The Embeddedness of Academic Online Groups in Offline Social Networks: Reputation Gain as a Stimulus for Online Discussion Participation?[*]

Abstract

This paper contributes to answering the question of under which conditions members of academic Internet Discussion Groups (IDGs) are motivated to provide help and answers to colleagues during group discussions on the Internet. It presents a simple micro-economic model that specifies mechanisms by which the embeddedness of IDGs in academic social networks provides incentives for active participation of members during group discussions. According to the reputation model the sending of public email messages is a means of the researcher to gain reputation in the academic community. Hypotheses of the models are tested with the help of data on the observed communication behavior of researchers in 49 academic IDGs. The results provide some support for the hypotheses of the model. Most important, they show that the embeddedness of academic IDGs in social networks stimulates the active participation in online discussions. The findings contribute to filling an important gap. They help understanding how offline networks influence online communication by showing which characteristics of offline networks have what impact. Conclusions are drawn for the social design of online groups.

key words: cooperation, knowledge management, online community, public good, social embeddedness, social networks.

Words without references: 8937, Words with references: 10016

[*] Please send correspondence to Uwe Matzat, Sociology Section, Department of Technology Management, Eindhoven University of Technology, P.O. Box 347, 5600 MB Eindhoven, The Netherlands, u.matzat@tm.tue.nl, http://www.tue-tm-soc.nl/~matzat. This research project was conducted while the author was a Ph.D. student at the Interuniversity Center for Social Science Theory and Methodology (ICS), University of Groningen, The Netherlands. I would like to thank Tom Snijders, Henk de Vos (ICS Groningen), three anonymous reviewers and the editors for valuable support and helpful comments.

1. Introduction

The use of Internet groups for the purpose of knowledge exchange is prominent and influential within the academic system, leading to new contacts between researchers (Matzat 2004a). Additionally, business companies can profit enormously from the use of so-called online communities of practice (Wenger and Snyder 2000). The potential of online communities, however, is not exhausted by simply providing up-to-date technology. The large number of high technology online communities that regularly fail is a warning (Ardichvili, Page, and Wentling 2003; Gal 2004). A successful online community needs motivated members who are willing to participate actively in online discussions and, in the case of knowledge sharing, who provide information and help to other members who are in need of it.

This paper contributes to answering two related questions. Firstly, under which social conditions are members of academic Internet Discussion Groups (IDGs) willing to provide help and answers to other members during group discussions on the Internet. Secondly, by which mechanisms do the social conditions influence the communication behavior of members? The term academic IDG describes all kinds of academic emailing lists. The answers to the questions are interesting from two points of view. First, for practical knowledge management efforts and the planned design of online communities it would be useful to know which conditions stimulate the information transfer between academics. Second, an adequate level of active participation in group discussions on the Internet contributes to a public good for the whole group (e.g. Dawes 1980). If members give *public* answers to the questions of other members during a group discussion, then *every* member can profit from this discussion contribution. Since the member gains this benefit independently of the degree of his own participation in the discussion, there is a free rider problem. While a too high amount of email messages may lead to capacity overflow problems, many mailing lists

do not produce enough messages to be interesting in the long run (e.g., Cummings, Butler & Kraut 2002), so that free riding indeed can be a serious problem. This paper not only points to conditions that facilitate overcoming the public good problem. Additionally, it contributes to enhancing our theoretical knowledge about the mechanisms by which the conditions stimulate active participation in online discussions.

Research gives two prominent answers to the question of which incentives members of online groups have for the provision of information. On the one hand, it is argued that reciprocity considerations motivate members to help other members (e.g. Constant, Kiesler, and Sproull 1994; Connolly and Thorn 1990; Rafaeli and LaRose 1993; Rojo and Ragsdale 1997; Thorn & Connolly 1987). On the other hand, it is argued that not only information incentives, but also social incentives could play an important role (Wellman et al. 1996; Kollock 1999; Wellman and Gulia 1999). Kollock (1999) argues that potential reputation gains could motivate members to provide help to other members in online discussions. Wellman & Gulia (1999) argue that online communication is affected by the members' social networks that exist offline. These arguments, however, have to be elaborated. It is unclear by *which mechanisms* the members' social networks that exist offline could influence the online communication and *which characteristics* of (what type of) offline networks show an impact.

This paper contributes to filling these gaps. It elaborates a general model that makes clear under which conditions the members' offline social networks stimulate active participation in discussions of online groups and by what mechanisms they do so.[1] The model makes use of Granovetter's (1985) idea about the importance of the social embeddedness of individual action which in the field of internet research can be found back in the statement that 'relations offline affect relations online' (Wellman et al. 1996). It is tested with data on the communication behavior of researchers in 49 academic IDGs. Section 2 presents the model. It

is used to deduce a number of hypotheses. In section 3 the design of the research project and descriptive findings are presented. Section 4 shows the results of the tests of the hypotheses. Section 5 summarizes the findings, proposes modifications of some weak points of the model, and draws conclusions for the planned design of online communities.

2. Social Embeddedness and the Provision of Incentives for Active Participation in Academic Internet Discussion Groups

In academic Internet groups often only a small minority becomes active in the discussion of mailing lists (Rost 2001; Zelman and Leydesdorff 1999; Stegbauer and Rausch 2001). In their study of scholarly IDGs Rojo & Ragsdale (1997) distinguish three different modes of use of IDGs: a fishing for information mode that is based on reciprocity considerations, an enjoying debate mode that is to some degree non-instrumental, and a social networking mode that views active participation as a means of creating new contacts. Matzat (2004a) showed that IDGs are indeed used for the making of new contacts. However, it is an open question whether interaction inside or outside of the emailing list is used for this purpose. Hoffmeister (2000) found significant bivariate associations between the mailing list members' assessment of the group's potential for making contacts and *gaining status* on the one hand and the self-reported participation behavior on the other hand. While both types of social incentives, status attainment and the making of contacts, may stimulate active participation in academic online discussion, this paper focuses on the role of status attainment.[2] The reputation model claims that some academic IDGs are embedded in integrated research communities and that the embeddedness provides opportunities for the researcher to gain reputation by becoming active as a public provider of information during online discussions.

The theory employs the arguments of the general price theoretical model. According to price theory individuals maximize utility by obtaining certain commodities (Hirshleifer and Glazer 1992). The model considers *academic status* as the relevant commodity or goal of a researcher. This is the relative standing of the researcher in the broader academic world. In the sociology of science it is widely agreed that the attainment of status is an important driving force for researchers (e.g. Bourdieu 1991; Edge 1990; Zuckerman 1988; Hagstrom 1965). The model also assumes that researchers maximize their academic status by spending their time on two different (baskets of) commodities that 'produce' status. On the one hand a researcher can pursue 'usual' research activities. This comprises routine activities such as reading and writing papers, conducting a literature search etc. I refer to all these activities as 'customary academic activity'. This activity basket thus contains a large number of activities that produce status in different ways. On the other hand a researcher can spend his time on writing an email and sending it to the academic IDG he uses. How does this help to produce status, either directly or indirectly? According to the *reputation model* a researcher can achieve reputation in his research community by actively participating in the online discussion. Reputation within one's community of peers, in turn, produces status in the broader academic world. The model focuses on conditions that change the *relative* 'adequacy' (see below) of one activity basket, namely the sending of email answers and questions to the IDG, compared to the other activity basket, namely customary academic activity, for status attainment. As long as these *relative* changes are uniform, it does not matter that customary academic activity is made up of a set of heterogeneous activities. Therefore, they are all treated as one basket. In addition, the model applies Becker's (1976) theory of social interaction which is a special version of the general price theoretical model. It allows the derivation of further hypotheses (see below).

The Reputation Model

The model makes a distinction between two kinds of decisions a researcher has to make as an IDG user, namely the decision whether to send an email question and, if somebody else has sent an email question to the IDG, the decision whether to send an email answer to the IDG or not. I first describe the general choice situation and the main idea of the model. Because of space limitations I then analyze the determining factors for active participation in more detail only for the theoretically more interesting issue of under which conditions a researcher is willing to provide *answers* in the group discussion. That is, I restrict the derivation of hypotheses only to the second decision problem.[3]

The basic assumption is that the researcher maximizes status. One means for this is to solve academic problems for which in turn the *reception of new information* (Z_1) is necessary. New information can be obtained by spending time on some customary academic activities (X) or by spending it on asking for help from other researchers (R_1). The amount of help (R_1) one receives from other researchers depends partly on one's efforts within his IDG. A researcher can write and send email *questions* to an IDG. At the same time, the amount of available help depends also on his basic endowment with respect to opportunities for asking other researchers and colleagues outside of the IDG (D_1). Accordingly, one of the researcher's decisions consists of choosing how much time to spend on these two (baskets of) commodities X (customary academic activity) and R_1 (sending questions to the IDG), given his time constraints and endowment with other opportunities for help.

