

Relation Between the Complexity of Individuals' Expression and Groups Dynamic in Online Discussion Forums

Luigi Lancieri*

France Telecom R&D, Orange Labs, France

Abstract: This article presents a study concerning the comprehension of interactions in discussion forums. In particular, we approach the role of the complexity of individuals' written expression in the various facets of these interactions, such as the level of activity or the fidelity of the group's members. The study is based on the analysis of traces of activities highlighted by related works in the field of social psychology. Among other things, the results show that the amount of inter-group relations is directly connected to the complexity of individuals' expression.

Keywords: Collective intelligence, expression, interaction, complexity.

1. INTRODUCTION

New communication and information technologies have changed habits in the fields of both personal and professional life. With the spread of computer networks basic usages has partially shifted to mediated usages where the media (device, software, content,...) become an intermediary between humans interactions. In this context of mediated environment, productivity and skills management goes far beyond the ordinary spatiotemporal or organizational frameworks generating emergent phenomena. As examples, let us cite the increase of online business or open-source free software initiatives. In addition to facilities related to the partial lifting of physical constraints, the second fundamental aspect of this finding is linked to the influence of collective behaviour. Though the act of purchasing an asset is often perceived as rather individual in "direct" retail, it becomes essentially collaborative when it is carried out on line. Thanks to collective intelligence, individuals can use free tools, such as shopbots or can consult forums, giving their opinions and benefiting more easily from other users' advice.

In various studies conducted over a period of 5 years, the forums are regularly quoted in 3rd position in the list of most commonly used communication tools, after web browsing and messaging software. These environments are not only a vector of interactions and a source of knowledge, like other media, but they have the distinction of providing a wide opening to the observation of collective behaviour. Collective behaviour can also be seen in other places on the Web, but in a much more indirect way and where special treatment is required for it to be observable. For example, the analysis of the links¹ between websites can identify influences, or shared interests or knowledge, but this information is not obvious by the simple reading of Web pages. By contrast,

regular attendance at a forum quickly gives clues concerning both the opinion leaders and the level of the community's cohesion or stability. In fact, the collective phenomena are central to a forum's life, the forum acting as mirror and memory of the complexity of human interactions. Unfortunately, the reading of forums demands a large investment of time and involvement and would appear to be difficult to approach in a scientific way.

In this context, our research problem is to better understand this complexity through the investigation of basic traces of user activity. Our research is situated in a multidisciplinary framework, at the crossroads of information technology (web-mining, knowledge management, etc.) and human sciences (cognitive science, social psychology, social networks, ethnology, etc.). In these areas, our work can be defined as the study of "collective mediated behaviours".

In this paper, we examine, as an example of collective mediated environments, the interactions in online discussion forums. We study more specifically the influence of users' individual expressiveness on forum groups and their collective activity. This raises several issues related to the individuals' motivations and to the "mechanisms" of a group's birth and development. For example, do individuals with expressive ability really communicate more? Do they send more messages and do they interact with more people? Is the level of group activity a factor of attraction or repulsion for new users or other groups? These considerations go further than merely saying that the acquisition of knowledge can be a major motivating factor in participating in online discussion groups. Knowledge is also the vector of perceptions of other people, even in the context of socially oriented interactions. This is quite clear, for example, in a group of friends where someone boasting about himself can be less frequented. Another important question is also to investigate the differences between face to face and mediated groups.

As a result, we observed that the complexity and length of messages posted by forums users correlates with the extent of their community of interactions. This conclusion is suggested especially by the remark that the groups generating very short or very long (i.e. not "optimal size") messages

*Address correspondence to this author at the France Telecom R&D, Orange Labs, France; E-mail: luigi.lancieri@orange-ftgroup.com

¹This principle is used in search engines (e.g. page rank popularized by Larry Page, co-founder of Google) to assess the value of a page. The links to a given page are considered as votes in favour of this page.

tend to have the greatest cross-posting activity (i.e. post toward other groups). In other words, it appears that individuals able to generate complex messages expand their circle of contacts more easily. Psychologists and sociologists have studied these issues for a long time, but not often from a computer science model point of view. Using the latter approach, the diffuse vision of the complexity of interactions perceived by the mere reading of the forum is clarified by the statistical model, which emerges from the trace analysis. The collaborative ecosystem takes on another dimension that can be characterized and observed more precisely.

This article is built around five other chapters. Chapter 2 describes the object of our study, the online discussion forum, in its historical, technological and sociological context. This chapter gives also examples of online discussion forums organisation and functioning. In chapter 3, we present the results of a survey and we discuss the motivation for a person to join a forum and to express himself in different cases. In Chapter 4 we present a statistical study related to the relationship between the group dynamics and its expressiveness. Both chapter 3 and 4 address the same main questions, evoked previously, but from a different perspective (questionnaire survey in chap 3 and log analysis in chap 4). Chapter 5 gives us the opportunity to put our findings into perspective with related social psychology works. This chapter also gives elements to highlight the differences between faces to face and mediated groups behaviour. Finally, we conclude in Chapter 6 by summarizing our contribution and presenting a few possibilities for future investigations.

2. WHAT ARE WE TALKING ABOUT?

We believe that collective mediated environments, partially reflecting face to face collective behaviour, are akin to complex systems. This implies that the appropriate way of understanding them is to consider them at the same time both as a whole and from a particular aspect of their operation. This section deals with the first part of this general framework: to study forums from a macroscopic point of view.

2.1. Object of Research: Online Forums of Discussion

Functionally, a discussion forum is a repository (centralized or not), for messages posted from a group of users at different locations. All posted messages are visible by the group during a more or less long period. As we will see, this basic concept can be declined in several ways from an organisational or a technical point of view.

We studied non-oral and asynchronous interaction in online groups of intermediate size. We therefore excluded exchanges in small groups (less than fifty individuals), and mass communication. The groups that we observed were big enough to limit the bonds of intimacy and small enough so that the individuals did not fall into total anonymity. Furthermore, the non-oral communication did not exclude multimedia information; but the beginning and structure of the exchange were based on writing, possibly supplemented with audio or video pieces. The asynchronous nature of the exchanges allowed memorization of the interaction and carrying the exchange of contents over long periods of disconnection. Therefore, tools like instant messaging (chat) were ex-

cluded. This is a central feature of our study because, as we shall see, memory has a significant influence on the emergence of collective intelligence and the gathering of activity traces.

Usenet discussion forum is a basic example of the media we studied, even if the principles that we are talking about are transferable to all mediated environments marked by a collective dominance of the group influence compared to that of the individual. This distinction between individual and collective influence is essential. The typical opposite example is that of a Weblog that, although it supports collective activity, is primarily influenced by the Weblog owners who orient the discussion. In this case, the effect of the merging of individual influences, as found in the forums, is less present.

2.2. Online Forums in the Evolution of Collaborative Environments

The concept of mediated forum can be viewed as a heritage of face to face forum as popularized by the ancient Rome. But, the advent of data networks and the associated tools has changed the conditions of non-oral communication in groups. In the "prehistoric" period of the electronic age, this spectrum of interaction, between the interpersonal mode (e.g. snail mail) and mass communication (e.g. newspapers, books), was little occupied. Nevertheless, it was possible, for example, to maintain correspondence within a group of friends or to transmit copies of small print run letters to a circle of insiders (e.g. political groups, hobby clubs, etc.). Small communities (scientists) had the means to disseminate their writings through media like acts of Congress. Although these modes of community communication have demonstrated their usefulness for several centuries, their use has diminished with the flood of electronic tools. Much research work has focused on the new dimensions that these tools have opened up for interaction (reduced transaction periods, ease of discovery of new contacts, new knowledge etc.) [1].

The history of online forums can be traced back to 1979. They are descended directly from the philosophy of the BBS (Bulletin Board System). Usenet (User to User Network) is probably the most prevalent realization of the forum concept. Technically, locally stored messages are relayed between servers *via* NNTP (Network News Transfer Protocol)², which also manages the synchronization between servers. Given the large quantity of messages, all servers do not include all news groups. Furthermore, the network of servers is not fully connected, and certain news groups are not always accessible. This is what justifies cross-posting, which acts as a bridge between news groups.

The messages (posts), originally designed as textual, quickly integrated binary files (images, music, videos, etc.) anticipated the success of file sharing services (Peer to Peer). Forum technology has also evolved over time, changing from both an ergonomic and a usage point of view. Like most specific Internet services, Usenet was federated on HTTP (HyperText Transfer Protocol), the core protocol of

²NNTP (see RFC 1036) is perhaps the dominant protocol in Usenet. This protocol is also linked to the mail protocol (RFC 822, etc.).

the World Wide Web. This tendency increased with tools like phpBB,³ which allowed anyone to operate their own forum *via* a unified user interface type. Soon after, this trend turned out new collaborative services, such as Wikis or Blogs. In 1995, the site "Deja News" (later bought by Google), began to archive Usenet posts. Some records still available on Google News date from 1981⁴. Forums now provide a very stable memory of Internet exchanges.

