Online Communities: A Longitudinal Analysis of Communication Activities

Thomas Schoberth
University of Bayreuth
95 500 Bayreuth, Germany
thomas.schoberth@uni-bayreuth.de

Jenny Preece
University of Maryland, Baltimore County
Baltimore, MD 21250
preece@umbc.edu

Armin Heinzl
University of Mannheim
68131 Mannheim, Germany
heinzl@bwl.uni-mannheim.de

Abstract

Online communities (OCs) are seen as important stimulus to electronic business. However, surprisingly little is known about how the communication activity of their users develops and changes over time. A longitudinal study bears the potential to better elaborate the enabling and inhibiting factors of the users' communication activity in OCs. To explore these phenomena, we aim to develop a conceptual framework that serves as a foundation to guide an explorative data analysis of real OCs. The notions of common ground, information overload, interactivity and social loafing will be used to explain the communication activity of the users in online communities. The empirically explored framework will help organizations to support the development of OCs and utilize them in an economically successful way.

Based on a literature review we develop a first conceptual framework. Then, we apply it to describe the development of the communication activity and its determinants in an OC hosted by a German financial service provider. The study examines over 33,000 participants and 1.03 million messages over a period of 3 years. We find a strong effect of external factors on the size of this OC. The size of the OC shows no direct influence on the communication activity of the users. But, in reaction to the increasing information load, communication strategies change and herewith influence the communication activity. The heterogeneity of the users' activity is growing over time and a small minority of users writes more and more of the postings.

1. Introduction

The origins of online communication and the notion of online communities (OCs) or virtual communities reach back to the beginnings of the Internet. Since 1975 e-mail based list servers and since 1979 newsgroups were used by scientists for the exchange of ideas and information [41]. There is no generally accepted definition of the term "online community" (OC), but we use the term to describe the communication and social interaction that is seen in Internet and web-based list servers, bulletin boards, Usenet newsgroups and chats.

OCs initially were seen as social phenomena without commercial focus [30]. With the expansion of the Internet and its accompanying commercialization, OCs were discovered as stimuli to electronic business [15]. Today, many organizations try to integrate sophisticated online communities into their web sites in order to retain existing customers by identifying the needs and believes of their customer base as well as to attract new clients [1, 4, 11, 34]. Actually, Brown et al. [3] found in a study of the usage of web sites, that users of community spaces "visit these sites nine times as often as nonusers do, remain twice as loyal – and buy almost twice as often."

As Online Communities do not exist for their own sake but serve as a medium for the meeting and the communication of their members [29], research has to focus on the communication activity of these users. This means not to consider solely the activity of the individuals but also their relations – their interactivity –, which establishes the communities [35]. Few research contributions have attempted to explain patterns and effect relations in mass communication in OCs over time. We try to close this gap by developing an appropriate conceptual framework of the factors determining the users' communication activity in a financial discussion OC. Such a longitudinal study bears the potential to better elaborate the enabling and inhibiting factors of the evolution of OCs and an empirically explored framework will help organizations to support the development of OCs and utilize them for their economic success.

In this article we draw particularly upon the publications about communication activity in OCs of three groups

of researchers (Whittaker et al. [40], Butler [5], Jones and Rafaeli [19-22]). None of these authors investigated longitudinal data but we use their work to identify factors that influence the users' communication activity in OCs, which we use then to develop a conceptual framework. We especially take into consideration the notions of common ground [6,7], information overload [13, 18, 24, 28, 39], interactivity [16, 29, 31] and social loafing [11, 23, 27]. We apply this framework to show the longitudinal development of the communication activity in the community of a German financial service provider. This community uses a web-based forum as platform and had over 33,000 participants during the 3-year period of the study, which contributed more than 1.03 million messages.

2. Research Questions

We aim to develop a better understanding of the users' communication activity in online communities. This study, represents a first step in which we will develop a conceptual framework of the users' communication activity and determinants, which we will apply to a OC hosted by a financial service provider. Accordingly, we address the following research questions:

- Which factors determine the communication activity of OCs?
- Are these factors related, and if so, how are they related?
- How do these factors, their relationship and their effect on the communication activity change over time?

3. Literature Review: The Basis for the Framework

In order to answer these research questions, a comprehensive literature analysis was undertaken. One field of research identified a lack of social cues in OCs and tries to compensate for this with visualizations of their activity and by presenting information about their participants [9, 10, 32, 33]. This research aims to enhance user interfaces and is concerned with the users' view of OCs. For example, Fiore et al. [12] compared participants' subjective evaluations of authors in newsgroups to quantitative metrics describing the same authors. They found a high correlation between the authors' posting activity with how the other participants perceived the social and informational value of their messages. So, they identified the need to add those metrics to OC's user interfaces in order to offer social cues to the participants.