There are other means for gaining status that are also related to active participation within IDGs. Academic *reputation* is gained within a not too large group of researchers, that is, the community of other researchers who are conducting research on his topic (Hagstrom 1965). Another crucial commodity for the researcher is thus *reputation within his research community* (Z_2). Whereas status refers to the relative standing of a researcher within the whole

research system, reputation refers to the evaluation of the researcher's qualities within a smaller community of researchers. The means for a researcher to gain reputation within his academic community is not only to solve academic problems as part of his customary academic activity, but also to make his academic community aware of his competence *(R_2)*. The competence awareness of the community naturally depends to a large degree on the prominence that the researcher already has. This is the basic degree of competence awareness he is already equipped with (D_2).

The crucial idea of the reputation model is now the following. Under some conditions, providing help to other researchers by sending public email answers to their questions in the view of all other colleagues who use the IDG can be a means of enhancing the collegial awareness of the researcher's competence. While often only little reputation can be gained, under some conditions, e.g. a high degree of network embeddedness, more reputation can be gained. Sending public email answers is then a way of gaining some reputation within the community in exchange against help. This argument rests on ideas of social exchange theory (Blau 1964). To maximize academic status (Z_1 and Z_2), the researcher can spend his time on three kinds of activities, namely customary academic activity (x), sending public questions (h_1), and sending public answers (h_2). Here only the analysis the decision of whether to send a public *answer* is examined. The more comprehensive model can be found in Matzat (2001). The decision of whether to send a public email answer is made after an email question has been sent to the IDG. For the attainment of reputation within his academic community (Z_2) the researcher now has to decide whether to spend his time on some customary academic activity (X) or on writing and sending a public *email answer* to his IDG. The sending of a public email answer (h_2) leads to more awareness of his competence and thereby increases his total degree of collegial competence awareness (R_2). His time constraints for this decision can be written as

$$P_x\, x + P_{R2}\, h_2 = I_2 \qquad (1),$$

where $P_x\, x$ represents the total amount of time spent on general academic activity and $P_{R2}\, h_2$ represents the total amount of time spent on writing and sending an email answer to the IDG.

The total degree of competence awareness (R_2) that a researcher obtains can be divided into a part that depends on his efforts within his IDG in providing answers to sent questions (h_2) and into a part consisting of the awareness he is already endowed with (D_2), that is $R_2 = D_2 + h_2$. A researcher's social income S_2 for the production of reputation, that is the overall value of all his assets, not only consists of the amount of time he can spend on academic activities (X) and competence awareness (R_2) for the production of reputation, but also of the time value of his endowment with competence awareness $P_{R2}\, D_2$.

$$S_2 = I_2 + P_{R2}\, D_2 \qquad (2)$$

The researcher maximizes his status if he chooses a combination of input factors X and R_2 in such a way that the ratio of their marginal products is equal to the ratio of their marginal prices. In addition, the researcher takes into account his time constraints (see equation 1) and his given endowment with collegial awareness of his competence (2). Accordingly, for the analysis one must investigate what influences the shape of the budget lines (1) and the shape of the output curves, that is, the relative marginal products and the relative marginal prices.

If reputation considerations play a role, then for IDG *members with little email experience* there are extra costs involved in answering questions. Members with little email experience need more time for presenting their competence in a favorable way to a whole group of

researchers with specialized norms of conduct. This price effect can be divided into an output and a substitution effect. Both effects lead to an increase in the number of answered emails for the experienced user. They increase the tendency to answer for experienced users. Other conditions influence the relative marginal product of sending public email answers. That is, they enhance the possibilities of acquiring a larger degree of awareness with the same number of sent email answers.

One crucial condition that varies between different IDGs is the *degree of network embeddedness*. The higher the frequency of the members' interaction outside of the IDG and thus the density of the members' network, the more a researcher gains awareness of his competence and thereby reputation in the whole academic community by sending public email answers. In social systems with a higher density it is more likely that information (about a member's competence) diffuses widely (Raub and Weesie 1990). So researchers in IDGs with a high degree of embeddedness face opportunities that make the sending of public email answers a more productive tool for the attainment of competence awareness, independent of their individual attitudes and preferences. A typical production function for the attainment of status of a researcher under a high degree of embeddedness is much steeper within the range of low levels of purchased competence awareness (see Figure 1b) than the production function of a researcher under a low degree of embeddedness (see Figure 1a). Under a high degree the researcher will invest more time in the attainment of competence awareness because of the larger marginal product of this activity. Accordingly, the lower the degree of embeddedness, the lower is the degree of competence awareness striven for within the IDG by a researcher with a given amount of reputation, and therefore the lower the propensity to send emails to the IDG.

A possible expectation might be that under the condition of a low degree of embeddedness, when reputation is less at stake, the difference between experienced and inexperienced members should be smaller. However, the reputation model makes exactly the opposite prediction. The concept of the elasticity of substitution can be used to derive a *negative interaction effect between email experience and embeddedness*. The elasticity of substitution is determined by the ratio of the percentage change in the relative quantity of two input factors to the associated percentage change in their relative prices (indicated by email experience). Price theory states that there is an inverse relation between the elasticity of substitution and the curvature of the production function (Baumol 1977). The reader can see this in Figure 1, which shows the effects of a price change for two researchers with different curvatures of their production functions for a given output level.

(Figure 1 here)

Figure 1a shows the situation for a typical researcher under a low degree of embeddedness, that is, a researcher who faces a lower marginal product of additional competence awareness for status attainment. He must attract a high degree of awareness in return for his forgone general activity in order to stay on the same output curve, that is, to receive the same degree of status. Figure 1b shows the output curve for a researcher under a high degree of embeddedness, which means that additional competence awareness has a higher marginal product for status attainment. He needs to attract much less awareness for his forgone academic activity in order to stay on the same output curve. One can see that, if the price ratio changes from high to low, one must move less far along the output curve before the new budget line is tangential to the output curve. The higher the degree of embeddedness, the smaller is the substitution effect of changing prices. If it is important for status attainment to obtain competence awareness, then it matters less how much effort and time it may cost to

obtain it. The higher the importance of attracting competence awareness, the less easily the public online activity is replaced with another activity when the costs of the first activity increase (see Hirshleifer and Glazer 1992). Therefore, changes in these relative costs have smaller effects on the optimal time allocation. The *difference* in active participation *between email beginners and experienced email users* should be larger, the lower the *relative importance of competence awareness*. The reputation model predicts that the effect of email experience should be larger under a low degree of network embeddedness than under a high degree.

In addition, the opportunities for the attainment of competence awareness are increased by the *existence of personally known colleagues* within the IDG. Their membership enables the researcher to gain more awareness of his competence because one notices and remembers much better useful contributions made by acquaintances and near colleagues. Emotionally significant information, such as an individual's name, can facilitate awareness even in situations where attentional resources are limited (Phelps 2006) and some degree of familiarity with a stimulus increases the likelihood that it will be processed more fluently and easier remembered (Yonelinas 2002; Whittlesea, Jacoby and Girard 1990). Researchers who know about the existence of personally known colleagues within their IDG thus have a higher probability of contributing to the discussion by answering questions.

The same effect occurs if a high-status researcher asks a question within an IDG. If the questioner has a high status within the academic community, then one can acquire more awareness of one's competence through the public provision of help than under the condition of a low-status questioner. A high status researcher often is a gatekeeper who spreads information to an important invisible college (Mulkay 1976, Crane 1969, 1972). For a high-status questioner the probability that his question will trigger an answer from other members

of the IDG is larger than for a low-status questioner. Moreover, researchers with an already established *high status* are endowed with a larger degree of awareness of their competence than researchers with a lower status ($D_{2high\text{-}status} > D_{2low\text{-}status}$). Figure 2 shows what happens if a researcher who faces constant relative prices and displays a constant level of customary academic activity is endowed with a larger basic degree of competence awareness D_{21}. This is reflected in Figure 2 by a move from D_{20} to D_{21} on the competence awareness-axis.