Before we address the question of interactions, it is interesting to dwell on the relatively formal mode of creating a Usenet group. This process, based on voting to reach a consensus, illustrates, in some respects, the roots of group dynamics. In order to create a group, there must be at least 100 more "yes" votes than "no" votes and more than 2/3 of "yes". By imposing a minimum of 150 supporting votes, this selection process tends to eliminate naturally (i.e. forget, see below) the small groups and the uncontrolled proliferation of forums. A study by J. C. Paolillo, running over a 12 year period (between 1989 and 2001) in the ".comp" hierarchy (subjects related to computers), showed 390 attempts to create groups, involving more than 88000 users and 117000 votes [2]. In his work, the author shows that over this period there was a trend towards increasing the number of positive decisions and that the votes were isolated acts. This meant that most contributors to the creation of a group (76,735 to 88,000) were concerned by that particular group and were not involved or little involved in the creation of another group. This observation lead the author to conclude that, even if it is expressed by collective action, the motivation of the voters is first individual. The author thus revealed that the form of democracy at work in the forums is not a plebiscite, but rather based on plurality, discussion and consensus (see Habermas theory:[3, 4]).

The use of forums, initially popular in public and open environments has been replicated in more closed areas, such as businesses. The case of forums is not isolated; many technologies (search engines, chat, VoIP, etc.) have been transposed in the same way.

2.3. Examples of Online Forums: Discussion Topic, Performances

Online discussion forums are used in a wide range of context and for all imaginable topics of discussion. Even if both issues are linked, it is important to make the difference between the functioning of collective environments and the result of this functioning. If we compare online forums as a tool exploiting collective intelligence, such tool is applied to a problem to solve (the discussion topic) and is expected to provide solutions. For example, it is clear that groups as "misc. Health. alternative" and "comp.os.linux.setup" will not have the same level of results. Indeed, if in most of the cases linux setup problems will find clear solutions, finding health "solutions" is not obvious at all. In this paper, we make the hypothesis that the mains processes of forums functioning (emergences, role of collective memory ...) are mainly identical whatever the problem to solve and that the

level of possible solutions are mainly linked with the problem itself and the competences and the motivations of the humans involved in the discussion. As a basic comparison, it is like a writer (novelist or not) who has to use a pencil (tool) to write simple text or an artwork (problem to solve). The functioning of the pencil is the same in both cases. In our context, the functioning of groups is much more complex than that of a pencil but we think that most of the rules involved are invariants. Nevertheless, one of the difficulties we have to tackle with is that complex systems are based on interdependency and thus cannot be segmented in clear parts.

Starting from this hypothesis we focus on online forums functioning but in order to distinguish each part of this complex system (problem, tool, solution), we give here after some examples of use of online forums.

The use of forums as productivity tools in enterprises has had mixed success depending heavily on the groups and the goal management [5]. The positive experiences, like those at DEC, Interquisa Canada or in the hospital sector, refer clearly to the solution of a specific problem [6, 7]. In this case, the forum of discussion is the medium of exchange for a particular community of practice [8, 9]. Not only does the problem to solve have to be clear, but the forum must be a part of the managerial means, in the sense that it helps to solve organizational constraints at a lower cost (e.g. communication between different teams at Interquisa corp., or with practitioners on remote hospital sites, etc.). Setting up an in-company forum, whose only objective is to encourage exchanges between people, but with no clear goal, often leads to failure. This need to channel collective action towards clear goals is confirmed by remarks made by several authors who highlight the role of the mediator in the success of a forum [5, 10]. In the case of forums without mediators, the most often mentioned causes of failure are the too low or too high number of messages. Observation of the behaviour of conventional teams (i.e. non-mediated) in decentralized companies confirms those observations made in mediated groups. For example, on examining the influence of the participation of employees depending on the size of project teams, Bradner *et al* observed that participation is higher in small groups compared to large ones [11]. The group coherence that is obtained by a strong initial and consensual motivation, by the definition of clear shared goals or by a mediator who drives the group, contribute to facilitating the process of individual decision, which feeds back on collective relation and maintains the involvement of the group members (see also section 4.2).

Another kind of forum use linked with companies, but outside their control, relates to the kind of activism or criticism posted by individuals in connection with the feeling of being misunderstood (power relations badly lived, etc.) For example, we can cite the case of company-dedicated forums organized by boursorama.com (online trader web site). The primary purpose of these forums is to allow shareholders to exchange ideas on the merits of buying or selling a company's shares. Regular observation of posts made it clear that some of these shareholders are also company members, who give details of their company's organization or its operations

³PhpBB website: <http://www.phpbb.com/> [accessed march, 3, 2008].

⁴Net.news: archive group net.news: <http://groups.google.com/group/net.news/about> [accessed march, 3, 2008].

to support their opinions on the value of the business. On some occasions, these forums are transformed into sometimes violent denunciations of management excesses or inconsistencies. This "safety valve" or regulation function, through free expression "outside" of the company, and under cover of anonymity, is hardly conceivable internally.

3. SURVEY: EXPRESSIVITY, MOTIVATION AND GROUPS DYNAMICS

Motivation is at the root of individual behaviour. It is what drives individuals to join a group, to express themselves and to play a more or less driving role in the creation of a network of relationships. A discussion on motivation is therefore useful to confirm and evaluate the meaning of certain causal relationships derived from the analysis of traces. To enlighten our discussion, we carried out a survey to get an idea about the motivations and practices of forum participants.

3.1. Survey Methodology and Respondents Profiles

The method that we used to obtain the users' points of views was an online survey based on a semi-open MCQ (Multiple Choice Questionnaire). Unlike the traditional MCQ (closed), which offers questions and set answer opportunities, the semi-open MCQ adds the opportunity to make comments, encouraging more nuanced feedback. It also helps in the identification of poorly phrased or biased questions. The mere observation of the number of comments made on a question reveals any distortions. This approach has the advantage of enabling a larger mass of people to be interviewed and of permitting an automated statistical treatment of the data. You may wish to refer to [12] for a study on the use of questionnaires in mediated environments. This poll is not meant to be exhaustive but simply to identify main opinion trends, so supplementing the trace analysis, which is then put into perspective in terms of social psychology.

The first step was to identify potential questions to ask. Having in mind the first raw results of our traces analysis, we discussed with a small group (6) of regular forums users and established the questionnaire. For the reader convenience we present in the annexe section the major part questions and average responses of this survey. The questionnaire was then put on a website in order make easy the typing, the collect and the management of responses (database,...).

62 individuals, of whom 81% were male, completed the survey. The average age was 28 years. Broadly speaking, the population has a higher level of education (63% at least Master of Science) and has a very good knowledge of ICT tools from their training period or their professional activity in this area. In addition, 63% of respondents indicated that their hobbies are also connected with computers. Furthermore, as part of their core business, 97% of the population accesses Internet daily, 74% for more than 2 hours per day. These proportions are slightly lower for personal activity (86% and 60%). In addition, all of those polled consult their mail daily and 81% of them several times a day. Furthermore, 60% of them have a Web page. Regarding the use of forums, 84% of the population has at least one year of experience (70% over

3 years), 94% of people consulting at least weekly (daily, 34%).

3.2. Expressiveness and Interactions in Online Groups

As confirmed by numerous studies, the majority of forum users are rather passive. Indeed, 44% consult but never post, while 17% post, but only in response, without initiating discussion threads. Only 30% are leaders, initiating discussion threads, but for less than 10% of their messages, while 8% initiate threads with more than 10% of their messages. Users familiar with instant communication (31%) are more likely to post than others. The post length is rather of moderate size, in the order of 7 lines (45% between 5 and 10 lines, 25% less than 5 lines and 12% more than 10 lines). It can also be seen that users making cross-posts (17%) tend to send the longest messages (43% longer on average). We believe that the cross-posting highlights the expansion of the radius of relationship. These results suggest that individuals with greater expressiveness (writing longer messages) tend to widen their circle of relationships. As we shall see below, this conclusion is confirmed by the statistical analysis of activity traces.

3.3. Satisfaction, Motivation and Groups Dynamic

Qualitative factors, such as the level of user satisfaction, are difficult to assess without opinion polls or direct interviews. The survey suggests that forum consultation reflects users' strong intentionality. In fact, 79% of the population consults for a specific reason, such as to find the solution to a problem, whether for professional use (65%) or for private matters (64%). This motivation is also significant in the rate of users (60%) indicating that they rarely surf the forums at random or in stroll mode, as they might do on the Web in general. We have seen in the last section that this kind of use makes sense in light of the factors of success or failure of forums in corporate contexts. This intentionality is also confirmed by the fact that few users (10%) do not have a regular group, while the majority tends to settle in a stable relationship with a rather limited number of groups (45%: 2 to 3 groups and 24%: 4 to 10 groups). Indeed, D. Maltz, remarked in his study that half of the users subscribing to fewer than 20 groups manage, in 90% of cases, to follow them [13]. The fraction of users who subscribe to more than 100 groups is 8%, and only 1% of them manage to follow their groups.

As might be expected, given the profile of the population, 73% perceive the benefits of their use of forums as satisfactory to very good. This use also changes the users' habits or decisions (frequently for 25%, sometimes for 68%). We also observed that satisfied users tend to post more. To better assess the reasons for satisfaction, we asked the respondents to rank in order of significance the reasons that could drive them to leave a forum after only a few visits. The first reasons cited are clearly related to a lack of supply of knowledge (not enough relevant messages, chats too basic...). Then there are reasons related to the interactions (the lack of answers to questions, groups closed on themselves or infiltrated, banishment or abusive censorship). In the same way, we asked them to rank the reasons that drive the users to

remain on the forum (fidelity). The responses confirm the previous evaluation, interest in terms of knowledge arriving in first position, then sympathy, social contacts, and the feeling of being useful to the group.