We identified only three groups of scientists, who attempted to explain the communication activity in OCs from a more general view and who explored their concepts with empirical data: Whittaker et al. [40], Jones and Rafaeli [19, 20], Jones et al. [21, 22] and Butler [5]. Now, we will present their work in chronologically order and we will show how it provides a foundation for our study.

3.1. Whittaker et al.: "The dynamics of mass interaction"

In their study "the dynamics of mass interaction" Whittaker et al. [40] derive a framework of OC behavior by using predictions from the common ground model [7] and the Netiquette guidelines (see newsgroup: news.announce.newusers). Common ground is defined as the sum of "mutual knowledge, mutual beliefs, and mutual suppositions" shared by individuals [7]. According to Clark and Brennan's [6] "model of language use" common ground provides the basis for interpersonal communication.

Whittaker et al. hypothesized that demographic factors (size of the community, familiarity of the members and moderation) influence the establishment of common ground and therefore conversational strategies and interactivity. Conversational strategies (FAQ production, message length and cross-posting) were expected to influence interactivity (thread depth). They tested their framework with a sample of 500 newsgroups consisting of 0.66 million posters and 2.15 million messages. Although their data covered a period of 6 month, they did not consider time as factor in its own right.

Their predictions of the relationship between demographics and conversational strategies were largely confirmed. Newsgroups that were larger in size had more cross-posting and shorter messages indicating difficulties to establish common ground. Newsgroups containing many repeated posters had less cross-posting and longer messages suggesting the effects of the familiarity of the posters. Moderated newsgroups showed more production of FAQs, less cross-posting and longer messages. However, the effect of the demographic variables on the interactivity could only partially be confirmed. The size (i.e., number of posters) and moderation of the newsgroups had no significant influence. As predicted the familiarity of the users had positive influence on the interactivity. However, the effects of the conversational strategies on interactivity showed contrary results compared to the predictions. Cross-posting, which the authors expected to indicate problems with common ground, actually seemed to increase interactivity and longer messages – signs of successful established common ground – seemed to decrease interactivity.

Whittaker et al. suggest two major modifications to their common ground model. The fact that shorter messages support interactivity maybe a reaction to information overload [19, 39]. Users have to handle large amounts

of messages so they prefer shorter postings. The positive effect of cross-posting on the interactivity indicates the benefits of diversity, because, although common ground encourages consensus, such conversation may also be perceived by the users as boring.

3.2. Jones and Rafaeli: "User Population and User Contributions to Virtual Publics: A Systems Model"

Jones and Rafaeli [19, 20] suggest a dynamic model of OCs. They examine the influence of critical mass, social loafing and information overload on the interactivity in OCs. Interactivity "is the extent to which messages in a sequence relate to each other, and especially the extent to which later messages recount the relatedness of earlier messages". As conversation requires interactive communication, interactivity is essential for computer mediated discourse and online communities [16, 29, 31].

In order to initiate a sustainable interactive discourse a critical mass of users is needed [17, 11, 29, 38]. But the number of users tends to influence the number of interactive contributions by a feedback loop influenced by two factors. On the one hand, an increase in the user population may not lead to an equal increase in interactive communication as the occurrence of social loafing generally increases, too. In the context of OCs, social loafing is also referred as free riding or lurking [11, 23, 27]. It describes users taking benefits from the community without contributing to it. Indeed, Nonnecke and Preece [27] found in a survey of dynamic list-servers a strong positive non-linear relationship between the number of lurkers and the number of members. On the other hand, with increasing user population and increasing communication load more and more users reach their individual limit for processing information [26]. This state is known as 'information overload'. Users react by filtering and ignoring information or by leaving the community [13, 18, 24, 28, 39]. Therefore, Jones and Rafaeli predict upper limits of sustainable interactive communication depending on the type of technology used. For example, chat systems lead to a higher communication load and therefore can sustain fewer active users than asynchronous systems.

In Jones et al. [21, 22] they presented initial findings, based on the examination of 600 news-groups over a 6-month period, which actually suggests that information overload has an impact on the discourse structure in OCs. They found "… for example, the larger the group, the smaller the number of words posted in interactive messages. Likewise, the relationship between number of threads and their depth supports the notion of a richness-versus-reach tradeoff strategy. Shorter messages were also more likely to generate discussion threads." [22]

3.3. Butler: "Membership Size, Communication Activity and Sustainability: A Resource-Based Model of Online Social Structures"

Butler [5] presents a resource-based model that focuses on the dynamics of social structure sustainability. According to Butler, social structures or communities can only be sustainable if they provide benefits that outweigh the costs of membership. Without resources, a community cannot provide benefits, and without benefits, it cannot attract new members and retain existing ones.

Butler measures membership size as an indicator of the community's resource base and uses communication activity, indicated by message volume and topic variation, to quantify the benefit provision process. Member gains and losses indicate the ability of an OC to attract and retain individuals. Based on this model, Butler predicts, "that social structures are faced with the fundamental problem of balancing the positive and negative consequences of size and community activity". Increasing the number of interaction partners makes the logistics of knowing each other more difficult. Furthermore, larger social structures are more likely subject to social loafing.