(Figure 2 here)

The researcher with more established status moves to a higher output curve. For him the relative marginal product of R_2 for the attainment of reputation (Z_2) and thus status is lower, because he has already acquired a high degree of collegial awareness of his competence R_2. He can still exchange some customary academic activity for competence awareness by active participation within his IDG, and moves thereby to his equilibrium point *e(i)*. Although his total degree of attention R_{21} is larger than for a low status researcher with a smaller degree of endowed competence awareness D_{20}, his degree of competence awareness purchased online h_1 is smaller than that purchased by the low status researcher ($h_1 < h_0$). His larger endowment of competence awareness - D_{21} versus D_{20} - implies a smaller relative marginal product of his public online activity within the IDG. The relative importance of additional competence awareness for status improvement is less for a researcher who already has a considerable degree of competence awareness than for a researcher who hardly enjoys any competence awareness at all. The higher the established status of a researcher, the lower is the probability that he will send an email answer to the IDG.

Finally, in IDGs the degree of reputation (Z_2) gained by a public answer is higher when the *number of group members with a high status* within the academic community is larger. If one

gains some collegial awareness within a more prestigious group of researchers, this leads to a larger degree of acquired reputation. If more prestigious researchers become aware of ones' competence it is likelier that this information will be spread to some relevant 'invisible colleges' and other institutions for which the prestigious researchers are gatekeepers (Mulkay 1976, Crane 1969, 1972). Accordingly, due to the larger marginal product of awareness, in such groups members should have a greater propensity to answer questions in public.

The predictions of the reputation model for the sending of public email answers can be summarized in the following hypotheses.

Hypothesis 1: The more email experience the researcher has, the higher the probability that the researcher sends public answers to questions in the IDG.

Hypothesis 2: The higher the degree of embeddedness of the researcher's Internet Discussion Group, the higher the probability that the researcher sends public answers to questions in the IDG.

Hypothesis 3: Under a high degree of embeddedness of the IDG, the effect of email experience will be smaller than under a low degree of embeddedness.

Hypothesis 4: The more personal colleagues the researcher has in the IDG, the higher the probability that the researcher sends public answers to questions in the IDG.

Hypothesis 5: A highly prominent researcher has a higher probability of receiving public answers to his sent question(s) than a less prominent researcher.

Hypothesis 6: The more prominent the researcher, the smaller the probability that the researcher sends public answers to questions in the IDG.

Hypothesis 7: The higher the number of very prominent researchers as IDG members, the higher the probability that a researcher sends public answers to questions in the IDG.

The hypotheses are to be tested under ceteris paribus conditions, which implies that a number of other relevant conditions must be controlled, such as pro-social attitudes/group attachment (Kollock 1999), time restrictions, problems related to accessing the IDG, gender differences (Herring 1999), the formal position, research experience, knowledge about the topic (Lewenstein 1995), being a native speaker of the relevant IDG language, and the number of IDG questions that have been sent in the past. The communication habits of some IDGs suggest that a public discussion of questions within the IDG is not always desired. In some IDGs the 'default reply-to' function is set in such a way that a public discussion is avoided. The tests of the hypotheses take this into account. For the testing of hypothesis 5 about the effect of prominence on the probability of receiving an answer to a sent question, one can additionally control for the relevance of the discussion topic to the question sender. By controlling for the topic relevance, the number of written papers, the knowledge of the question sender, and additional factors (see above), one ensures that it is the prominence of the sender and not just the intellectual content of the sent question that stimulates other members to answer the question.

3. Design of the Study, Measurements, and Descriptive Findings

There are two kinds of data used. Firstly, the communication behavior of members of 49 academic emailing lists was observed. In May and June 1999 all the emails that were publicly sent to these 49 IDGs were collected. These data include the information about the participation behavior of researchers. Secondly, a random sample of *all active and passive members* of the 49 mailing lists were sent an online questionnaire in order to obtain the information needed for the measurement of the explanatory concepts.[4] Both kinds of data were combined into one data set with the help of the email addresses of the respondents and were then made anonymous. As a result, the combined data set contains information about the

(non-)participation behavior of the respondent and about the background of the researcher, his social network, his academic community, etc. About 60% of the survey data were collected in May and June. However, to increase the response rate, a final reminder was sent in September so that 40% of the respondents answered the questions after the data collection of the dependent variable was finished. While this is a limitation to assess causality, additional data analyses not presented here that take into account the distinction between early and late responders strengthen the results to be presented in section 4 (see Matzat 2001: Appendix 6.4).

The 49 IDGs were selected in the following way. In the first step, a 'traditional paper-and-pencil-questionnaire' was sent to a random sample of English and Dutch university researchers in eight different disciplines in the humanities, the social sciences, and the natural sciences. This data set is described and used somewhere else (Matzat 2004a). Every respondent to this questionnaire who was an IDG user was asked to fill in the names and some additional information about the five most important IDGs that he used for his research. This resulted in a list of emailing lists and newsgroups used by a random sample of university researchers. From the list of emailing lists, a random sample of 49 emailing lists was selected.

The sample size of the used data set is 4562, which corresponds to a response rate of 35.1% of all 12996 randomly selected active and passive members who were subscribed in the middle of April 1999 to at least one of the 49 mailing lists. A total of 11.2% of the respondents sent at least one email message to their mailing list during the two months of observation. The proportion of active participants among all 12996 selected members is 7.9%, which is slightly smaller. The sample thus has a slight bias in favor of active participants. This is reflected by the fact that those who filled in the questionnaire earlier (measured in months) have a slightly higher probability of having sent a message (r=0.069, p<.01, n=4517). The bias is disturbing,

but additional data analyses (Matzat 2001) suggest that the slight bias does not influence the conclusions drawn from the data analysis.

3.1 The Measurements and the Method of Data Analysis

Dependent variable: The participation behavior

The dependent variable of the data analysis is the public sending of email answers. In contrast to many others, this study does not rely on self-reported assessments of the contribution behavior of the respondents. Rather, the *participation behavior of the respondents was directly observed* by reading the email messages that were sent to the mailing list during the observation period spanning two months.

Every email message of a respondent was coded as a question, an answer, or a different kind of message. If the email contained a request for information then it was coded as a 'question'. If an email contained a reaction to such a request for information then it was coded as an 'answer'. If neither coding was relevant then it was coded as 'another kind of message'. A small number of email messages contained both a question and an answer to another question. These messages were coded both as question and answer. For the measurement procedure it did not matter whether the message contained several questions or several answers grouped together. In order to minimize the effects of a potential bias in the coding of the emails, the two interesting categories of participation were each measured by a dichotomous variable. A respondent was categorized as a 'question sender' if at least one of his sent emails was coded as a question. A respondent was categorized as an 'answer sender' if at least one of his sent emails was coded as an answer. As a consequence, the empirical analyses examine under which conditions researchers become active as senders of answers. Since only a very small minority of researchers sent large numbers of email messages, differences in the intensity of

the discussion contributions among those who were already active participants were not analyzed.

Independent variables

Embeddedness consists of the frequency of the members' social interaction with each other as a research community outside of the discussion group. Since the development of a comprehensive scale that directly measures the embeddedness of the IDG was impossible, the measurement proceeds in two steps. First, it is measured to what extent the group of researchers constitutes an *integrated research community that shares many activities and interests in common.* Second, it is shown that this is associated with three separate indicators that measure dyadic interaction between members outside of the IDG. In step 1 network embeddedness of the IDGs was measured with the help of an unfolding scale analysis (Van Schuur and Post 1990) of five six-point Likert-scaled items, which resulted in a Mudfold scale. The following five items had to be answered (options ranging from '*disagree completely*' to '*agree completely*') by those mailing list members who had been a member for at least one year and who had been active researchers for at least two years in a field related to the mailing list.

To what extent do the following statements describe adequately the group of researchers in your Mailing List?' 'a) It is more a group of unconnected individuals than a research community.' 'b) It is a set of groups and cliques with their own interests and activities, but not much in common as a research community.' 'c) It is a set of groups and cliques with their own interests and activities, but also some common interests as a research community.' 'd) It is a moderately integrated research community that shares some interests and some activities in common.' 'e) It is a well-integrated research community that shares many interests and activities in common.

Note that every respondent was selected for a specific mailing list that was explicitly mentioned in the email invitation so that it was clear which mailing list the respondent had to assess. The items uni-dimensionally cover a wide range of the degree to which the group of researchers in the IDG constitutes an integrated research community that shares many interests and activities in common. The scale analysis showed that the items form a good unfolding scale. The coefficient of scalability is H=.65; none of the five items has a scalability value of H<.55.