3.4. Discussion on the Relationship Between Motivation and Groups Dynamic

We can see that there are various motivations for an individual to participate in a forum. We support the idea that one of the common roots among all these motivations is linked to the "value perception" associated with the constraints of cognitive load. Basically, this means that individuals get involved in groups if they feel that the difference between the necessary "effort" and the expected "profit" will be in their favour.

Broadly speaking, the authors agree on the fact that there are 3 main reasons which push an individual to join a group [14]. The utility model postulates that the interest of membership is to meet a certain need. The model of social cohesion maintains that the membership decision is made on the basis of interpersonal attraction. The model of social identification is based on a perception of identity and not on emotional criteria as in the model of social cohesion. From a certain angle, these models involve a perception of power or capacity (social ability, competence, etc.) and, more generally, of value, which gives rise to phenomena of categorization and discrimination, leading to the formation of groups. Ng proposed an experimental method to highlight the relationship between "value perception" and group formation [15]. As a first step, the author observed discrimination when each member of 2 groups had to divide money between all the members of the 2 groups. In a second stage, the experimenter made each member believe that he was the only one to decide what amount of money had to be given to the members of his own group and to those of the other group. In this case, the discrimination in favour of his own group was practically absent (see also [51, 52]). This means that the feeling⁵ that his group (and therefore himself) may be disadvantaged pushes an individual to selfishness and discrimination (i.e. restores the balance of earnings). The irrationality or the subjectivity of this feeling often leads to a biased evaluation of the role of group members, stimulating injustices and conflicts.

In utility models, this concept of value may be achieved through gaining greater skill or knowledge. From an evolutionary perspective, the link between knowledge acquisition, development and research of identity makes sense. At the edge of reasoning, the logic is to exist (i.e. to be acknowledged) in order to survive. This quest for individual appreciation can also motivate a strategy of differentiation, a search for originality or even for inconsistency. Some of these mechanisms are highlighted in the "open source" community, where the status of experts is valued [9, 16]. The authors structure these groups into 8 levels of non-fixed, weakly hierarchical roles, ranging from project manager to passive members. Most of the time, passive members ac-

counted for the majority (99% in the Apache⁶ Project). We can see the relation between these environments and discussion forums, where only a minority is really active. However, contrary to what some may think, these passive members are very far from being useless because they contribute to valorising the most active members by their mere presence, a bit like spectators clapping in a theatre.

The forum members' activity level does of course have a link with motivation and the theory of value, but it also introduces many questions regarding the interaction, the expression between individuals and the role of the media (the Forum). In the same way that two individuals will certainly tend to limit their contact if they have difficulty understanding each other, forums could alter expression, as well as the amount and the structure of interactions. Having discussed forum dynamics in a general way, we now approach these issues in terms of trace analysis.

4. FROM TRACES TO MODELS: EXPRESSIVITY, INTERACTIONS AND GROUPS DYNAMIC

The traces refer to data generated by user action. Specific analysis of the traces then gives us information about various aspects of user activity (e.g. user #1 posted a message at 10 pm). This approach, taken from a systemic point of view, aims at combining quantitative and socio-cognitive models. It allows the observation and understanding of short-lived, cross-related, or hidden phenomena, which are difficult to catch by direct observation of these virtual milieus. If mediated environments are very rich in digital traces, it remains tough firstly, to identify and collect relevant data, then, to choose the direction of the analysis, and, finally, to interpret and apply the results. Specific difficulties appear at each of these stages and often require consideration of the balance between the effort needed to obtain data of good quality and the benefit of the investigation. The informative level of traces, for example, is sometimes limited by the need to take into account ethical, legal or simply technical constraints. However, as in most areas of science, traces analysis is often a key method for understanding phenomena. The purpose of the socio-metric approach is to analyze quantitative traces of interactions in order to extract meaning by analogy to cognitive or social models. This principle has been addressed in many works [17-21].

4.1. Methodology and Data Descriptions

In the study that follows, we analysed Netscan⁷ traces of 1800 French-speaking Usenet groups over a period of one year (April 2004 to April 2005). For each of these 1800 groups we obtained the characteristics of user activity (number of messages, number of posters etc.) over the one-year period. These groups cover all kinds of discussions topics in French and are not specially related, except for hazard, with the newsgroup evoked in the questionnaire survey (section 3). As an example, the following table gives a view of the

⁶The open source Apache project produced a widely used Web server as well known as Linux is the operating system field.

⁷Netscan: Usenet social accounting search engine from Microsoft; <http://netscan.research.microsoft.com/> [accessed march, 3, 2008].

⁵ See the theory of the bounded rationality of Herbert Simon.

raw data as provided by Netscan for 10 groups of the 1800 groups.

fr.soc.politique	104415	3286	898	93090	2874	4383	31	28564	356
microsoft.public.fr.windowsxp	29169	4123	1247	23267	2661	827	30	932	112
fr.soc.religion	26945	1245	399	24745	1110	1024	33	12615	131
fr.rec.moto	25877	1016	492	24429	799	258	17	291	64
fr.sci.philo	23726	753	231	22120	628	405	30	3778	81
fr.soc.feminisme	19725	1345	211	17635	1285	1582	26	7776	101
fr.lettres.langue.francaise	18158	611	232	16980	482	305	19	720	117
fr.rec.photo.numerique	15107	1534	645	13677	1249	304	22	833	79
fr.rec.bricolage	14167	1798	700	11921	1301	357	23	371	51
fr.rec.aviation	12816	786	314	11288	645	620	21	1070	77

Each record contains 10 values which are the following meaning:

- Name of the newsgroup:
- # Posts (per group)
- # Posters: Contributors of at least one message
- # Loyal contributors (i.e. having also contributed during the previous period)
- # Responses to a previous message
- # Responders to a message
- # Messages without responses
- # Average lines per message containing at most 40 lines (see below)
- # Cross-post out (i.e. from a group toward other groups)
- # Cross-post in (i.e. from other groups toward a group)

The following table gives some key statistics regarding these characteristics:

	Posts	Posters	Loyal	Reponses	Responders	Without Resp.	Lines	Cross out	Cross in
Min	1	1	0	0	0	0	1	0	0
Max	104415	4123	1247	93090	2874	4383	40	28564	356
Avge.	548,51	76,98	23,14	451,34	47,55	42,28	19,10	66,30	15,51
Total	1005421	141108	42416	827314	87165	77498	35012	121535	28437

In this study we made an assumption that will be discussed ahead. We believe that the ability of an individual to process complex messages can be evaluated quantitatively by the size and number of messages he sends. Although this vision seems simplistic because there are, as we shall see, other criteria of linguistic complexity (e.g. lexical redundancy), we believe that this hypothesis is true, statistically speaking. Moreover, the great advantage of trace analysis is being able to work on a large data set. In our case, we used an analysis of more than one million posts, which allows reasonable statistical conclusions.

4.2. Relationship Between Individual Expressivity, Activity and the Growth of Groups

Several questions arise regarding forum life. For example, did group size have an influence on activity and dynamics? Various authors have studied this issue and have given apparently contradictory answers. Some studies in the context of collaborative work report that an increase of group size tends to reduce individual participation [11, 22, 23]. In a more detailed way, Avouris *et al* remark that larger groups produce better results and generate greater activity, but that this activity is less homogeneously distributed between different members of the group [24]. Valacich *et al* reported an even more precise finding, observing that the activity rate per member increases with the size of the group when it is composed of members with diverse skills, whereas it decreases when the group is homogeneous [25].

These studies suggest that the background and the composition of the group strongly influence the interdependence between the size of the group, its effectiveness and the individual participation. In a company, where groups are typically rather homogeneous (same objectives, skills,) because the context is very structured, large groups are rather a brake on productivity. Hence, a control (leadership, inciting contributors to participate, etc.) is often essential in these groups. In less structured environments, an increase in productivity can happen as the size of the group grows, *via* a form of creativity generated by the meeting and even the clash of diverse skills⁸.

So, what did we observe in Usenet news groups that are relatively poorly structured (little or no control, freedom of presence, speech, etc.)? We first observed that large groups tend to generate the most loyal users. This is shown by the following figure, which represents the percentage of users posting for a certain period, who had posted in the previous period, as a function of the number of posters⁹ in the group.

We also observed that individuals participating in high-traffic groups tend to be more expressive. This is shown in the Fig. (1), where we see that groups with many posters tend to produce longer messages.

On this subject, our findings contradict the results of Q. Jones and S. Whittaker, who indicated that the individuals involved in active groups (high number of messages or post-

⁸See the role of chance and random phenomena in the evolutionary perspective.