The predictions were tested with four month of data from 206 dynamic list-servers. As expected, community size and communication activity had both negative and positive influence on the sustain-ability of online communities. Larger and more active list-servers attract more new members but also result in a higher member loss.

3.4. Summary and conclusions

The framework of Whittaker et al. [40] is tested with an impressive amount of data, but as a pure cross-section study it does not consider time explicitly. Based on their results common ground is not suited to entirely explain the phenomena in OCs. Clark and Brennan's "model of language use" is derived from direct natural dyadic (between two persons) conversation and does not consider aspects of asynchronicity and "written-ness" [14, 36]. For example Lee et al. [25] argue that "over-hearers … indirect participants who might be watching, listening or reading information" are also able to reach understanding "without directly contributing" to the conversation.

The model of Jones and Rafaeli [19, 20] is thoroughly based on theoretical foundations and time is an implicit factor in its feedback loop. Their initial findings [21, 22] support their predictions of upper limits of sustainable interactive communication. Until now they did not publish longitudinal studies, but we are looking forward to their future results.

Butler [5] also uses empirical data and he claims to take longitudinal data into consideration. However, four months of data does not seem enough for that purpose. Additionally, he aggregates the data in monthly steps resulting in only four points per time series. Maybe, because of this his detailed mathematical models are static and consider time dependencies only in the "… error structure …".

All three articles have things in common: For example Butler's argumentation of the negative effects of community size and activity have strong resemblance to Jones' and Rafaeli's argumentation of the effects of information overload and social loafing. The concept of a resource pool is similar to Jones' critical mass of users. Likewise, Whittaker et al.'s focus on common ground leads to the conclusion that the growing size of OCs has not only positive effects, as often claimed in literature [15, 11]. But in spite of these overlapping constructs further work is still needed to develop a robust theoretical core on which to base future studies.

4. Research Design and Methodology

In the study that follows we build on the theoretic work reviewed in section 3 by developing a conceptual framework of the communication activity in OCs, which we us to guide the exploration of longitudinal data from OCs. In addition to the concepts of common ground, information overload, interactivity and social loafing we add other factors from literature. Like the authors of the reviewed articles we want to investigate communities of different types in order to identify general patterns. By explicitly considering the factor "time" in the analysis of longitudinal data we will study the evolution of the communication activity and its determinants.

Because of the technological differences between OCs we restrict our investigations to factors that are common to all OCs. We sample the messages' creation date, author ID, length and thread structure. From this raw data we derive all our measured indicators. In order to conduct a longitudinal analysis we divide the period of observation into equidistant intervals (in the example presented later: 140 weeks).

As OCs often have a big number of users, we analyze communication activity not at the individual level, but at an aggregated activity level. When possible we use relative measures in order to eliminate obvious correlations between the indicators and to make different communities comparable. For example, there is obviously a strong correlation between the overall number of messages and the number of active users at any particular time. Whereas, the relative measure – the mean number of messages per user – is not expected to have an obvious connection to the number of active users.

The OC of the German financial service provider serves as a first unit of analysis and its development will be characterized and presented in the following sections of this paper by applying an early framework resulting in a set of time series. A typical characteristic of time series is their autocorrelation. That means the value of a time series at a given time t is at least partially dependent on the value at time t-1 or even earlier. Conventional correlation methods like Pearson assume the independence of successive data points and therefore cannot be applied here. Instead, we reason about relationships of the indicators from eye-catching events in the development of the time series. We assume that obvious extrema and turning points that occur at the same points in time at different indicators are causally connected. To strengthen this approach, we take more detailed samples at the interesting points in time. Although this approach has to be seen as preliminary and will be further developed in future work with detailed multivariate time series analysis [2], it offers first explorative dependencies between the indicators.

5. Conceptual Framework

Using the concepts discussed above as a basis we will now present an initial conceptual framework for determining the users' communication activity level in online communities (Table 1 and 2). According to Stegbauer [35] we distinguish between relational and attributive user activity.

Table 1: Indicators of the users' relational and attributive communication activity level in OCs.

Relational communication activity	Attributive communication activity
- Mean number of messages per thread - Ratio of established threads (i.e., with more than one message)	- Mean number of messages per user - Relative standard deviation of messages per user

5.1. Relational Communication Activity

Relational activity (i.e., interactivity) is the extent to which individuals refer to each other's messages [19] and is indicated as threads. These threads are a tree-structured visualization of the discussions that show the sequence and relationship of the messages. Like Whittaker et al. [40] we measured the mean number of messages per thread in a week.

In OCs many attempts to start discussions fail. Posters of messages, which are not replied by other users, may

have low common ground [40]. Because of that, their messages may be filtered and ignored by the other users in reaction to information overload [18, 28, 39]. Our measure for this phenomenon is the ratio of established threads in a week. An established thread is a thread with more than one message. The value of this measure represents the probability of an initiating message to be answered and to successfully start a thread.