The assumption underlying the measurement procedure is that the more the group of researchers in the IDG constitutes an integrated research community that shares many activities and interests in common, the higher is the degree of network embeddedness, which consists of the frequency of the members' social interaction outside of the online group. The following associations support this assumption. I asked the same respondents to what extent they agreed/disagreed with the following statements (six-point Likert scales). *'It is very likely that two arbitrarily selected researchers who (would) regularly send messages to this Mailing List, will ... a)...meet each other at least once every two years at a conference, a professional meeting or another academic gathering. b) ...publish an article in the same academic journal within the next two years. c) ...be members of the same academic committee within the next two years, such as an editorial board of the same academic journal, a board of an academic society, a review panel for grant applications, a dissertation defense panel etc.'* The bivariate associations between the member's perceived degree of integration and their answers to the three statements (r_1=.27, r_2=.22, r_3=.24, all p-values <.01 [two-tailed], n=1022) confirm the assumption that there is more social interaction outside of the IDG and thus a higher degree of embeddedness for a higher degree of integration of the community.[5] The embeddedness of the IDG is indicated by the mean scale value of the emailing list members' scores on the Mudfold scale.

Other independent variables

Details about the measurement of the theoretically interesting independent variables, in addition to those about the measurement of the control variables, can be found in Matzat (2001). Here only the most crucial aspects of the measurements are dealt with.

concept	**measurement**
email experience	natural logarithm of the months of use of email
interaction effect email experience & high degree of embeddedness	dichotomization of embeddedness indicator and multiplication with email experience; cut-off point for dichotomization arbitrarily chosen so that upper third of groups has 'high' embeddedness
prominence	'Consider the research field that is most closely related to the topic(s) of this Mailing List. How well known is your work by other researchers in this field?' ['1 (practically unknown)' to '7 (very well known)']
proportion highly prominent members	proportion of members who scored five or higher on the scale
number of personal colleagues in the list	'With how many researchers of your Mailing List do you have a personal contact? With this we mean mutual acquaintance and at least some regular interaction with each other outside of any Mailing Lists.' [natural logarithm plus one]

The main analysis method consists of two-level multiple logistic regression analyses (Goldstein 1995). They take into account that there are violations of the assumptions of the standard regression model because of the clustering of the data (Snijders and Boskers 1999). The p-values presented in the multivariate analyses are one-sided p-values.

3.2 Descriptive Findings

A total of 4562 respondents filled in the questionnaire (70.2% are male). A total of 68.8% reported that they had been a member of their mailing list for more than 12 months, while 37% had been a member of the same mailing list for at least three years. 34.7% had their main professional position in the USA, 22.3% in the UK, 5.2% in the Netherlands, 4.5% in Canada, 4.4% in Australia, 4.3% in Germany and 2.2% in France. The selection of emailing lists covers a wide range of different kinds of mailing lists. The multivariate data analyses for testing the hypotheses of the reputation model include only the data of 47 lists. Two moderately small lists had to be excluded because of missing values. Group size varied from 16 to 2100 members (arithmetic mean of membership size=425, median=295, standard deviation=441). Most of the mailing lists were located on servers in England or the United States. For the hypothesis testing only the answers of researchers (excluding librarians, students, consultants etc) are used. The descriptive findings (Table 1) make clear that very different groups of researchers are included in the sample.

(Table 1 here)

In line with the findings of other empirical studies (Stegbauer & Rausch 2001; Zelman & Leydesdorff 1999), only a minority of the respondents were active participants in the list discussions during the two months of observation, with 11.2% of them sending at least one email-message. Additional analyses not presented here show that approximately 73.5% or 36 of the groups suffer from too low a volume of messages, as indicated by the group mean scores of subjective assessments of the number of sent messages. For many IDGs within the sample the public good problem of reaching a high number of messages is in fact a serious problem (Matzat 2005). Naturally, a high level of activity has its price. The number of list answers is positively correlated with the mean assessment that the number of messages is too

high (r=0.40, n=49). So to some degree, answering activity is an indicator of the general posting activity in the IDG including the sending of information about conferences etc.

4. Results of the Tests of the Models

Table 2 presents the results of the multivariate analyses. In model 1 (first column) only the six theoretically interesting variables and a few control variables are included. In model 2 more control variables are added. Table 2 shows the following notable results.[6]

(Table 2 here)

The number of personally known colleagues in the mailing list has the predicted positive and significant effect. Those who have more colleagues in the list have more incentives to become active as a public information provider. Moreover, the effect of embeddedness is positive and significant. The higher the degree of embeddedness, the higher the likelihood that a member is willing to spend some time on sending public answers to questions. The main effect of email experience only finds weak evidence (p=0.06 in Model 2). The strength of the price effect of email experience depends on the degree of embeddedness. In IDGs that have a higher degree of embeddedness, the effect of email experience is weaker than in IDGs with a lower degree of embeddedness. This negative interaction effect between embeddedness and email experience reaches significance. The three significant effects are in accordance with the predictions of the reputation model.

The effect of prominence is only weakly significant (p=0.06 in Model 2). Moreover, the sign of the effect is contrary to the prediction of the reputation model. More prominent researchers seem to be somewhat more likely to become active as public information providers than less

prominent researchers. This finding implies that the expected endowment effect of a higher degree of prominence predicted on the basis of Becker's (1976) extension of the price-theoretical model does not find evidence. So hypothesis 6 is not supported. Furthermore, the results show that a higher proportion of very prominent researchers in the IDG reduces significantly the likelihood that members become active as public answer senders. The reputation model, however, predicted a positive effect. A higher proportion of prominent researchers thus does not present additional opportunities for gaining reputation by active participation. Hypothesis 7 does not find support in the data.

A total of 3 of the 6 hypotheses are confirmed by the data. Additionally, the effect of email experience found weak support. The confirmed effects of embeddedness and of the number of colleagues in the mailing list are of special relevance. They cannot be explained by alternative mechanisms based on group attachment or self efficacy, which are ruled out by the control variables (see also Matzat (2001) for this interpretation).

Table 3 shows the results of the multivariate test of hypothesis 5 of the reputation model. The reputation model predicts that highly prominent members have a higher chance of receiving public answers to their questions than less prominent members. This analysis is based on the sample of researchers who sent at least one question to the IDG. Unfortunately 45 of these 178 researchers did not give a self-assessment of their prominence. The exclusion of all cases which have missing values in one of the other control variables leaves only a small sample of n=105 researchers, who are members of 20 mailing lists, for the multivariate test. Since only 16 of these researchers have a prominence score above the scale midpoint, the variable is dichotomized above this point. The bivariate association between being highly prominent and receiving a public answer to a question is only weak (r=.051, p>.05, n=131). However, this does not take into account differences in the embeddedness or in the proportion of highly

prominent members of the IDG, which may have an effect on the likelihood of receiving an answer. The following table shows the results of a 2-level multiple logistic regression analysis that controls for such other effects.

(Table 3 here)

The results show that highly prominent researchers have a higher likelihood of receiving a public answer. Moreover, the group embeddedness has a significant effect, which is remarkable because of the small number of data points (N=20, df_2=15). If the questions are sent to an embedded IDG, the researcher has a higher likelihood of receiving an answer. In addition, researchers who sent more questions, those who have less knowledge, and researchers in a group with a lower proportion of prominent researchers have a higher likelihood of receiving an answer. At first sight, an alternative explanation might be that those who are more prominent write 'better' questions in the sense that their questions are intellectually more challenging. However, the model controls for the knowledge of the question sender, whether the discussion topic is of special relevance to him, and whether the question sender has more contacts in the IDG. The negative effect of the knowledge of the sender does not suggest that those who write 'better' questions have a higher chance of receiving an answer. The prominence effect cannot be explained by this assumption.

General interpretation of the findings

Table 2 and Table 3 presented tests of the seven hypotheses of the reputation model. Preliminary evidence was found to support four of the seven hypotheses. Researchers in embedded IDGs and those who have more contacts in the list have additional incentives to become active as a public help provider. The size of the price effect, that is, the effect of email experience, depended on the degree of embeddedness of the IDG. Moreover, highly

prominent researchers and researchers in embedded IDGs have a higher chance of receiving answers to their questions. This evidence is in accordance with the claim of the reputation model that the interaction in academic IDGs is steered by incentives to gain reputation in the academic community by active online participation. Other explanations such as group attachment, self efficacy and further alternatives are ruled out by the model tests. Furthermore, the results shown offer weak evidence for a fifth hypothesis about the main effect of email experience. The preliminary conclusion is that the reputation model offers an explanation of the researchers' public answering behavior that finds evidence.