⁹Limited to 600 here. Beyond this limit, the number of elements in each class is unrepresentative. Indeed between 1 and 100 posters, there are 1500 items when there are only 5 between 600 and 700 posters.

ers) tend to post shorter messages [26, 27]. The main difference between these works and our approach is the maximum size of messages taken into account. We excluded messages greater than 40 lines, whereas Whittaker and Jones included messages up to 2000 lines. This disparity raises several issues. First of all, we may wonder what the realistic length limit of a message is and why the inclusion of very large messages changes the conclusion so radically. Indeed, if we reproduce the methods of Jones / Whittaker (taking into account the very long messages) on our data set, we find a decrease in the number of posters (ord) depending on the number of lines (abs) (see Fig. (3)).

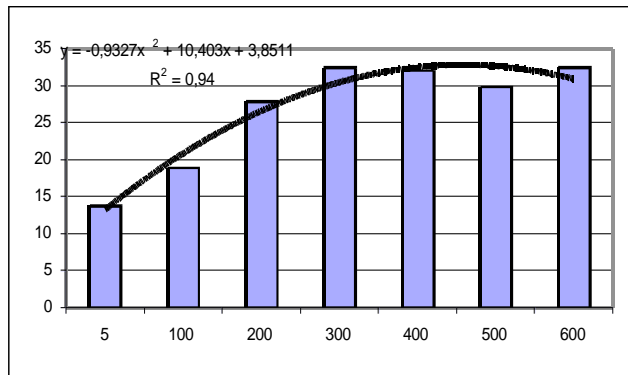


Fig. (1). Average percentage of loyal posters (per poster) (ord) as a function of the number of posters in each group (abs).

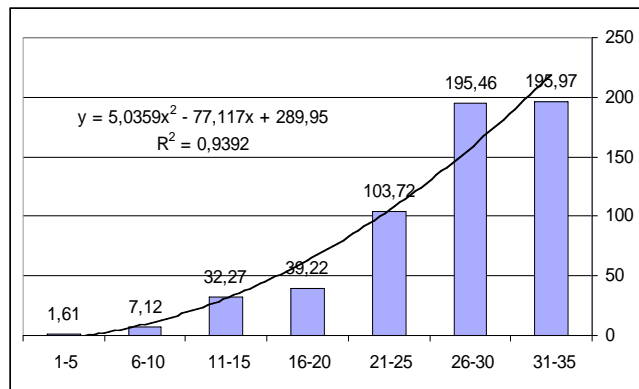


Fig. (2). Average number of habitual posters per group compared to the average number of lines per message.

We believe that the Whittaker / Jones approach is not realistic. Indeed, as anyone with any experience can testify, it is quite logical to assume that messages entered at the keyboard are rather short, especially on tools such forums. This finding is confirmed by statistics from our survey (7 to 15 lines). The analysis of traces is even more categorical and more accurate. Indeed, the Fig. (4) histogram shows that out of 1800 groups, 1465 (i.e. 81%) produce messages of less than 25 lines. In fact, very long messages are often copy-pastes of text not especially written for the occasion (e.g. FAQ¹⁰, external documentation, encoded binary files etc.). In

addition, the two studies have different sizes of experimental sets (578 groups for 5 months against 1800 groups over one year for our study). In fact, our chart (Fig. 2) corresponds to a zoom-in of the beginning of Fig. (3).

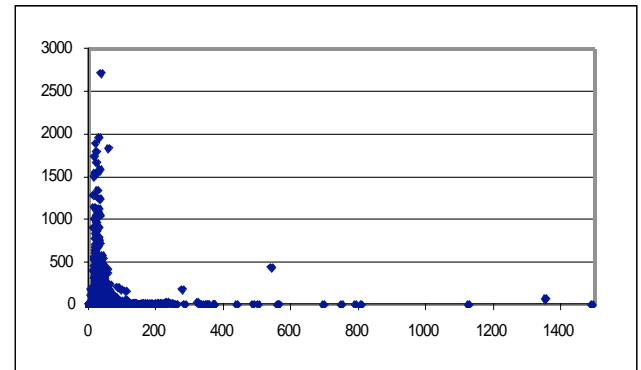


Fig. (3). Number of ordinary posters per group depending on the number of lines per message following the method of Jones and Whittaker.

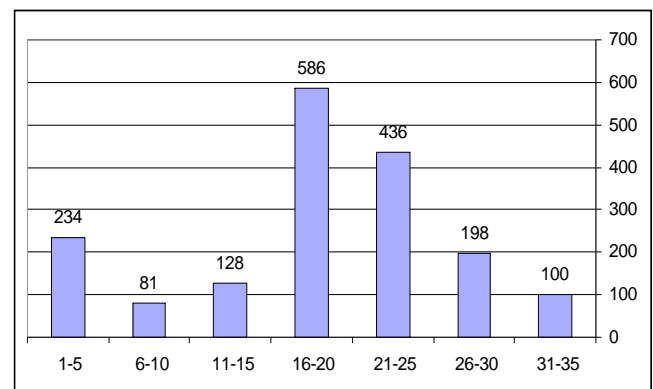


Fig. (4). Number of groups (ord) producing messages of given average length (abs).

In terms of motivation, the fact that active groups produce longer messages can also be explained by normative group pressure. Placed in front of a large amount of information, the user tends to replicate its context and also produce long messages. It is a bit like in a classroom, when the teacher asks students to make a presentation in front of their peers, without specifying the length of the speech. The first students may have a tendency to make short presentations in order not to expose themselves, but gradually, subsequent students will tend to make presentations at least as long as their colleagues', so as not to lose face (theory of value). Jones also reported this gradual increase in expressivity in his next study, which showed that the messages starting a thread tended to be shorter than the following ones [27].

We may also wonder whether the complexity of individual expression correlates with the activity of the group. The quantitative study of the posts seems to show that this is indeed the case, as shown in the Fig. (5), which represents the average number of posts per group (reduced to the poster) based on the average number of lines per post.

¹⁰ Frequently Asked Questions.

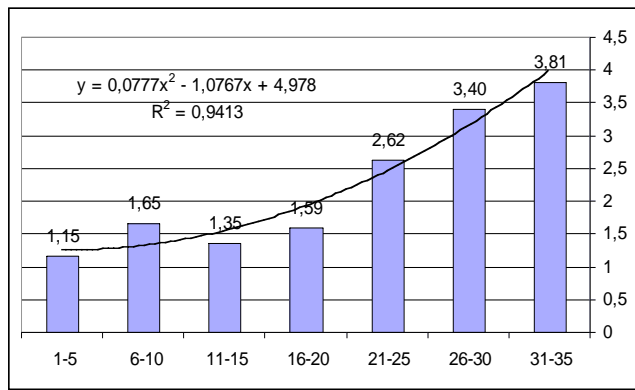


Fig. (5). Number of posts by poster compared to the average number of lines per message.

4.3. Relationship Between Expressiveness and Interactions Between Groups

As in direct social relationships, mediated communities show interactions between groups. This can be, for example, a cross relation between professional teams having to participate in large corporate projects. In this part of our study, we try to describe these crossed interactions in forums by observing cross-posting according to the size of groups and their expressiveness.

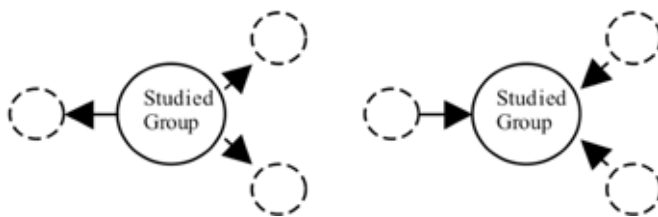


Fig. (6). Outgoing and incoming cross-posting.

We notice, firstly, that the groups' population tends to influence the number of interactions between groups. This is brought out by the chart below, which shows the number of cross-posts per poster as a function of the number of posters in the group.

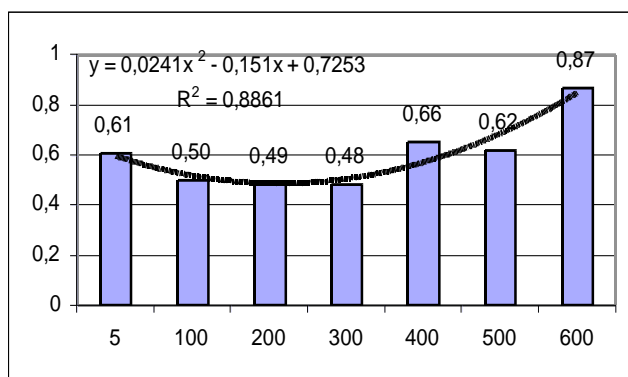


Fig. (7). Number of cross-posts (per poster) depending on the number of posters.

From a methodological point of view, let us note the clumsiness of some authors, who assess the absolute number of cross-posts in relation to the number of posters, pointing out that this quantity, grows almost exponentially (from 6 to 474 cross-posts for 100 to 600 posters). In reality, it is obvi-

ous that there will be more cross-posts in absolute terms, if you increase the number of posters, which makes the result insignificant. By reporting the number of cross-posts per poster (normalization), the result also increases, but in a more subtle way (Fig. 7). This does not change the conclusion but makes it much more realistic.

If the social theory - which we will develop further on - can be transposed to mediated groups, we should find a link between the level of cross posting and the size of messages. The following graph shows that this is indeed the case. We see that groups producing long or very short messages tend to contact other groups more.