5.2. Attributive Communication Activity

In contrast to relational activity, attributive activity is not concerned with the relations of the users but with their individual attributes [35]. These attributes are sampled from all active individuals, and then aggregated for each time interval. E.g., like Whittaker et al. [40] we measure the mean number of messages per active user in a week.

In the literature there often are remarks about an unequal distribution of the users' activity. A small minority of the users usually posts the majority of the messages [40, 27, 35]. Information about such imbalance and heterogeneity of the communication activity are not revealed in the mean number of messages. Therefore, we add the relative standard deviation of the number of messages per active user in a week indicating the deviation from the mean activity level.

Table 2: Presumed determinants and according indicators of the users' communication activity in OCs. (Indicators in brackets are not applied to the financial community.)

Community size	External influences	Communication strategies	User experience
- Number of posters	- NEMAX All Share index - (Season) - (Media presence) - (Advertising expense)	- Mean message length - Mean thread duration - Ratio of subject messages	- Mean duration of membership

5.3. Community Size

According to the papers reviewed in section 3 the size of a community (number of members) tends to play a decisive role in its activity. However, with increasing size, problems of establishing common ground tend to emerge [40] and social loafing and information overload increase [40, 19, 5]. On the other hand the resource pool of the community and consequently its potential to attract members also increases [5]. As an indicator of the community size we measure the number of active (posting) users in a week.

5.4. External Influences

External influences like advertising or publicity in mass media may also stimulate the communication activity [11, 15]. Some possible indicators are the community owner's advertising expenses or the number of times the community is mentioned in newspapers or TV. Other influences are more specific to the individual OC. For example a community of amateur gardeners could show seasonal variations in the users' activity. As the main subject of the OC hosted by the financial service provider is stock trading, we expect the German stock index to externally influence the activity in that particular community. Therefore, we look at the average daily closing price in a week of the NEMAX All Share index.

5.5. Communication Strategies

As Whittaker et al. [40] point out the individuals' struggle to gain common ground are reflected in aggregated communication strategies. Likewise, the findings of Jones et al. [21, 22] show that information overload influences the communication strategies of the users. Changes in communication strategies affect the users' capacity to deal with the information load and therefore may lead to changes in communication activity.

To cope with information overload users tend to adopt one of two basic strategies: first, leaving the community, second, filtering or ignoring messages [13, 18, 24, 28, 39]. A style of messages that generate less information load could evolve to a prevailing strategy. For example, Whittaker et al. [40] reported that communities with shorter messages tend to have higher interactivity, so we will measure the mean message length (in characters) in a week.

The message headers give excellent opportunities for filtering. They indicate content and offer information about creation time and author [28, 39]. I.e., with a daily load of hundreds of messages the users may focus on newer messages and may ignore older ones. Expression of that is the mean thread duration (measured in days) in a week. We calculate this measure from the difference be-

tween the creation times of a thread's first and last message. Threads consisting of a single message are not considered.

Messages with headers containing expressive subject lines should be more attractive than messages that demand the user to read their whole content, as they are faster to grasp and therefore reduce the users' information load. Subject messages are messages without an actual message body and with the whole content in the subject line. Therefore the subject lines of such messages should be exceptionally expressive and we measure the ratio of subject messages to all messages in a week.

5.6. User Experience

Users with a longer history of communication (i.e., more experience) behave differently than newer users [4, 11, 15, 18]. As indicator for the users' experience we introduce the mean membership duration of the active users in a week. We calculate this by ascertaining at a week k for all active users i the week x_i of their very first message, then we average the individuals' membership duration $k - x_i$ over all that users in week k.

6. Explorative Results

The first case study for an analysis using our framework is the OC of a German financial service provider. Membership to the OC is not restricted to customers of the service provider. The members' main motivation for belonging to this community is to gather information about stock trading and to talk about that common interest. Therefore, according to Hagel and Armstrong [15] this is a community of interest. The OC uses a web-based forum as its technological platform. The data covers 987 days (about 3 years or 140 weeks) since the start of the community. In this time 33,315 users contributed over 1.03 million messages.

The OC's development is characterized by applying the factors and determinants discussed above. The time axis of the figures (Figure 1 to 10) is divided into 140 weeks and each point along the axis corresponds to a week of aggregated data. As some of the measures are quite noisy, the solid curves in the figures represent the centered four-week floating averages in order to more clearly illustrate the development of the data.