However, there are also some weak points. Hypotheses 6 and 7 of the reputation model found no support. What does that mean for the assumptions of the model? For the unexpected finding about the negative effect of the proportion of highly prominent members I offer the following two potential explanations. Firstly, it might be the case that 'ordinary' members of an IDG with a high proportion of very prominent members expect the more prominent members to answer the incoming questions. This reasoning assumes a bystander effect caused by a high proportion of prominent researchers. Bystander effects have been found in chat groups of a very large size (Markey 2000), but not (yet) in academic groups. Secondly, it might be the case that members feel inhibited when the IDG membership includes a high proportion of very prominent researchers. In this case, researchers are concerned about their reputation in the academic community, just as the reputation model implies. Apparently, the assumption that in such an IDG there are additional opportunities for the average member to gain reputation is wrong. Contrary to this assumption, in an IDG with a high proportion of very prominent researchers, members fear loosing a measure of reputation through inappropriate answers. This reasoning thus assumes an inhibiting effect of the presence of prominent members. Further research has to show which of the two explanations is true.

The hypothesis about the expected endowment effect also found no support. One explanation could be that the used indicator, a self-assessment, was not a good one. However, this explanation is weakened by the findings of Table 2, showing that other proxies, such as the number of written papers or the research experience, tend to have positive and significant effects. So, more prominent researchers tend to be more often public answer senders than less prominent researchers. One possible explanation of this finding is the following one. Customary academic activity is a necessity for (almost) every researcher in the academic world. All other activities are of less relevance. However, gaining immediate competence awareness in the *online* world may be of special importance to those who publish a lot and are highly prominent in the circles of their academic community. A researcher's status in the academic system is very much a competitive good, and a highly prominent researcher competes with other highly prominent researchers. While customary academic activity in the form of publishing is still the most important way to secure established reputation and thereby status, *especially* for highly prominent researchers it might be tempting to spend a little bit of the precious time to reach a larger audience in the online world that does not only include the peers of their research community. Mailing lists tend to be very large and include not only the traditional academic community (see Matzat 2001). It is likely that by sending an email to the mailing list a highly prominent researcher reaches a broader audience and the 'academic externals' will remember him especially when his name is well-known. It is tempting for highly prominent members to utilize the mailing list to gain more competence awareness in the external academic world. When you compete with other very productive researchers such awareness in the outer system may make a difference for your overall academic status, as long as you stay productive and thereby gain reputation within the academic community of peers. Technically spoken this means that the output curves are no longer of the same shape for high and low status members which is a precondition for using Becker's (1976) theory of social interaction. So the endowment effect (hypothesis 6) can no longer be expected. In price

theoretical terminology reputation gains respond relatively more to immediate academic externals' competence awareness inputs for highly prominent researchers than for less prominent researchers (see Hirshleifer and Glazer 1992: 298f). The reader should note that this argument only changes hypothesis 6 about the endowment effect, but leaves the core of the model untouched.

5. Summary and Conclusions

This paper presents a model that explains the active participation behavior (sending answers and questions) of researchers in discussions of academic emailing lists on the Internet. The sending of adequate answers to an academic Internet Discussion Group (IDG) is a public good for the whole group. Group members have to solve a cooperation problem in order to overcome the problematic social situation which gives incentives to free ride and to hope that other members will invest time and effort in the discussion. The problematic social situation of how to ensure that members provide help and answers in online discussions is typical for different kinds of online groups.

Existing theories about conditions fostering active participation in the group discussion focus on *information incentives* (e.g. Thorn and Connolly 1987). Other researchers argue that *social incentives* could also stimulate active participation, but they leave open under which conditions and by what mechanisms they do so (Kollock 1999). Network analysts claim that social networks existing offline have an impact on relations and interaction online (Wellman and Gulia 1999; Wellman et al. 1996). However, they leave open what kind of social incentives could stimulate active participation, by which mechanisms offline social networks could have an impact on the online discussion, and which characteristics of offline networks have an impact. The reputation model fills these gaps. It describes mechanisms that explain

how the embeddedness of academic IDGs in social networks of researchers provides social incentives for active participation in online discussions. It supposes that active participation is motivated by the goal of gaining reputation within one's academic community. It is used to derive predictions of which conditions motivate researchers to send questions to the IDG and which conditions stimulate them to send public answers to questions. The reputation model highlights the relevance of the *network embeddedness* of academic IDGs for stimulating active participation. That is, it highlights the effects of a *high density* of the informal offline network.

The hypotheses were tested with data on the archived public email communication of researchers in 49 academic e-mailing lists, supplemented by questionnaire data of the same researchers. The results of the data analyses showed that the reputation model explains the public sending of answers in a reasonable fashion. The evidence supporting the hypotheses about the effects of network embeddedness is of special relevance. The reputation model argues that researchers gain reputation in the academic community by providing answers during the online discussion. When the IDG is embedded in a well-integrated research community, this should provide an additional incentive to become active in the online discussion. In an IDG that is embedded in a well-integrated research community much more reputation can be gained than in an IDG that is embedded in a fragmented or loosely connected research community. The results make clear that the embeddedness in offline social networks has (at least) two effects. It has a direct positive effect on the probability to send a public email answer during the online discussion and it modifies the size of the effect of email experience on the probability to send a public email answer. The model tests also revealed two unexpected findings that suggest two minor modifications. First, it was originally expected that a high proportion of highly prominent researchers provide an incentive to become more active as a public answer sender. According to the data, however, a high

proportion rather decreases the likelihood to become active. This may be explained as an inhibitive effect. Some researchers may fear to loose reputation in such an online group. Second, originally the model predicted a negative endowment effect of being prominent, that should reduce the likelihood to become active as a public answer sender. Again, the data suggest that being highly prominent increases the likelihood to become active. This may be explained as an effect of the larger reputation gains that highly prominent researchers receive through sending public email answers. They may utilize an academic mailing list for reaching broader academic circles outside the community of peers that tend to pay more attention to answers of highly prominent researchers. According to this idea IDGs provide tempting additional opportunities for highly prominent researchers to gain status. Further research has to test the modifications. In any case, they leave the core arguments of the model untouched.

The next step in this line of research is to replicate the found effects in other studies that additionally analyze whether information or other social incentives stimulate participation in online discussions. Should it be confirmed that the social incentives show also effects when the analysis controls for the hypothesized effects of information incentives, then the findings have important implications for practical knowledge management and the planned design of online communities in addition to implications for theoretical research on the Internet and online groups. The results would underline the *importance of different kinds of social networks for online community design*. Firstly, the wider *social network of the research community* that exists around an Internet group can have important effects by providing an important incentive for active participation in online discussions to every member of the group. Secondly, *the personal network of the individual researcher* also could be important by providing an incentive for active participation to the individual researcher. Both aspects can be used for managing the online discussion of an online group.

Three additional remarks are necessary. First, the empirical findings can strictly only be *generalized* to the academic members of the emailing lists used in this study. Further research is needed to find out whether embeddedness and reputation mechanisms facilitate the solution of public good problems in other academic online groups such as online communities that consist of a number of different tools such as public bulletin board systems, emailing lists, chat groups, weblogs etc. and in non-academic online groups. Second, future research that investigates the incentive structures for active participation in a variety of different online groups would allow knowledge to be gained about the 'right' formal and informal governance structure for overcoming problematic social situations that many online groups face. Such research on the *governance structure of online groups* would be useful for reaching desirable outcomes of interaction in online groups and for our theoretical understanding of cooperation-inducing mechanisms that in general work on the Internet (see Matzat 2004b). Third, the findings weaken arguments that emphasize that the uniqueness of interaction on the Internet requires *special* theories for analyzing online interaction (see e.g. Preece 2000). The results suggest that the use of well-known *general* behavioral models might not be a bad strategy for analyzing new phenomena. Mechanisms of social control and reputation effects have been shown to work in many interaction situations. They may be more relevant than early research on computer-mediated communication has suggested (e.g. Kiesler, Siegel, and McGuire 1984). The study of problems of online interaction might profit from the insights of other sociological research on groups that must be adapted to the new online environment. Social network analysts claim that offline networks would have a fundamental impact on the outcomes of online interaction (e.g. Wellman et al. 1996; Wellman & Gulia 1999). The claim is appealing and intuitively seems reasonable. However, strong empirical evidence and our understanding of the mechanisms underlying the influences of offline social networks are limited (Matzat 2004b). Future research should elaborate the claims and evaluate them to find

out which characteristics of social networks do have what effects by which mechanisms. [7] The present study is an example of such an elaboration.