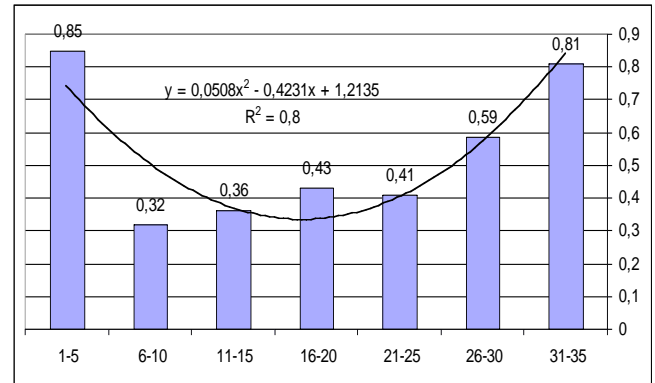


Fig. (8). Average of outgoing cross-posts per group (per poster) (ord) relative to the number of lines per message (abs).

In the same way, we observe that the average message size in groups which often use cross-posting is 63% higher than that of those doing no cross-posting (19.9 vs 12.6 lines). This trend is also confirmed by the survey (on average 8 lines vs 5.6). The cross-posters, however, are still a minority. Out of the 62 users of our survey, only 11 (20%) cross-posted (43 did not). Note that this ratio is 12% in the trace analysis. The symmetry between long and short messages is less clear for ingoing cross-posting (Fig. 9). One possible reason for this lack of symmetry is probably a methodological bias. Indeed, logically, we should normalize this plot like the previous one. This was not possible because, as opposed to cross-posting out, where the number of posters was known (the posters of the studied group), for ingoing cross posts, the number of posters has no real sense, since there can potentially be an infinite quantity of posters outside of the group (all posters of all external groups). To try to reduce this bias, we computed the cross-post-in ratio relative to the total number of incoming posts. In any case, in the following plot we can observe a clear relationship between message length and the cross-posting-in ratio.

The Figs. (8,9) indicate that the groups producing intermediate messages sizes, (i.e. of "optimal"¹¹ size) tend to look inward, if we take the low level of cross-posting as an indication, and also appear to be contacted little by other groups.

¹¹In anticipation of explanations given in the next section, we believe that short messages are difficult to interpret because they are too succinct, implying the readers' good understanding of the context. In the same way, longer messages are complex because of their structure. The optimal size is intermediate, in the order of about fifteen lines in our measurements.

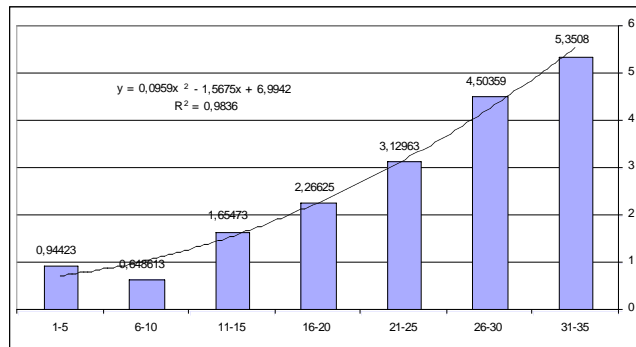


Fig. (9). Average ingoing cross-posts per group (according to total number of posters of incoming posts) (ord x1000) as a function of the number of lines per message (abs).

4.4. Discussion

Taken together, these results can be summarized in the following diagram, which compares the relationship between group characteristics and those of individuals.

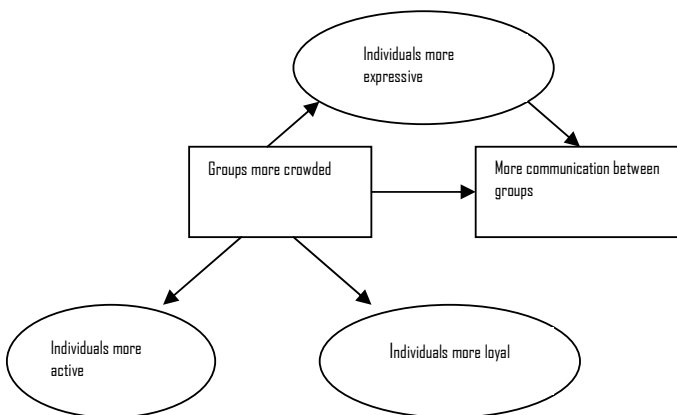


Fig. (10). Relationship between group and individual characteristics.

This figure highlights, to some extent, the interdependencies between the functioning of the collective and the individuals who compose it. The relationships mentioned are, of course, only partial, since the studied system is complex and involves many different parameters. Some relationships that appear unidirectional are perhaps bidirectional, probably being linked to a mutual influence. For example, if a popular group causes more inter-group communications, it is likely that the resulting publicity in other groups causes feedback to the original group. These issues are open and certainly need specific investigation.

After having observed the collective phenomena from a statistical point of view, in the following section we go into the reasons underlying the mechanisms that we are talking about.

5. ROLE OF EXPRESSION IN GROUPS CONSTRUCTION AND DYNAMIC

Studies of the influence of individual expression on the formation of groups are not entirely new, and researchers in human sciences have already considered such relationships. We highlight the fact that such investigation is made more understandable, thanks to new information systems that favour the collection, management and operation of a large

quantity of traces witnessing human interactions. Not only are large populations involved, but the study can be made in an almost transparent way on wide space - time scales. This allows us to view the analysis and modelling of social phenomena in a new light, and in particular, to identify laws or emerging changes, which may not be apparent without the observation of such a large amount of data.

This part of our study aims to clarify our vision of the role of the complexity of individual expression on collective dynamics, through related work from multidisciplinary fields. In Part 4 we have taken into account only message length to study this complexity. We show how this can be justified, in particular *via* various reduction mechanisms operated through the "mediator", regardless of the nature of that mediator (Internet, the forum device, the group, the language, etc.). "Mediator" refers here to the extended role of intermediary played by the media in its generalized concept. The mediator is more than a basic transmission line, playing an active part as a go-between in human interactions.

Above all, let us note that an online forum does not reflect an ordinary group, but, to be precise, a group whose interactions are mediated. The action of the medium in question is not neutral, and before discussing the relationship between expression and interaction, we must take into account the special role of such media.

5.1. The Role of the Media in Expression

From this point of view, the modality of expression refers to those media which vehicle "thought" directly, such as the voice, symbols, writing, but also books and other physical or electronic media. Each of these media has a more or less reductive effect on the original thought or intent. This effect of reducing, akin to that of forgetting, is especially sensitive through human-computer interfaces (command line, mouse-windows interface, multimedia, and so on.), which translates to a greater or lesser smoothness and completeness of the interaction capabilities desired by the software's author [1]. The electronic forum, as a medium, is no exception to this rule.

In this context, the work of Q. Jones showed that the type of technology used has an impact on the stability of individuals' contributions over several consecutive periods [28]. The authors reached this conclusion by comparing this stability in the forum mode and in the list of diffusion mode. In the latter mode, the contributions are more stable (50% of users who posted in a given month posted the following month, against 11% for forums¹²). In a broader context, C. Jensen showed that the contribution of users to a group activity, such as an online game, was higher when the mode of communication was more evolved (voice, speech synthesis, text chat, no communications¹³) [29]. The author points out

¹² It may be assumed that the reasons for this difference are linked to the larger involvement of individuals in the mailing lists. In fact, this medium is often more "confidential" than forums and requires specific action prior to subscription and less anonymity, which may discourage less motivated users.

¹³ In the context of one-way communication, lurking behaviour is sometimes abusively translated as "voyeurism", but in fact involves passive consultation (few or no posts). The study shows that lurkers' dominant motivation is

that artificial speech generates more contributions than the equivalent text (chat), but less than human conversation. He attributed this relationship to the level of trust; trust being established more easily with realistic forms of communication. In the same way, the work of I.E. Morley shows that in situations of limited communication (e.g. telephone), compromise is more difficult than in face-to-face negotiations, and that it is usually the power relations that existed before the communication that gain in strength [30, 31]. The role of the media in the phenomena of influence and conformity was treated by S. Cilesic, who compared face-to-face opinion polls with online surveys, studying contributors' responses to 4 versions of the same poll [32]. The result suggests that online polls are rather reliable tools. The author shows that some phenomena of bias and conformance that are well known in real social relations, apply differently online. For example, online, individuals tend to maintain their original opinion, resisting "manipulation", even in cases where the author of the survey is not neutral and suggests his opinion. In the same manner, M.R. Parks has studied the emergence of real relationships from virtual contacts on the forums [33]. The author reports a survey of 176 users of newsgroups (contacted by their email, visible in the posts; 176 responses over 528 requests). 60.7% (72% women, 54% men) of respondents reported having formed a personal relationship with someone they met online. This frequency is correlated strongly with that of the participation in focus groups. In 69% of cases, the virtual relationship lasted less than a year before continuing in the real world. In 30% of the cases, the frequency of the virtual relation was 3 to 4 times a week, and for 50%, once a week.

According to the results of our study, all of this work confirms that the media¹⁴ are changing the interactions between individuals and that expressiveness tends to have a positive influence on making contact and group construction.

5.2. The Group, Another Mediator of Expression

Like the individual or the media, the group, as a homogeneous entity, intervenes in different ways as a mediator of expression. We analyze this role of active filter *via* group productions but also in terms of collective influence.

