was to study whether there exists a dominant genre in a thread, ‘dominant’ meaning that a genre appears in the majority of the e-mails of the thread. Here we have only considered the 204 threads consisting of more than one e-mail and we found that among these threads each of the eight used genres is dominant as many times as indicated in the following chart:



It is interesting to see whether the genre of the root (first) e-mail in the thread persists or disappears through the subsequent e-mails of the thread. In fact we found that in 62 threads (30.39%) the genre of the root e-mail is dominant too while in 142 threads (69.17%) the genre of the root e-mail is succeeded by a different dominant genre.

Finally, the following table presents the number of genres employed by each sender of the mailing list e-mails:

	G-1	G-2	G-3	G-4	G-5	G-6	G-7	G-8
S-1	61	6	29	9	1	22	49	11
S-2	72	28	3	10	4	22	61	30
S-3	59	17	4	10	0	29	76	15
S-4	40	23	4	21	0	9	44	11
S-5	31	8	2	6	0	12	41	5
S-6	30	7	8	14	2	12	41	7
S-7	16	13	1	13	1	13	32	11
S-8	24	6	4	9	2	9	26	5
S-9	17	3	8	1	0	1	12	1
S-10	11	15	3	10	3	1	24	6
S-11	7	0	6	1	0	1	10	1
S-12	9	14	0	6	9	2	18	2
S-13	7	10	1	5	1	0	4	1
S-14	7	2	2	6	0	2	14	2
S-15	6	2	6	1	0	0	2	1
S-16	5	0	1	0	0	1	5	0
S-17	4	6	1	3	0	0	5	0
S-18	3	0	0	0	0	0	4	1
S-19	0	0	0	1	0	0	2	0
S-20	1	0	1	2	0	0	3	0
S-21	1	0	0	0	0	0	1	0
S-22	0	0	0	2	0	0	2	0
S-23	1	0	0	0	0	0	1	0
S-24	0	0	1	0	0	0	0	0
S-25	1	0	0	0	0	0	0	0
S-26	0	0	0	0	0	0	0	0
S-27	0	0	0	0	0	0	0	0
S-28	0	1	0	0	0	0	0	0
S-29	0	0	0	0	0	0	1	0

## 2b. Networks of Genres

In this section we are going to consider two networks of genres emerging from the mailing list e-mails: (i) the network of genres coexisting in an e-mail and (ii) the network of genres appearing in threads of e-mails.

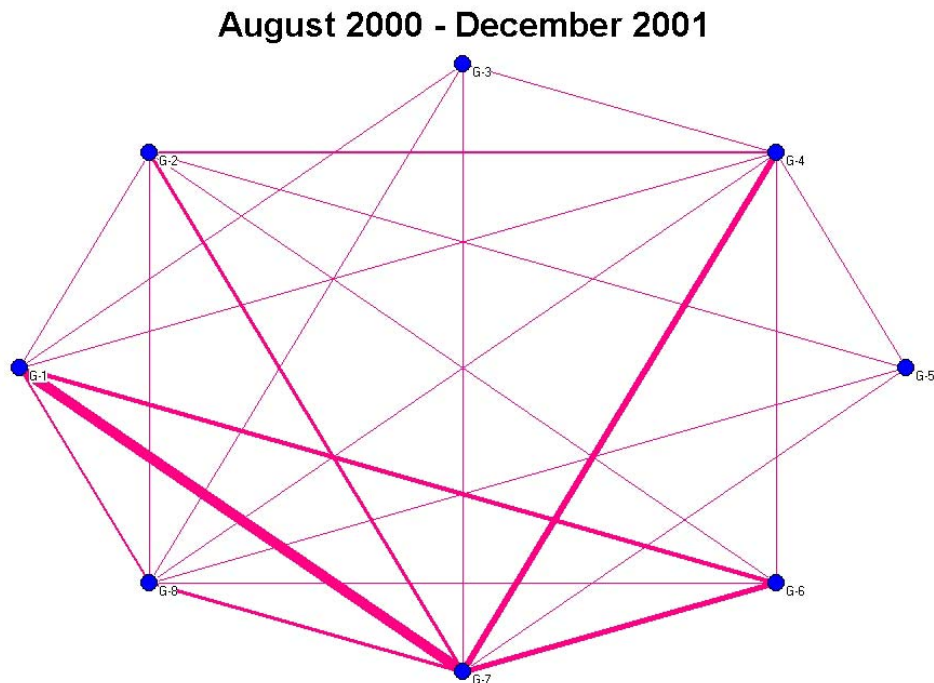
In the *network of genres coexisting in e-mails*, the actors are the eight genres: Two genres are said to be linked in this network, whenever they both appear in an e-mail and the weight of their link equals the number of different e-mails where these genres appear. The underlying idea in the definition of this network is that the coexistence of two genres in some e-mail should not be accidental but it should be related to some communicative affinity that these genres should possess so that they might employed simultaneously in an e-mail.