References

Ardichvili, Alexander, Vaughn Page, and Tim Wentling. 2003. 'Motivation and barriers to participation in virtual knowledge-sharing communities of practice.' *Journal of Knowledge Management* 7 (1): 64-77.

Baumol, W. J. 1977. *Economic Theory and Operations Analysis.*Prentice/Hall international.

Becher, Tony. 1989. *Academic Tribes and Territories. Intellectual Enquiry and the Cultures of the Disciplines.*The Society for Research into Higher Education & Open University Press.

Becker, Gary S. 1976. 'A Theory of Social Interactions'. In *The Economic Approach to human Behavior.* eds Gary S. Becker, pp. 253-281. Chicago and London: The University of ChicagoPress.

Blau, Peter M. 1964. *Exchange and Power in Social Life.* New York, London, Sidney: John Wiley & Sons.

Bourdieu, Pierre. 1991. 'The Peculiar History of Scientific Reason.' *Sociological Forum* 6 (1): 3-26.

Connolly, Terry and Brian K. Thorn. 1990. 'Discretionary Data Bases: Theory, Data, and Implications'. In *Organisations and Communication Technology.* eds Janet. Fulk and Charles. Steinfield, pp. 219-233. Newbury Park, CA: Sage.

Constant, David, Sara Kiesler, and Lee Sproull. 1994. 'What's mine is ours, or is it? A study of attitudes about information sharing.' *Information Systems Research* 5 (4): 400-421.

Crane, D. 1969. 'Social Structure in a Group of Scientists: A Test of the 'Invisible College' Hypothesis.' *American Sociological Review*, 335-352.

Crane, D. 1972. *Invisible Colleges. Diffusion of Knowledge in Scientific Communities*. Chicago: The University of Chicago Press.

De Vos, Henk. 1990. 'The welfare state and unemployment: A theoretical analysis'. In *Social Institutions: Their emergence, maintenance and effects.* eds Michael Hechter, Karl Dieter Opp, and Reinhard Wippler, pp. 205-222. Berlin.

Edge, David. 1990. 'Competition in modern science'. In *Solomon's House Revisited: The Organization and and Institutionalization of Science.* eds Tore Fraengsmyr, pp. 208-232. Nobel Symposium 75, Canton: Science History Publications and The Nobel Foundation.

Gal, Yoav. 2004. 'The reward effect: A case study of failing to manage knowledge.' *Journal of Knowledge Management* 8 (2): 73-83.

Granovetter, Mark. 1985. 'Economic Action and Social Structure: The Problem of Embeddedness.' *American Journal of Sociology* 91 (3): 481-510.

Hagstrom, Warren O. 1965. *The Scientific Community.* New York: Basic Books.

Herring, S. C. 1999. 'The Rhetorical Dynamics of Gender Harrassment On-Line.' *The Information Society* 15 (3): 151-167.

Hirshleifer, Jack and Amihai Glazer. 1992. *Price Theory and Applications.* Prentice Hall: Englewood Cliffs.

Hoffmeister, Helge. 2000. 'Kollektiv gut? Bedingungen fuer die aktive Beteiligung an Mailinglisten. Eine empirische Studie.' Master Thesis. Institut fuer Journalistik und Kommunikationsforschung - Hannover.

Kiesler, S., J. Siegel, and T. W. McGuire. 1984. 'Social Psychological Aspects of Computer-Mediated Communication.' *American Psychologist* 39 (10): 1123-1134.

Kollock, P. 1999. 'The Economies of Online Cooperation: Gifts and Public goods in Cyberspace'. In *Communities in Cyberspace.* eds M. Smith and P. Kollock, pp. 220-239. London: Routledge.

Lewenstein, Bruce V. 1995. 'Do Public Electronic Bulletin Boards Help Create Scientific Knowledge? The Cold Fusion Case.' *Science, Technology, & Human Values* (2): 123-149.

Markey, P. M. 2000. 'Bystander intervention in computer-mediated communication.' *Computers in Human Behavior* 16 (2): 183-188.

Matzat, U. 2001. *Social Networks and Cooperation in Electronic Communities: A theoretical-empirical Analysis of Academic Communication and Internet Discussion Groups.* Amsterdam: Thela Publishers. Online available at: http://dissertations.ub.rug.nl/faculties/ppsw/2001/u.matzat/

Matzat, U. 2004a. 'Academic Communication and Internet Discussion Groups: Transfer of Information or Creation of Social Contacts?' Social Networks, 26, 3: 221-255

Matzat, U. 2004b. 'Cooperation and Community on the Internet: Past Issues and Present Perspectives for theoretical-empirical Internet Research.' Analyse & Kritik 26(1): 63-90.

Matzat, U.. 2004c. 'The Social Embeddedness of Academic Online Groups as a Norm Generating Structure: A Test of the Coleman Model on Norm Emergence.' *Computational and Mathematical Organization Theory* 10 (3): 205-226.

Matzat, U.. 2005. 'Die Einbettung der Online-Interaktion in soziale Netzwerke der Offline-Welt: Möglichkeiten der sozialen Gestaltung von Online-Gruppen' (The Embeddedness of Online Interaction in Social Networks of the Offline World: Opportunities for the Social Design of Online Groups.) in: Jäckel, Michael; Mai, Manfred (Hg.): *Online-Vergesellschaftung? Medien-soziologische Perspektiven auf neue Kommunikationstechnologien*. Verlag für Sozialwissenschaften, Wiesbaden: 175-199.

Mulkay, M. J. 1976. The mediating role of the scientific elite. *Social Studies of Science,* 445-470.

Phelps, E. A. 2006. 'Emotion and Cognition: Insights from Studies of the Human Amygdala.' *Annual Review of Psychology, 57,* 27-53.

Preece, Jennifer. 2000. *Online Communities: Designing Usability, Supporting Sociability.* Chichester, New York, Weinheim, Brisbane, Singapore, Toronto: John Wiley & Sons, LTD.

Rafaeli, S. and R. J. LaRose. 1993. 'Electronic Bulletin Boards and 'Public Goods' Explanations of Collaborative Mass Media.' *Communication Research* 20 (2): 277-297.

Raub, W. and Weesie, J. 1990. 'Reputation and efficiency in social interactions: An example of network effects'. *American Journal of Sociology* 96: 626-654.

Rojo, Alejandra and R. G. Ragsdale. 1997. 'Participation in Electronic Forums: Implications for the Design and Implementation of Collaborative Distributed Multimedia.' *Telematics and Informatics* (1): 83-96.

Rost, Martin. 2001. 'Mailing Lists und ihre Strukturen.' retrieved 5-6-2001 from the WWW: http://www.netzservice.de/Home/maro/home/mlstudie10.zip.

Snijders, Tom A. B. and Roel J. Boskers. 1999. *Multilevel analysis : an introduction to basic and advanced multilevel modeling.* London etc.: Sage.

Stegbauer, Christian and Alexander Rausch. 1999. 'Ungleichheit in virtuellen Gemeinschaften.' *Soziale Welt* 50 (1): 93-110.

Stegbauer, Christian and Alexander Rausch. 2001. 'Die schweigende Mehrheit-'Lurker' in internetbasierten Diskussionsforen.' *Zeitschrift fuer Soziologie* 30 (1): 48-64.

Thorn, Brian K. and Terry Connolly. 1987. 'Discretionary Data Bases: A Theory and Some Experimental Findings.' *Communication Research* 14 (5): 512-528.

Van Schuur, Wijbrandt and Henk A. L. Kiers. 1994. 'Why factor analysis often is the incorrect model for analyszing bipolar concepts, and what model to use instead,' *Applied Psychological Measurement* 18 (2): 97-110.

Van Schuur, Wijbrandt and Wendelina Jantina Post. 1990. *Mudfold User's Manual.* Groningen: iec ProGAMMA.

Wellman, Barry and Milena Gulia. 1999. 'Net surfers don't ride alone: virtual communities as communities'. In *Networks in the Global Village.* eds Barry Wellman, pp. 331-367. Boulder, CO: Westview Press.

Wellman, Barry et al. 1996. 'Computer Networks as Social Networks: Collaborative Work, Telework, and Virtual Community.' *Annual Review of .Sociology* 1996 (22): 213-238.