The most visible sign of group action is certainly its written productions. In his work, Whittaker studied the influence of collective activity (number of FAQs, quantity of cross-posting, thread size, etc.). These elements of expression are seen as "collective" factors that increase the cohesion of the group (common ground) [26]. The author observes that, regardless of the number of messages, the group's activity is mainly the result of a core of a very active minority. Some authors, such as F.B. Viagas, use individual contribution intensity as a measure of the individuals' aptitude [34]. His studies have shown that, in forums, these indicators of group

activity influence participants' choice of which messages to read, creating a feedback loop on the group dynamics. Viagas also measured the level of interactivity in the group by studying the variation in size of discussion threads. Such data is valuable to assess the depth of group shared memory, which is directly involved in the group dynamics [1]. He observed that a thread contains, on average, 1.8 messages, while 33% of threads contain more than 2 messages, this corresponding to a successful initiation of conversation. The initial message with no follow-up communications (single message in a thread), corresponds to 44% of messages. The author also calculated that 54% of groups produce FAQs, viewed as productions of a more complete and structured knowledge.

Regardless of productions, the group itself has an impact different from that of the individual. For example, G.M. Stephenson notes that during collective bargaining, compromises are more difficult to develop than in an individual-to-individual situation [35]. According to the author, the discussion can become more difficult when it is collective, sometimes with inflexible positions, because there is less personal risk. We believe that, in addition, collective expression is by nature less consistent and more disorderly than individual expression. For example, interlocutors sharing the same point of view can contradict each other, due to a different level of knowledge on the subject, blurring the "message of their camp" rather than clarifying it. Achieving consensus or processes of agreement, not applicable to an individual, are sources of noise or misunderstanding in groups. The process of convergence of group thinking is also longer and less reactive.

Nevertheless, collective expression also has its advantages; even if it dilutes the common understanding of individual expression, it is more stable and reduces the manifestation of differences, thus lessening the scope of conflicts. W. Doise transcribes Ross' ideas that: "A society ravaged by dozens of oppositions, whose lines of division run in many directions, may know less violence and erupt slower than another society divided by a single line of division. Each new division helps to diminish a general rift so that it may be said that a society is sewn into a whole by its internal conflicts" [31]. According to Doise, purposes or projects establish the belonging to a particular social group, in the same way that such membership can facilitate adherence to a common project or to the experience of a specific fate. These findings confirm our remarks regarding the relation between the success of a forum and the clear objective of the group.

Among the key elements of the formation of these groups, we find social representation. Broadly speaking, social representation is rooted in language and forms "collective reconstructions of reality" [36]. Ranging between expression and action, social representation influences the behaviour of individuals and groups, which in turn influence social representation. Consequently, there is always a little of the group in individual expression. This also means that, in some ways, social representation is involved in the development of identity and in acceptance of the individual by the group (See P. Moscovici's work [37]).

one of utility, based on the acquisition of knowledge (of the functioning of the group or the content of the exchanges), but they choose to maintain their distance so as not to expose themselves (reputation to maintain, lack of social ability, avoidance of spam and solicitations, and so on.).

¹⁴ It would probably be more appropriate to speak of a communication channel, which, in the terminology of social psychology, refers to something broader than just a support (telephone, writing, etc.) and taking into account the physical aspects of the interaction (gestures, postures, and so on.).

5.3. Complexity and Quantity of Expression

Clearly, the media, whether artificial or human, individual or collective, act as a filter and reduce the expression of thoughts and reality. We have seen that the mechanisms underlying this reduction are complex in nature. In discussion forums, where the mode of expression is written words, this complexity can be broken down in terms of lexical redundancy or amount of information (see Shannon information theory).

A lot of work, especially in the language theory field, has for a long time analyzed the influences of text complexity and structure on understanding. These works show that text length, linked with the lexical redundancy level, is directly related to the text comprehensibility. Furthermore, the literature shows us that the memory is the key element of this relation. The PISA report, for example, classes the understanding of text in 5 levels and shows that the most difficult levels (4: understanding and 5: deep understanding) involving longer texts are achieved respectively by only 20% and 8% of 15-year old children in the OECD countries [38]. Lexical redundancy is a key element of this understanding, and is intimately linked to text length. The analysis of the lexicon shows, in fact, that the level of redundancy and of lexical span tends to increase with the length of a text [39-41]. Various studies have shown, moreover, that a too low or too high lexical redundancy, consistent with a too short or too long text, impedes understanding [42, 43]. Jean Veronis¹⁵ shows in a very illustrative way that texts difficult to read, such as the EU constitution, contain a high level of lexical redundancies that he evaluates in a simple manner by the text compression ratio (the newspaper "Le Monde" or "L'Humanité": 61%, the journal "Science": 63%, the EU constitution: 75%). The effect of redundancy on understanding can be perceived differently from one author to another [44]. Recent work, however, regularly evokes the theory of cognitive load, introduced by John Sweller and based on the 70s information theories [45, 46]. Broadly speaking, the idea is that the amount and the structure of information that people need to process, and the limitations of human working memory may limit understanding. The importance of memory in this process has been highlighted by O. Le Bohec who shows, through a memory recall test, that documents with low and high redundancy tend to be less well remembered than documents with an intermediate rate of redundancy [43]. This trend is reversed if the individuals are allowed to take notes on repeating the same test (i.e. external Memory).

On a more general level, in his work on social cohesion, M.R.A Chance suggested, following B. Berstein, that there is a strong link between verbal code complexity (vocabulary span or richness) and the motivation for setting up or joining groups with a strong social unity and clear lines of authority [47, 48] (see also [50]). In the case of rudimentary expression, the main concern of individuals in a group is to maintain the stability of social relations, which reduces their motivation to contact other groups. Conversely, a rich language

able to convey objective communication promotes social relations beyond those closed groups, where a rigid system imposes a determined role on each individual. B. Berstein relates the example of the link between simplicity of language and communitarian withdrawal among Hooligans. One can also observe a comparable tendency in sectarian groups, when they reject without discussion anything that contradicts the group ideology. This creates the followers' addiction to the group, even if open (i.e. diversity of) expression is not allowed.

This relationship between the richness of expression and social interaction is also reciprocal, as shown by the anthropologist M. Jousse [49]. Indeed, as soon as an individual is less focused on social constraints, he tends to abandon traditional means of communication and returns to primary forms of expression. This condition is marked by gestures and onomatopoeia that provide a direct expression of an experience. P. Moscovici shows also that, in small groups, the verbal code tends towards poorer language, reducing the vocabulary and augmenting the lexical redundancy [37]. It is in this context that jargon, or more generally, private languages appear. In fact, some authors (Walloon, Werner, Kaplan, Moscovici) have distinguished 2 extremes in expression: a direct, primitive and internal form, marked by gesture and oral production, poorly communicable; an adult, external form, marked by writing and organization, more easily communicable. According to all the authors, all forms of expression exist somewhere on a continuum between these 2 extremes. The position on this continuum reveals the role of the individual in the collective. Individuals, who express themselves "optimally" (without superfluity) and who cannot or do not wish to bear a significant cognitive load, tend to want to remain isolated and control their circle of contacts.

6. CONCLUSION

In our study, after having considered the complexity of the forum environment from a general point of view, we have approached the relationship between the expressiveness of individuals and collective behaviour in an empirical way. In summary, 2 contributions of this study can be highlighted.

First, we support a methodology based on activity trace analysis, putting forward the interest of this technique as a complement to the socio-cognitive sciences. The benefit of this approach in the understanding of complex phenomena is to illustrate and supplement theoretical reflections thanks to purely quantitative data. It is well known that group' motivations, behaviour or performances obey to multiples inner or outer influences. This complexity imposes a mixed approach, *via* the study of both inter-group relations and the internal changes observable through individual behaviour. This cross investigation is hardly possible through direct observation. We support the idea that the instrumentation based on the collection and analysis of traces could play a leading role in better understanding collective phenomena. The use of these tools is just beginning to emerge, thanks to the widespread use of personal computers and the possibilities for individual communication. All things considered, one could compare these benefits to those conferred by investigative tools (microscope, telescope, etc.) in the under-

¹⁵ Jean Veronis Weblog: Jean Veronis, Language technologies, 2005 <http://aixtal.blogspot.com/2005/04/texte-la-constitution-europenne-pour.html> [accessed March, 3, 2008].

standing of the physical world. At their origin, and even now, these tools only allow lightening a small part of the universe complexity. Similarly, even if we think that it is illusory to measure precisely all aspects of collective behaviours, it is certainly possible to improve our understanding of these phenomena.

The second contribution of this work is that, in terms of results, we can say that the role of memory and forgetfulness, whether individual or collective, physical or artificial (carried by the media), is crucial for collective momentum. Forgetting is present in the phenomena of information space reduction involved at all moments of a group's life, from its creation by the voting mechanism (see Chapter 2), through the filtering effect operated by the media, the group and individual expressiveness. We believe that, paradoxically, this phenomenon is inseparable from collective intelligence mechanisms. We have developed the idea that individuals capable of managing complex expression find it easier to expand their circle of contacts. This corresponds to the ability or the willingness of individuals, for example, for reasons of individual worth, to support a more or less high cognitive load.