6.1. Attributive Communication Activity

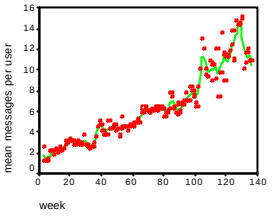

Figure 1: mean number of messages per user

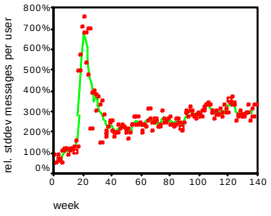

Figure 2: relative standard deviation of messages per user

The mean number of messages per user and week (Figure 1) increases nearly linear (R^2_{cor} = 0.863) from about 2 to 12 messages. No obvious extrema or turning points are visible and therefore it is not possible to draw safe conclusions about other indicators. Whittaker et al. [40] reported 3.1 messages per user during a 6-month period. We found a mean of 40 messages per user in the last 6 months of our data. The attributive communication activity in this web-based forum is therefore much higher than that of the newsgroups in Whittaker's study.

The relative standard deviation of the number of messages per user and week (Figure 2) also increases continuously (except for week 17 to 40). On closer examination, we see in the first weeks, that the 10% most active users contribute about 40% of the messages, whilst at the end of our data the same percentage of users provides 70% of the messages. So, the heterogeneity of the users' attributive communication activity is high and is increasing over time. According to Butler [5] and Jones and Rafaeli [19] the growing proportion of passive users (i.e., lurkers or social loafers) could be due to the growth of the community. But this theory works only for the first half of the data, because even with the drastic decline of the users' number after week 75 (Figure 5) the heterogeneity still increases.

Between week 17 and 40 we see an obvious peek in the heterogeneity with its maximum at week 22. Sampling from that week reveals a single, extremely active user dominating the OC for nearly half a year. This user writes 34% of all contributions in that week and produces an incredible number of 449(!) messages. He copies from news sources like TV or financial newspapers. Typical subjects of his messages are "US car sales rising in February" or "Ad hoc service: Fortec Electronic". His messages are five times longer (1,849 characters) than the average length produced by the rest of the population (343 characters). His dominance is reflected in the time series of the relational communication activity (Figure 3 and 4), the message length (Figure 7) and in the ratio of subject messages (Figure 9). A sample at a later time (week 50) still reveals high activity by that user (3% of all contributions), but he

lost his dominant position and the community is now more balanced.

6.2. Relational Communication Activity

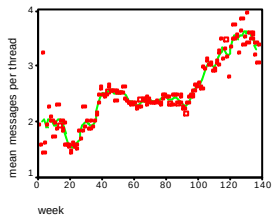

Figure 3: mean number of messages per thread

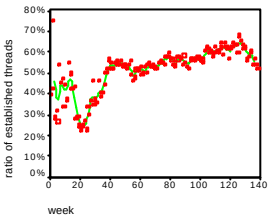

Figure 4: ratio of established threads

The mean number of messages per thread and week (Figure 3) develops in four phases. In the forum's first weeks the measure heavily fluctuates, possibly an expression of the early users' struggle to establish common ground, but maybe simply because of the low numbers of users at the beginning of the community's existence. In the second phase (week 17 to 40) we see a minimum of the number of messages per thread, caused by the above-mentioned dominant user. Even though this user was highly active in an attributive way, he was nearly not at all relationally active: he wrote no replies and only 3% of his messages received replies from others. In a third phase the number of messages is nearly constant at about 2.5 per thread. In the following, final phase (after about week 100) there is a sudden increase to a value of 4.5 messages per thread. An explanation of that last phase will be given in section 6.4; it is caused by a simultaneous increase in the ratio of subject messages (Figure 9).

The ratio of established threads (Figure 4) follows the first two phases of the number of messages per thread. In a third and last phase (after week 40) we see a tendency to an increased probability of successful attempts to start threads, rising from about 55% to 65%. These values are much higher than that of the Usenet study (39%) of Jones et al. [21]. But, a recalculation of the data from Fiore et al. [12] shows ratios of 17% to 86% established threads. So the here measured values are not extraordinary.

6.3. Community Size and External Influence

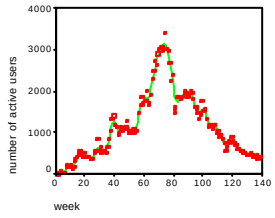

Figure 5: number of active users per week

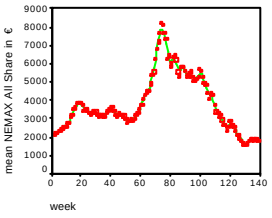

Figure 6: NEMAX All Share Index

The size of the community – represented by the number of active users (Figure 5) – follows a nearly triangular shape, almost synchronous with the NEMAX All Share (Figure 6). The NEMAX – simultaneously with the number of posters – increases strongly up to a maximum at week 75 and decreases rapidly afterwards. The stock market index proves to be a strong external influence on the size of the financial community.

According to Whittaker et al. [40] and Jones and Rafaeli [19] changes in the community size should have direct influence on the attributive and relational communication activity. Therefore, the massive changes the size of the community is undergoing around its maximum at week 75 also should strongly influence the indicators of the communication activity. This assumption cannot be proved here, as these indicators (Figure 1 to 4) show no changes around that time.