Wenger, E. C. and W. M. Snyder. 2000. 'Communities of Practice: The Organizational Frontier.' *Harvard Business Review* 78 (1): 139-145.

Whittlesea, B. W. A., Jacoby, L. L., & Girard, K. 1990. Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language, 29,* 716-732.

Yonelinas, A. P. 2002. The nature of recollection and familiarity: a review of 30 years of research. *Journal of Memory and Language*, 46, 441-517.

Zelman, Andres and Loet Leydesdorff. 1999. 'Threaded Email messages in Self-Organization and Science & Technology Studies oriented mailing lists.' *Scientometrics* 48 (2): 361-380.

Zuckerman, H. 1988. 'The Sociology of Science'. In *Handbook of Sociology.* eds N. Smelser, pp. 511-574. Newbury Park: Sage.

Figure 1: differences in price elasticity

Figure 1a: weak embededdness: Low curvature = high elasticity of substitution customary academic activity

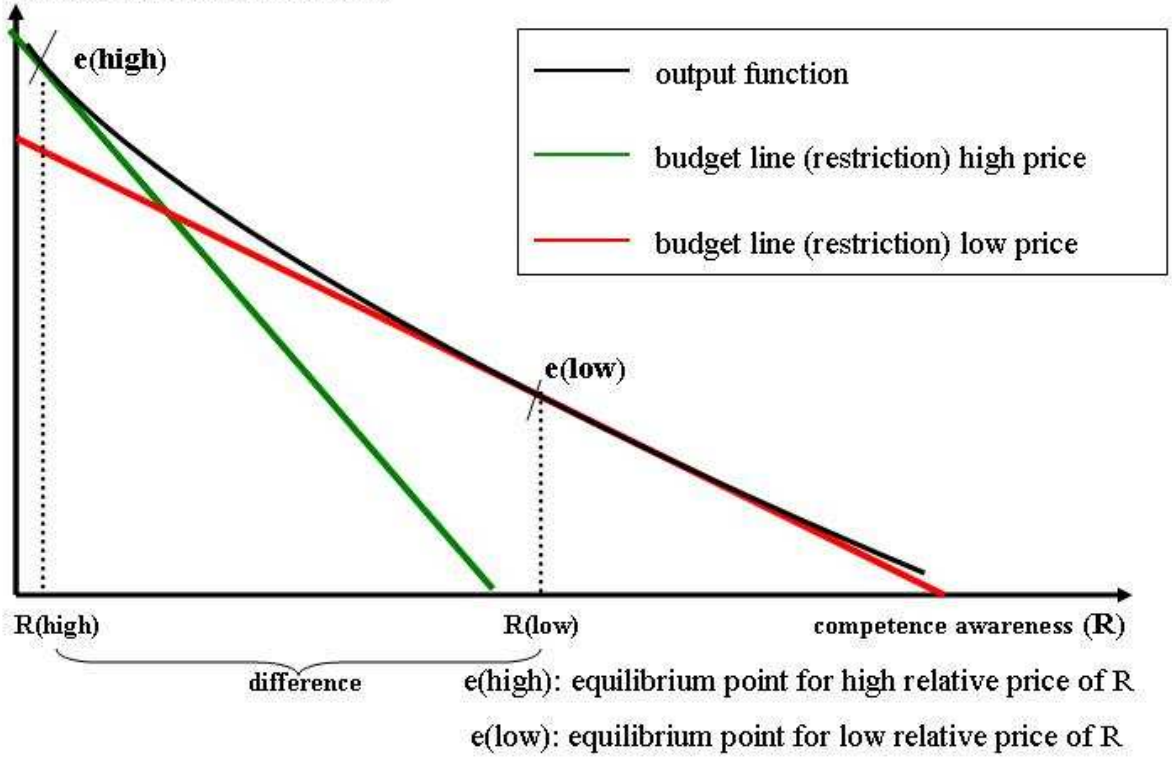
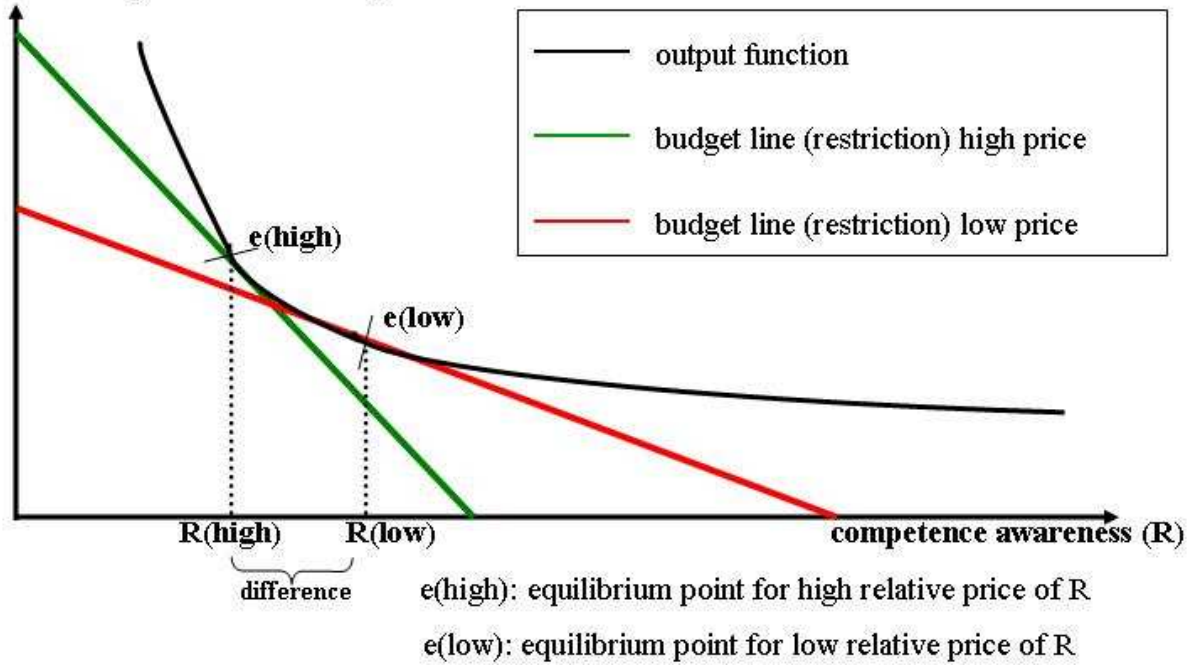
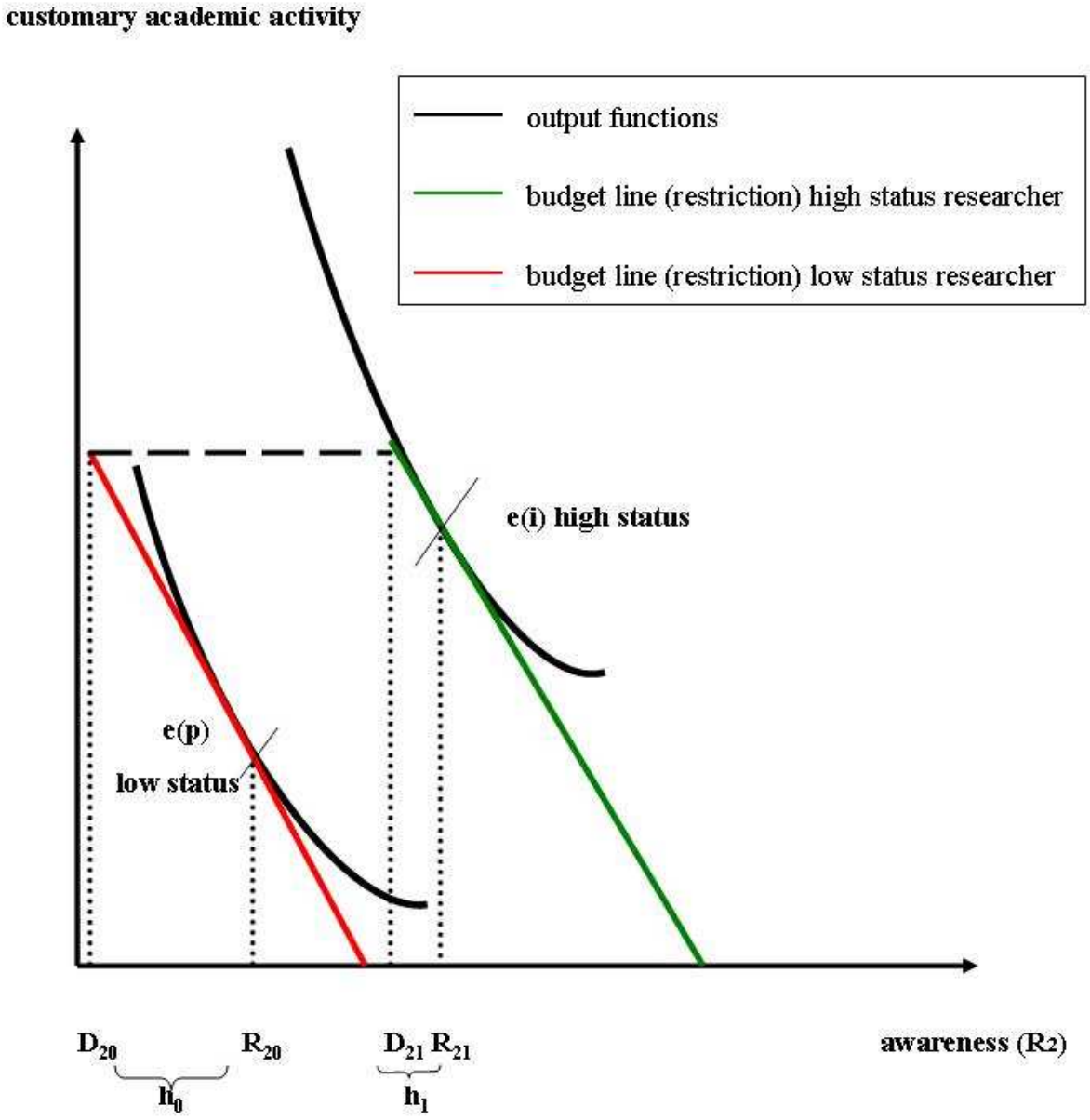