In the following animation, we are showing successively the networks of coexisting genres for each of the 17 months (from August 2000 until December 2001):



[Link to an Animation of the Network of Coexisting Genres from August 2000 to December 2001](#)

Over the whole period from August 2000 until December 2001, the network of coexisting genres appears in the following figure:



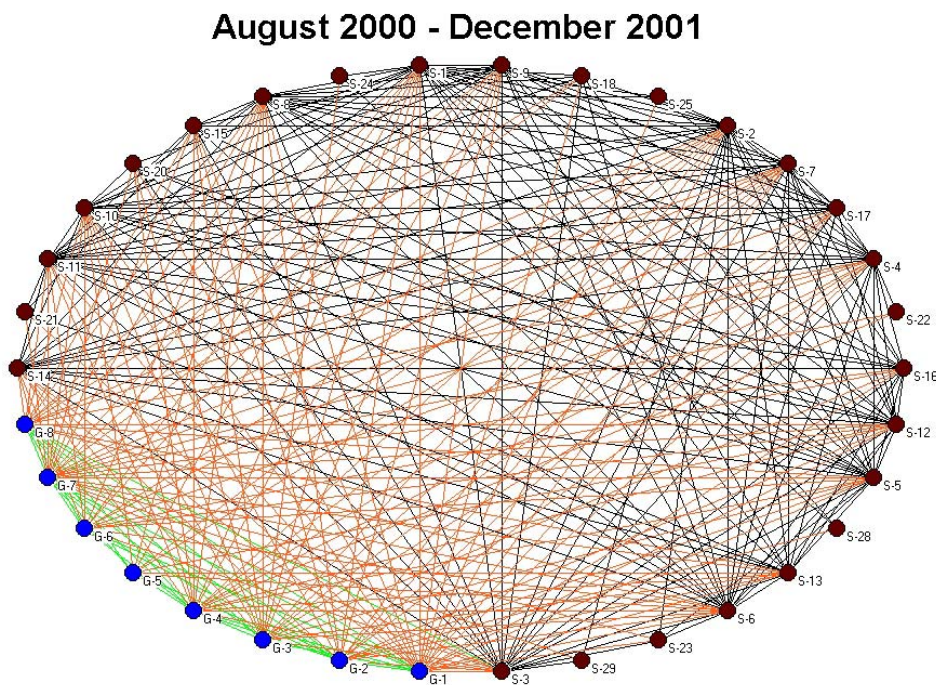
In each month, three important properties of the network of coexisting genres (density, degree centralization and betweenness centralization, all computed for the corresponding binary networks) are given in the following table:

	Density	Degree Centralization (%)	Betweenness Centralization (%)
Aug2000	0.3571	47.62	15.65
Sep2000	0.0985	28.33	3.58
Oct2000	0.6071	33.33	8.16
Nov2000	0.1071	23.81	0.00
Dec2000	0.2143	28.57	12.93
Jan2001	0.0000	0.00	0.00
Feb2001	0.2857	38.10	4.08
Mar2001	0.6071	33.33	8.16
Apr2001	0.3214	33.33	17.01
May2001	0.6786	42.86	19.73
Jun2001	0.5714	38.10	14.97
Jul2001	0.5714	57.14	27.21
Aug2001	0.3571	47.62	12.93
Sep2001	0.4643	52.38	22.45
Oct2001	0.4643	33.33	20.41
Nov2001	0.4286	38.10	27.21
Dec2001	0.2857	38.10	37.41
All	0.8214	23.81	3.85

Furthermore, the coexisting genres (as actors in this network) possess the following network properties during the whole period (August 2000 to December 2001):

	No of Mails	% Mails	Degree	Closeness	Betweenness
G-1	414	37,50	85.714	87.500	2.381
G-2	163	14,76	85.714	87.500	2.381
G-3	85	7,70	57.143	70.000	0.000
G-4	129	11,68	100.000	100.000	6.349
G-5	22	1,99	57.143	70.000	0.000
G-6	136	12,32	71.429	77.778	0.000
G-7	479	43,39	100.000	100.000	6.349
G-8	110	9,96	100.000	100.000	6.349

Furthermore, over the whole period from August 2000 until December 2001, the heterogeneous network of threaded senders and coexisting genres appears in the following figure:

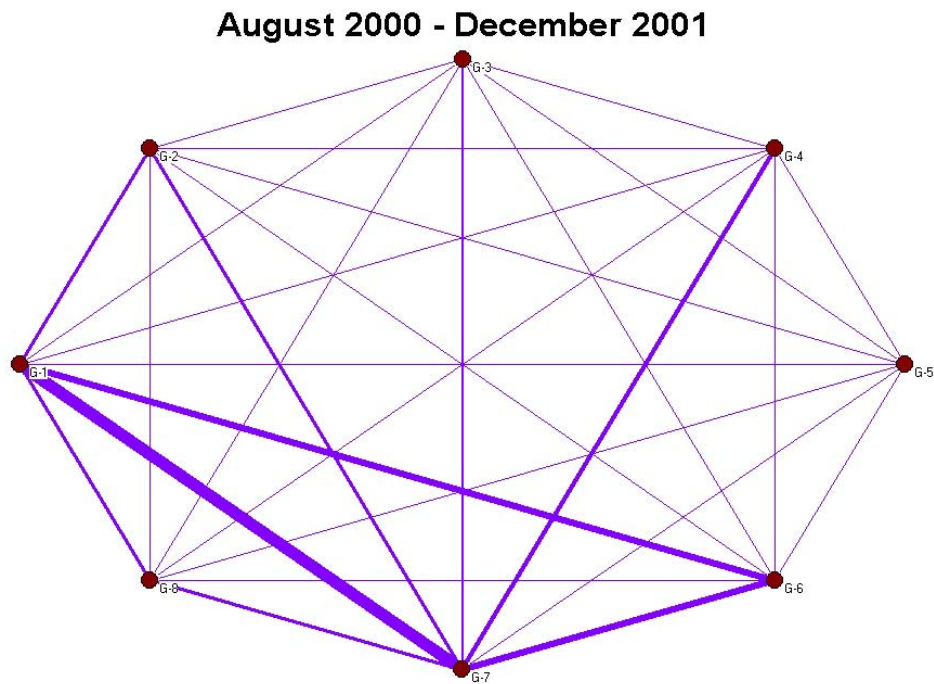


In the *network of genres appearing in threads*, the actors are again the eight genres: But now two genres are said to be linked in this network, whenever they both appear in a thread of e-mails but in different e-mails of this thread and the weight of their link equals the number of different threads where these genres appear (always in different e-mails). The underlying idea in the definition of this network is that the appearance of two genres in some thread of e-mails should not be accidental but it should be related to the fact that any thread of e-mails represents a sort of communicative unit which only supports certain genres (pertinent to the discursive character of the thread).