6.4. Communication Strategies

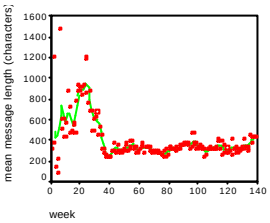

Figure 7: mean message length (in characters)

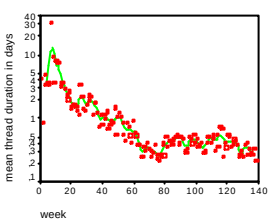

Figure 8: mean thread duration (in days)

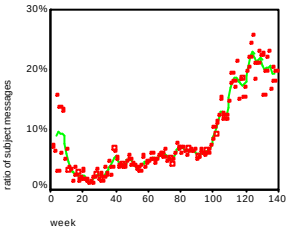

Figure 9: ratio of subject messages

The mean length of the messages (Figure 7) develops in three phases. After initial fluctuations and the effects of the hyper active user in the second phase, this indicator stays – after week 40 – nearly constant on an average level of 326 characters per message. This is equivalent to 8.2 lines (40 characters per line). Compared with the 44 lines in Whittaker et al.'s [40] study, it is a five times lower value.

The mean thread duration decreases steeply – even on the logarithmic scale shown in Figure 8. In the first weeks the mean duration of the threads is up to 30 days, after week 60 the thread duration is less than the half of a day. This steep decline could be interpreted as the users' reaction to the increasing size of the OC (Figure 5) and the increasing information load. The users seemingly focus on

the newest and most current threads. With decreasing pressure from the information load because of decreasing user numbers, the decline of the thread duration flattens after week 75.

The ratio of subject messages (Figure 9) again shows a four phased behavior: First, strong fluctuations, then the effects of the dominant user, after week 40 a slight increase in the proportion of the subject messages and at about week 100 a jump from less than 10% to more than 20%. In order to determine the strong jump in the popularity of this type of message, we investigated a sample after this change (week 135). The proportion of subject messages was 23%. Are these messages a kind of secret language or jargon of a subgroup within the community? 33% of all threads contained subject messages, but only 4% consisted solely of them. So subject message are mixed with normal messages and are not the conversation strategy of a subgroup. Interestingly, 99% of all threads containing subject message, had one of those messages as initiator. The chance of a subject message to initiate a thread containing more than one message was 92%, while a normal message had only a success rate of 42%.

The astonishing attraction of subject messages is because the complete content of the message is contained in the header. Therefore their subject lines have to be more expressive than those of conventional messages. This leads to a higher attraction of discussions starting with subject messages, because the users are able to decide that joining a thread is worthwhile without having to open the thread's first message. This could be interpreted according to Jones and Rafaeli [19] as a change in the communication technology, lowering the information load and enabling higher relational communication activity.

Indeed, threads initiated with subject messages are more interactive (6.3 messages per thread) than threads with conventional messages (2.1 messages per thread). Because of that, the increase in the ratio of subject messages also leads to an increase in the mean number of messages per thread (Figure 3) after week 100.

6.5. User Experience

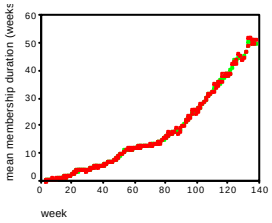

Figure 10: mean membership duration of the active users (in weeks)

The mean duration of the active users' membership (Figure 10), which represents their experience, increases steadily over time and eventually reaches a value of circa 50 weeks. This is more than a third of the community's age (140 weeks), showing a remarkable loyalty from the users. At the maximum number of users at week 75 (Figure 5) there is a slight change in the slope. After that week, the stream of new users weakens and therefore the slope of the mean duration of membership steepens.

The steady increase in the users' experience may be one of the reasons for the steady growth in the number of messages per user (Figure 1) and the users' heterogeneity (Figure 2). Experienced users aggregate over time, write more and more messages, show more loyalty and are less influenced by the size of the community and external factors like the NEMAX (Figure 6) [4, 11, 15].

7. Summary, Conclusions and Future Work

In a review of the articles of Whittaker et al. [40], Jones and Rafaeli [19, 20], Butler [5] and others, no common theoretical model was found to explain longitudinal aspects of the users' communication activity in online communities (OC). For that reason, we have chosen an exploratory approach in which we have developed and applied a conceptual framework based on expanding relevant factors identified by these authors. We especially took into consideration the notions of common ground, information overload, interactivity and social loafing. We distinguish between relational and attributive communication activity in OCs. As determinants the size of the community, external influences, communication strategies and the experience of the users were discussed.

In a comparison of time series data from an OC, hosted by a German financial service provider, we found evidence of a strong effect of external factors, (which in this study was the NEMAX All Share Index) on the community size, which together with the information load influences the users to focus on newer threads. In strong contrast to Whittaker et al. [40] and Jones and Rafaeli [19] there was no visible correlation between community size and the communication activity.