By being present both in individuals and the media, the memory and forgetting effects activate phenomena of "merging" and of emergence that are already partially visible in the direct observation of real world collective phenomena. With the spread of artificial media, such as forums, which also have mechanisms of memory and forgetting (see chap. 5), the merging and emerging phenomena take on an entirely different dimension. These phenomena are difficult to estimate in a rational way using our senses, but have very interesting research prospects. In economics, for example, the analysis of trading traces, particularly in financial markets, makes it possible to envisage a better understanding of operators' behaviour (rationality, herd phenomena, etc.). Emergence plays a role of amplifier in collective phenomena. It can, according to its intensity and its adaptation, be disruptive or have a facilitating effect. From this point of view, the study of the influence of group heterogeneity is also a promising research track. Indeed, we have seen that the potential productivity of groups, depending on their size, can be very different in different contexts of heterogeneity.

We believe that online collective environments are not only exciting to study but that they also hold great potential in terms of technical innovation. We hope that this study will spread our enthusiasm on these subjects and inspire other researchers.

ANNEXE

Part of the Survey (62 respondents): questions asked to French peoples are translated here in English.

Are You Male or Female?

Male 80.6%
Female 19.4%

How Old are You?

Less than 20 years: 4.8%
21 to 25 : 548%

26 to 30!: 14.5%
31 to 35: 6.45%
36 to 40: 4.8%
41 to 45: 4.8%
45 to 50: 4.8%
Higher than 50: 0%

What is Your Higher Study Level?

High school or less: 1.61%
Associate's degree: 24.19%
Bachelor's degree: 11.29%
Master's degree: 58.06%
Phd degree or higher: 4.8%

Are Your Hobbies Related to Computer Science?

Yes deeply (website building, programming, software or hardware testing,...): 27.4%
Little, as a user (gaming, chat, mail,...): 35.5%
No: 37.1%

Do You have a Personal Web Site?

Yes: 37.1%
It is in project: 17.74%
I had but not presently: 22.58%
No: 19.35%
No answer: 3.23%

How Often Do You Check Your Email?

Several times per day: 80.65%
Once a day: 19.35%
Once a week or less frequently: 0%

How Many Times Do You Use Internet for Your Main Activity (Professional, School,..) ?

More than 2 h per day: 74.19%
From 15 min to 1 h per day: 22.58%
1 h per week or less: 1.61%

How Many Times Do You Use on Internet for Your Hobbies?

More than 2 h per day: 59.68%
From 15 min to 1 h per day: 27.42%
1 h per week or less: 9.7%
No response: 3.23%

You Use Online Forums from:

Less than 1 year: 12.9%
Between 1 and 2 years: 14.52%
Between 3 and 5 years: 41.94%
Between 5 to 10 years: 25.8%
More than 10 years: 3.23%
No response: 1.61%

How Often Do You Uses Online Forums in Your Hobbies?

Every day: 32.2%
One to 2 times per week: 30.6%
One to 2 times per month: 17.7%
Less to 5 times per years: 16%
No response: 3.23%

How Many Groups Do You Follow Regularly?

Zero: 9.68%
 One: 8.06%
 From one to two: 45.16%
 From 4 to 10: 24.19%
 More than 10: 4.84%
 No response: 8.06%

What is the Size of Your Favourite Groups?

Large size (hard to evaluate the number of participants): 62.9%
 Around 30 participants: 16.13%
 Around 10 participants: 8.06%
 Less than 5 participants: 4.84%
 No response: 8.06%

Do You Use Forum in Your Professional Activities (If Applicable)?

Very often and regularly: 37%
 Sometimes: 27%
 Rarely: 22%
 Never: 8%
 No answer: 4.8%

Do You Use Forums for School or University Activities (If Applicable) ?

Very often and regularly: 11%
 Sometimes: 14%
 Rarely: 24%
 Never: 37%
 No answer: 12.9%

Do You Use Forums for Learning Purpose (Refreshing or Acquiring Knowledge)?

Very often and regularly: 29.1%
 Sometimes: 31.8%
 Rarely: 25.83%
 Never: 6.5%
 No answer: 6.5%

Do You Use Forums in Case of Specific Problem to Solve (e.g Computer Setup Problems, Bad Functioning of Home device,...) ?

Very often and regularly: 48.1%
 Sometimes: 32.26%
 Rarely: 11.61%
 Never: 4.8%
 No answer: 3.2%

Do You Surf Randomly on Forums as You Could Do on Websites?

Very often and regularly: 8.06%
 Sometimes: 24.19%
 Rarely: 40.32%
 Never: 19.35%
 No answer: 8.06%

Do You Visit Known Forums to check for Novelty (Interesting Discussion,...) ?

Very often and regularly: 11.26%

Sometimes: 29.03%
 Rarely: 27.41%
 Never: 24.19%
 No answer: 8.06%

What is Your Forums Consultation Attitude?

I read messages but rarely or never make answer: 43.5%
 I post but only in response, no thread initiation: 17.74%
 I start thread but for less of 10% of my posts: 29%
 I regularly start threads (most of 10% of my posts): 8.06%
 No answer: 1.6%

What is the Average Length of Your Posts?

Less than 5 lines: 24.2
 Between 5 and 10 lines: 45.16%
 More than 10 lines: 12.9%
 No answer: 17.7%

Do You Think that the Size of the Group Influence Your Posting Attitude?

No: 69.35%
 Yes more the group is large, more I tend to post: 11.3%
 Yes, less the group is large, less I tend to post: 11.3%
 No answer: 8.06%

From Your Experience, Do You Think that Large Groups (More than 50 Contributors) are :

More efficient than small groups: 38.7%
 Less efficient than small groups: 6.5%
 The size of the group no matter with its efficiency: 45.16%
 No answer: 9.7%

How Often, Forums Change Your Decisions, Point of View, Activities (e.g Decision, Opinion on a Topic,...) ?

Often: 24.2%
 Sometimes: 67.8%
 Never: 4.8%
 No answer: 3.2%

Do You Make Cross-Posting (Post in Several Groups in the Same Times)?

No: 70.97%
 Yes: 17.74%
 No answer: 11.29%

From a General Point of View are You Satisfied of Your Use of Forums?

Enthusiast: 12.9%
 Satisfied: 59.68%
 Average: 20.97%
 Not satisfied: 4.84%
 No answer: 1.61%

From Your Point of View is "Knowledge Acquisition" Your Major Motivation, Second Order, or Third or Lower Order Motivation to Stay in a Discussion Group?

Major: 58.06%
 Second: 9.68%
 Third or higher: 22.58%
 No answer: 9.68%

From Your Point of View are "Social Contacts, Sympathies, Friends Relations" Your Major Motivation, Second Order, or Third or Lower Order Motivation to Stay in a Discussion Group?

Major: 17.74%
 Second: 58.06%
 Third or higher: 14.52
 No response: 9.68%

From Your Point of View is "the Feeling To Be Useful to the Group" Your Major Motivation, Second Order, or Third or Lower Order Motivation to Stay in a Discussion Group?