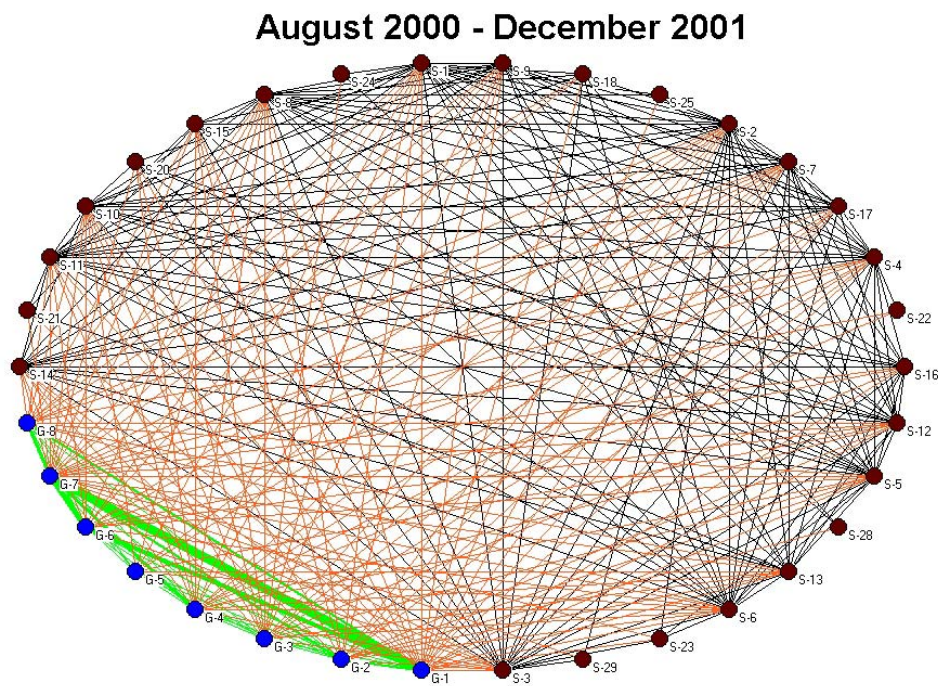
In the following animation, we are showing successively the networks of coexisting genres for each of the 17 months (from August 2000 until December 2001):

[Link to an Animation of the Network of Genres Appearing in Threads from August 2000 to December 2001](#)

Over the whole period from August 2000 until December 2001, the network of genres appearing in threads is shown in the following figure:



Finally, over the whole period from August 2000 until December 2001, the *heterogeneous network of threaded senders and genres appearing in threads* is shown in the following figure:



## References

- Lewis, D.L., Knowles, K.A. (1997). Threading electronic mail: A preliminary study. *Information Processing & Management*, vol. 33, no. 2, pp. 209-217.
- Orlikowski, W., & Yates, J. (1994a). Genre repertoire: Structuring of communicative practices in organizations. *Administrative Science Quarterly*, vol. 39, pp. 541-574.
- Orlikowski, W., & Yates, J. (1994b). Genre repertoire: Norms and forms for work and interaction. MIT Sloan School Working Paper #3671-94. Centre for Coordination Science Technical Report #166.
- Whittaker, S., Terveen, L., Hill, W., & Cherny, L. (1998). The dynamics of mass interaction. *Proceedings of CSCW '98*, pp. 257-264.
- Yates, J., & Orlikowski, W. (1992). Genres of organizational communication: A structural approach to studying communication and media. *Academy of Management Review*, vol. 17, no. 2, pp. 299-326.
- Yates, J., Orlikowski, W., & Okamura, K. (1999a). Explicit and implicit structuring in genres in electronic communication: Reinforcement and change of social interaction. *Organization Science*, vol. 10, no 1, pp. 83-103.
- Yates, J., Orlikowski, W., & Okamura, K. (1999b). Explicit and implicit structuring of genres: Electronic communication in a Japanese R&D organization.