The communication strategy to use subject messages (messages with the whole content in the subject line) lowers the information load and leads to increased interactivity (i.e., relational activity). Communication strategies evolve over time as a reaction to information overload and in that way influence – in accordance with Jones and Rafaeli [19] – the communication activity.

For half a year, a single user negatively influenced the relational activity with his dominance and unusual communication strategy (a lot of long messages but no interactivity). Unusual heterogeneity proves to be negative for the development of OCs and should be closely observed and if possible damped by the community's manager.

The users' heterogeneity seems to inevitably increase with the age of the community, which continuously changes the attributive communication activity and leads to a steady increase in the mean number of messages per user. These tendencies could be related to the continuing aggregation of experienced users as their activity and their proportion of the total activity increases more and more over time. It is difficult to reduce or at least delay these increasing imbalances, which make it hard for new inexperienced users to gain a foothold in the community. One possibility could be to lure highly active users with special features into separate closed areas of the community space to reduce their visible dominance and give new users a better chance to get established. Another possibility would be to grant experienced users moderation rights so they could care for inexperienced users [11, 29].

At a certain size, the population of OCs forms subgroups within the community [35]. The subgroups' evolvement and attributes should influence the users' activity. Therefore, in future, we will use methods associated with social network analysis [37, 35] to add factors of social changes to our framework.

The results described here were found with a rather graphical comparison of time series of the framework's indicators with data from a single OC. From eye-catching events, visible as clear extrema or turning points at the same points in time as different measures, we drew conclusions about dependencies of indicators. To fortify this procedure, we took more detailed samples at the time of these events. In a next step, more accurate analysis will be conducted by correlating the time series to examine the relations of the proposed framework. Conventional correlation methods like Pearson cannot be applied because the time series are autocorrelated and therefore successive data points are not independent of each other. Instead, methods like Box-Jenkins Time Series Analysis [2] will be used. By comparing communities of different types, and themes, we will test the generality of the present results and gain new insights in to the users' communication activity.

8. References

[1] P.J. Boczkowski, "Mutual shaping of users and technologies in a national virtual community", *Journal of Communication*, 49 (2), 1999, pp. 86-109.

[2] G.E.P. Box and G.M. Jenkins, *Time series analysis: forecasting and control*, 2nd Edition. Holden-Day, San Francisco, 1976.

[3] S.L. Brown, A. Tilton, and D.M. Woodside, "The case for online communities", *The McKinsey Quarterly*, 1, http://www.mckinseyquarterly.com/article_page.asp?ar=1143&tk=273538:1143:24&L2=24&L3=45&pagenum=1, fetched 2001-12-19.

[4] J. Brunold, H. Merz, and J. Wagner, *www.cyber-communities.de: Virtual Communities: Strategie, Umsetzung, Erfolgsfaktoren* , Landsberg, 2000.

[5] B.S. Butler, "Membership Size, Communication Activity, and Sustainability: A Resource-Based Model of Online Social Structures", *Information System Research*, 12 (4), 2001. pp. 346-262.

[6] H.H. Clark and S. E. Brennan, "Grounding in communication", L.B. Resnick, J.M. Levine, and S.D. Teasley Eds., *Perspectives on Socially Shared Cognition*, American Psychological Association, Washington, DC, 1991, pp. 127-149.

[7] H. Clark, *Arenas of language use*, University of Chicago Press, Chicago, 1992.

[8] S.L. Condon and C.G. Cech, "Profiling Turns in Interaction: Discourse Structure and Function", *Proceedings of the 34th Hawaii International Conference on System Sciences*, Hawaii, 2001.

[9] J. Donath, "A Semantic Approach to Visualizing Online Conversations", *Communications of the ACM*, 45 (4), 2002, pp. 45-49.

[10] T. Erickson, C. Halverson, W.A. Kellog, M. Laff, and T. Wolf, "Social translucence: designing social infrastructures that make collective activity visible", *Communications of the ACM*, 45 (4), 2002, pp. 40-44.

[11] C. Figallo, *Hosting Web Communities: Building Relationships, Increasing Customer Loyalty, and Maintaining a Competitive Edge*, New York, 1998.

[12] A.T. Fiore, S. Leetiernan, M.A. Smith, "Observed Behavior and Perceived Value of Authors in Usenet Newsgroups: Bridging the Gap", *ACM digital library and CHI 02*, 2002.

[13] H. Franz, "The Impact of Computer Mediated Communication on Information Overload in Distributed Teams", *Proceedings of the 32nd Hawaii International Conference on System Sciences*, Hawaii, 1999.

[14] A. Giboin, "Conversations and Correspondences for Cooperation", *COOP'98 Workshop # 1, The Use of Herbert H. Clark's Models of Language Use for the Design of Cooperative Systems*, Online Working Notes, 1998.

[15] J. Hagel and A. Armstrong, *Net Gain: Expanding markets through virtual communities*, Boston, 1997.