Major: 14.52%
 Second: 20.97%
 Third or higher: 53.23%
 No answer: 11.29%

REFERENCES

- [1] L. Lancieri, "Interactions humaines dans les réseaux," Ed. Hermes 2005 ISBN : 2746211084
- [2] J. Paolillo and D. Heald, "Democratic Participation in the Discursive Management of Usenet," in Proceedings of 35th Annual Hawaii IEEE International Conference on System Sciences (HICSS'02)- Vol. 4 - Vol. 4, 2002.
- [3] J. Habermas, "Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society," MIT press reprint 1991.
- [4] L. Goode, and J. Habermas, "Democracy and the Public Sphere" (Modern European Thinkers), Paperback, 2005.
- [5] C. Bachelet and M. L. Caron-Fasan, "Introduction de la messagerie et des forums dans l'entreprise, quel impact pour l'organisation : une approche exploratoire," In Proceedings of 5th conference AIM 2000 (Association information et management).
- [6] L. Dubé, "S'outiller pour e-collaborer," actes du colloque "la gestion du savoir" Quebec 2003. <http://www.cefrio.qc.ca/pdf/atelier3.pdf>, [Accessed march, 3, 2008].
- [7] J. Curran-Smith, S. Sibte Raza Abidi, and P. Forgeron, "Towards a Collaborative Learning Environment for Children's Pain Management: Leveraging an Online Discussion Forums," *Health Informatics J.*, vol. 11, no. 1: 2005, pp. 31-43.
- [8] E. Wenger, R. McDermott, and W. Snyder, "Cultivating communities of practice A guide to managing knowledge," Harvard Business School Press 2002.
- [9] J. Lave and E. Wenger, "Situated Learning: Legitimate Peripheral Participation," Cambridge University Press, 1990.
- [10] O. ZARA, "Le management de l'intelligence collective," M2 éd., 2005.
- [11] E. Bradner, G. Mark, and T. D. Hertel, "Team size and technology fit: participation, Awareness, and rapport in distributed teams," *IEEE Transactions on Professional Communication*, Vol. 48, Issue 1, Mar 2005 pp. 68 - 77.
- [12] L. Lancieri, A. Lavallard, and P. Manson, "E-Brainstorming: Optimization of collaborative learning thanks to online questionnaires", proceedings of CELDA IADIS International conference 2005.
- [13] D. A. Maltz., "Distributing Information for Collaborative Filtering on Usenet Net News," SM Thesis, Massachusetts Institute of Technology, Cambridge, MA. 1994. Available as MIT/LCS/TR-603 and Xerox PARC CSL-94-5, <http://www.lcs.mit.edu/publications/pubs/pdf/MIT-LCS-TR-603.pdf>, [Accessed march, 3, 2008].
- [14] S. Guismond, "Les groupes sociaux," in RJ Vallerand, les fondements de la psychosociologie, Gaétan Morin edition, 1994.
- [15] S. H. Ng, "Equity theory and the allocation of rewards between groups," *Eur. J. Soc. Psychol.*, vol. 11, pp. 439-443, 1981.
- [16] Y. Ye, and K. Kishida, "Toward an Understanding of the Motivation of Open Source Software Developers," In proceedings of the 25th international conference on software engineering 2003 (ICSE03).
- [17] M. Smith, "Invisible Crowds in Cyberspace: Measuring and Mapping the Social Structure of USENET," in Communities in Cyberspace, edited by Marc Smith and Peter Kollock. London, Routledge Press, 1999; <http://research.microsoft.com/~masmith/Invisible%20Crowds%20in%20Cyberspace.doc>, [Accessed march, 3, 2008].
- [18] Y. Saito, J. Mogul, and B. Verghese, "A Usenet Performance Study," Sep 1998. project report; HP Labs, <http://www.ysaito.com/usenet.ps>, [Accessed march, 3, 2008].
- [19] G. Flake, S. Lawrence, and C. L. Giles, "Efficient Identification of Web Communities," KDD International conference 2000.
- [20] L. Lancieri and N. Durand, "Internet User Behavior: Compared Study of the Access Traces and Application to the Discovery of Communities," in *IEEE Trans. Syst. Man Cybern.* (T-SMCA) Jan 2006.
- [21] C. Reffay and L. Lancieri, "Quand l'analyse quantitative fait parler les forums de discussion," Journal STICEF (Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation) 2007.
- [22] B. Mullen, C. Johnson, and E. Salas, "Productivity Loss in Brainstorming Groups: A Meta-Analytic Integration," *Basic App. Soc. Psych.*, vol. 12, pp. 3-23, 1991.
- [23] M. Diehl and W. Stroebel, "Productivity Loss in Idea-Generating Groups: Tracking Down the Blocking Effect," *J. Pers. Assess.*, vol. 61, pp. 392-403, 1991.
- [24] N. Avouris, M. Margaritis, and V. Komis, "The effect of group size in synchronous collaborative problem solving activities," Proceedings AACE Conf. ED-MEDIA 2004, Lugano, June 2004. http://hci.ece.upatras.gr/pubs_files/c80_Avouris_Margaritis_Komis_2004_ED_MEDIA.pdf, [Accessed march, 3, 2008].
- [25] J. S. Valacich, B. E. Mennecke, R. Wachter, R. Wachter, and B. C. Wheeler, "Computer-mediated idea generation: the effects of group size and group heterogeneity," *System Sciences*, 1993, Proceeding of the Twenty-Sixth Hawaii International Conference on vol. iv, 5-8 Jan. 1993 pp. 152 - 160 vol. 4.
- [26] S. Whittaker, L. Teeveen, W. Hill, and L. Cherny, "The dynamics of mass interaction", In Proceedings of Conference on Computer Supported Cooperative Work, 257-264. New York: ACM Press (1998). <http://dis.shef.ac.uk/stevewhittaker/cscw98-published.pdf>, [Accessed march, 3, 2008].
- [27] Q. Jones, G. Ravid, and S. Rafaeli, "An Empirical Exploration of Mass Interaction System Dynamics: Individual Information Overload and Usenet Discourse," In: Proceedings of the 35th Annual Hawaii International Conference on System Sciences, IEEE, Big Island, Hawaii. (2002). <http://modiin.njit.edu/docs/hicss%20on%20system%20dynamics.PDF>, [Accessed march, 3, 2008].
- [28] Q. Jones, "Applying Cyber-Archaeology", Proceedings of the Eighth European Conference on Computer Supported Cooperative Work, 14-18 September 2003, Helsinki, Finland. Kluwer Academic Publishers, Dordrecht Hardbound, ISBN 1-4020-1573-9. http://modiin.njit.edu/docs/jones_ecscw03.pdf, [Accessed march, 3, 2008].
- [29] C. Jensen, S. Farnham, S. Drucker, and P. Kollock, "The Effect of Communication Modality on Cooperation in Online Environments," In Proceedings of CHI 2000, The Hague, Netherlands March 2000. <http://research.microsoft.com/scg/papers/dilemmachi2000.pdf>, [Accessed march, 3, 2008].
- [30] I. E. Morley and G. M. Stephenson, "The social psychology of bargaining," *J. Ind. Relat.* JIR.1979; vol. 21, Issue. 508.
- [31] W. Doise, "Les relations entre groupes" dans *Psychologie sociale*, Livre collectif dirigé par Serge Moscovici, 2003, Puf edit.
- [32] S. Cilesiz and R. Ferdig, "Expressiveness and conformity in Internet-based polls," *First Monday Journal*, vol. 8, No. 7 (July 2003), URL: http://firstmonday.org/issues/issue8_7/cilesiz/index.html, [Accessed march, 3, 2008].
- [33] M. R. Parks and K. Floyd, "Making Friends in Cyberspace," *J. Comp.-Mediated Comm.*, vol. 1, no. 4, March, 1996.
- [34] F. B. Viagas and M. Smith, "Newsgroup Crowds and AuthorLines: Visualizing the Activity of Individuals," in Conversational Cyberspaces, Proceedings of the 37th Hawaii International Conference on System Sciences 2004, <http://research.microsoft.com/~masmith/Newsgroup%20Crowds%20and%20Author%20Lines.pdf>, [Accessed march, 3, 2008].

- [35] G. M. Stephenson and C. J. Brotherton, "Social progression and Polarisation. A study of discussion and negotiation in groups of mining supervisors," *Br. J. Soc. Clin. Psychol.*, vol. 14, pp. 241-252, 1975.
- [36] D. Jodelet, "Représentation sociale phénomène concept et théorie dans Psychologie sociale", Livre collectif dirigé par Serge Moscovici, 2003, Puf edit.
- [37] S. Moscovici, "Communication process and language," in L. Berkowitz edition *Advance in experimental social psychology* vol. 3 pp. 224-270 New York academic press 1967.
- [38] PISA report, "Profil de performance des élèves en compréhension de l'écrit et en sciences" extrait du rapport PISA/OCDE 2003 (apprendre aujourd'hui réussir demain). <http://www.pisa.oecd.org/dataoecd/48/43/34473525.pdf>, [Accessed march, 3, 2008].
- [39] F. Rastier, "Sens et textualité," Université. Paris: Hachette ed., 1989.
- [40] M. Hoey, "Patterns of lexis in text," Oxford University Press, 1991.
- [41] K. W. Church, "Empirical estimates of adaptation: The chance of two Noriegas is closer to $p/2$ than $p2$," 18th International Conference on Computational Linguistics 2000. Ed. ACL. Vol. 1: 180-86.
- [42] D. Peraya "Communication écrite", rapport 2000, http://tecfa.unige.ch/proj/cvs/doc_ress/peraya_com_ecrite.pdf, [Accessed march, 3, 2008].
- [43] O. Le Bohec and E. Jamet, "Effets de redondance et prise de notes", Archive edutice 2004; <http://hal.ccsd.cnrs.fr/docs/00/00/16/66/PDF/n028-64.pdf>, [Accessed march, 3, 2008].
- [44] M. A. Walker, "Informational Redundancy and Resource Bounds in Dialogue," PhD Dissertation 1993.
- [45] J. Sweller, "Cognitive Load During Problem Solving: Effects on Learning," *Cogn. Sci.* vol. 12, no. 2: 257-285.
- [46] J. Sweller, "Instructional Design in Technical Areas," Australia: ACER press, 1999.
- [47] B. Berstein, "A socio linguistic approach to social learning," In J.Gould (ed) Penguin survey of the social sciences. Harmondsworth. Penguin Books 1965.
- [48] M. R. A. Chance, "Anthropologie bio-sociale," ouvrage collectif dirigé par Robin Fox, édition complexe, distribuée en France par Puf 1978.
- [49] M. Jousse, "l'anthropologie du geste", Ed. Gallimard 1955.
- [50] A. Maass, D. Salvi, L. Arcuri, and G. Semin, "Langages in intergroups context: the linguistic intergroups bias". In *J. Pers. Soc. Psychol.* vol. 53, pp. 41-52, 1989.
- [51] J. M. Rabbie, and M. Horwitz, "L'effet discriminatoire entre groupes en fonction d'une réussite ou d'un échec au hasard," In W Doise (ed), *Experiences entre groupes*, Paris Mouton 1979.
- [52] M. Conway, M. T. Pizzamiglio, and L. Mount, "Status communality and agency: implication for stereotypes of gender and other groups," *J. Pers. Soc. Psychol.*, vol. 71, pp. 25-38, 1996.