[16] S. C. Herring, "Interactional coherence in CMC", *Proceedings of the 32nd Hawaii International Conference on System Sciences*, IEEE Press, Hawaii, 1999.

[17] S.R. Hiltz and M. Turoff, *The network nation: Human communication via Computer*, Addison-Wesley, London, 1978.

[18] S.R. Hiltz and M. Turoff, "Structuring Computer-Mediated Communication Systems to avoid Information Overload", *Communications of the ACM*, 28 (7), 1985.

[19] Q. Jones and S. Rafaeli, "User Population and User Contributions to Virtual Publics: A Systems Model", *GROUP 99*, Phoenix Arizona, 1999.

[20] Q. Jones and S. Rafaeli, "What Do Virtual 'Tells' Tell? Placing Cybersociety research Into a Hierarchy of Social Explanation", *Proceedings of the 33rd Hawaii International Conference on System Sciences*, Hawaii, 2000.

[21] Q. Jones, G. Ravid, and S. Rafaeli, "Information Overload and Virtual Public Discourse Boundaries", *Eighth IFIP conference on Human-Computer Interaction*, Tokyo, JAPAN, 2001.

[22] Q. Jones, G. Ravid, and S. Rafaeli, "Empirical Evidence for Information Overload In Mass Interaction.", *ACM digital library and CHI 01*, 2001.

[23] P. Kollok and M. Smith, "Managing the virtual commons: cooperation and conflict in computer communities", S. Herring, Ed., *Proceedings of Computer-Mediated Communication: Linguistic, Social, and Cross-Cultural Perspectives*, John Benjamins, Amsterdam, 1996, pp. 109-128.

[24] D.M. Lang and M. Luketick, "Mining for Gems in an Information Overload", *SIGDOC '96*, 1996.

[25] J. Lee, J. Mckendree, R. Cox, K. Stenning and F. Dineen, "Vicarious Learning: More than Common Ground", *COOP'98 Workshop # 1, The Use of Herbert H. Clark's Models of Language Use for the Design of Cooperative Systems*, On-line Working Notes, 1998.

[26] G.A. Miller, „The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information", *The Psychological Review*, 63, 1956, pp. 81-97.

[27] B. Nonnecke, and J. Preece, "Lurker demographics: Counting the silent", *Proceedings of CHI 2000*.

[28] B. Nonnecke, and J. Preece, "Persistence and Lurkers in Discussion Lists: A Pilot Study", *Proceedings of the 33rd Hawaii International Conference on System Sciences*, Hawaii, 2000.

[29] J. Preece, *Online Communities: Designing Usability, Supporting Sociability*, John Wiley & Sons, New York, 2000.

[30] H. Rheingold, *The virtual community: homesteading on the electronic frontier*, MIT Press, Reading, Massachusetts, 1993.

[31] S. Rafaeli and F. Sudweeks, "Networked Interactivity", *JCMC*, 2 (4), 1997.

[32] S. Sack, "Conversation Map: A Content-Based Usenet Newsgroup Browser", *IUI 2000*, New Orleans, 2000.

[33] M.A. Smith and A.T. Fiore, "Visualization Components for Persistent Conversations", *CHI 2001*, Seattle, WA, USA, 2001.

[34] C. Romm et al., "Virtual Communities and Society: Toward an Integrative Three Phase Model", *International Journal of Information Management*, 17 (4), 1997, pp. 261-271.

[35] C. Stegbauer, *Grenzen Virtueller Gemeinschaft – Strukturen Internetbasierter Kommunikationsforen*, Westdeutscher Verlag, Wiesbaden, 2001.

[36] D. Traum, "On Clark and Schaefer's Contribution Model and its Applicability to Human-Computer Interaction", *COOP'98 Workshop # 1, The Use of Herbert H. Clark's Models of Language Use for the Design of Cooperative Systems*, On-line Working Notes, 1998.

[37] B. Wellman, "An Electronic Group is Virtually a Social Network", S. Kiesler, Ed., *Culture of the Internet*, Lawrence Erlbaum Associates, Mahwa, New Jersey, 1997.

[38] S. Whittaker, "Talking to strangers: An evaluation of the factors affecting electronic collaboration", *Proceedings of CSCW'96*, ACM Press, Cambridge, Mass, 1996.

[39] S. Whittaker and C. Sidner, "Email Overload: Exploring Personal Information Management of Email", S. Kiesler, Ed., *Culture of the Internet*, Lawrence Erlbaum Associates, Mahwa, New Jersey, 1997.

[40] S. Whittaker, L. Terveen, H. Hill, and L. Cherny, "The dynamics of mass interaction", *CSCW 98*, Seattle Washington, 1998.

[41] R.H. Zakon, *Hobbes' Internet Timeline v5.3*, Uppsala University. http://www.zakon.org/robert/internet/timeline/, fetched 2001-06-26.