Figure 1b: strong embeddedness: high curvature=low elasticity of substitution customary academic activity



(adapted from: Baumol (Baumol 1977))

Figure 2: endowment effect of established high status



e(p): equilibrium point low status researcher

e(i): equilibrium point high status researcher

$h_0=R_{20}-D_{20}$ $h_1=R_{21}-D_{21}$ $h_1<h_0$

Table 1A: Descriptives of the most important variables

Variable	arithmetic mean	Standard deviation	Minimum	Maximum
Continuous variables (N=2688)				
Email experience	4.4	.6	1.4	6.7
prominence	1.8	1.7	0	6
Contacts in IDG	1.1	1.1	0	6.9
Papers	3.2	2.9	0	30
Conferences	2.2	2.2	0	43
Knowledge	-0.1	1.5	-3	3
Years of research experience	7.6	6.8	0	50
Group variables (n=47)				
Embeddedness	24.7	3.3	16.5	30.9
Proportion of highly prominent members	.19	.11	0	.67
Number of list questions	18.4	39.5	0	199
Dichotomous variables (N=2688)		**Proportion**		
Male		.7		
Native speaker		.6		
Being an answer sender		.05		

Table 1B: Bi-variate correlations between the most important variables

Individual characteristics+

variables	Email exp.	prominence	Contacts	papers	Conf.	Know.	Exp.	Male	Native	Answer
Email experience	1	.18**	.24**	.07**	.08**	.09**	.24**	.15**	.14**	.03
Prominence		1	.45**	.18**	.26**	.39**	.44**	.17**	-.04*	.06**
Contacts in IDG			1	.16**	.21**	.34**	.33**	.10**	.06**	.06**
Papers				1	.37**	.16**	.15**	.10**	-.03	.03
Conferences					1	.17*	.14**	.04*	-.09**	-.01
Knowledge						1	.30**	.11**	-.05*	.07**
Years experience							1	.19**	.07**	.07**
Male								1	-.09**	.04*
Native									1	.07**
answer										1

+: N=2688, *: p<.05, **: p<.01 (2-sided)

Group characteristics++

Variable	Embeddedness	Number questions	Proportion prominent members
Embeddedness	1	.26	.21
Number questions		1	-.15
Proportion prominent members			1

++: n=47, *: p<.05, **: p<.01 (2-sided)

Table 2: Results of 2-level multiple logistic regression analyses of help provision (answers)

Variable	Model 1 Estimated value (standard error)	Model 2 Estimated value (standard error)
Individual level effects		
Email experience	0.36* (0.19)	0.291 (0.194)
Prominence of sender	0.14* (0.07)	0.11 (0.07)
Contacts in IDG	0.06** (0.02)	0.049** (0.019)
Group level effects		
Embeddedness	0.28** (0.10)	0.30** (0.09)
Proportion of highly prominent members	-11.95** (2.55)	-11.39** (2.43)
Cross level interaction effect		
High embeddedness × email experience	-0.16 (0.12)	-0.211* (0.113)
Control Variables		
Group level effects		
Group size		-0.0011* (0.00060)
Number list questions	0.006** (0.003)	0.011* (0.055)
Default reply to group	0.76* (0.39)	0.92** (0.37)
Individual level effects		
Knowledge	0.17** (0.07)	0.11 (0.076)
Years research experience		0.027* (0.014)
Interruption in Access to list		-0.410* (0.22)
Male		0.75** (0.26)
Non-university researcher		0.16 (0.32)
Other university researcher		0.48 (0.34)
Professor		-0.18 (0.30)
Native speaker		0.35 (0.22)
Number visited conferences		-0.07 (0.06)
Number written papers		0.054* (0.03)
Importance of contacts		0.110 (0.077)
*: p<=0.05 **: p<=.01 (one-sided) N=2688 n=47	τ_0^2 = 0.16 (0.13)	τ_0^2 =0.11 (0.10)

N: level 1 sample size (number of individuals); n: level 2 sample size (number of groups); τ_0^2: level 2 variance of the intercept

The interaction variable is defined as the product of the email experience index with the dichotomous 'high embeddedness' variable (0 = no high embeddedness, 1 = high embeddedness). This implies that the estimated main effect of email experience is that for researchers in not highly embedded IDGs, the main effect of 'embeddedness' being that for researchers with email experience index = 0 or with the dichotomous high embeddedness variable = 0.

Table 3: Results of the 2 level logistic regression analysis of the probability of receiving a public answer

Variable	Model 1 Estimated value (standard error)	Model 2 Estimated value (standard error)
Individual level effect:		
Question sender is highly prominent	2.154* (0.914)	3.152* (1.19)
Control Variables:		
Group level effects:		
Default reply to group	0.092 (0.615)	0.76 (0.97)
Embeddedness	0.184* (0.093)	0.23* (0.13)
Proportion of highly prominent members	-9.050* (4.50)	-10.59* (6.27)
Group size		-0.0007 (0.00075)
Individual level effect:		
Knowledge	-0.355* (0.171)	-0.38* (0.19)
contacts in IDG		-0.080 (0.072)
Number of questions sent by the researcher over two months		3.76** (1.50)
Male		0.14 (0.64)
Native speaker		-0.67 (0.58)
Relevance of the topic for question sender		0.24 (0.32)
*: $p<=.05$ **: $p<=.01$ (one-sided) N=105 n=20	τ_0^2 =0.00 (0.00)	τ_0^2 =0.30 (0.45)

N: level 1 sample size (number of individuals); *n*: level 2 sample size (number of groups); τ_0^2 : level 2 variance of the intercept

endnotes

[1] For a critical assessment and an empirical test of the reciprocity model of Thorn & Connolly (1987) see Matzat (2001) who compares whether a model based on information incentives or a model based on social incentives can better explain the communication behavior within academic IDGs. As a consequence, reciprocity considerations are not included in this paper.

[2] See Matzat (2001) for the development of a model based on the idea that active participation in online discussions can be used to facilitate the making of contacts.

[3] Additional hypotheses about the sending of questions can be found in Matzat (2001).

[4] Dependent on the size of the mailing list and on the privacy concerns of the list owner either a random sample of the members' email addresses or all email addresses have been selected. An email version of the questionnaire was sent to those mailing list members who had no access to the world-wide-web during the time of the data collection.

[5] The reader can find more details of the Mudfold scale analysis in Matzat (2001). In addition, the three items are highly correlated with each other (all *r*'s between 0.53 and 0.65). Nevertheless, they do not form a good scale for the measurement of embeddedness. The first principal component that resulted from a principal component analysis of the three items leads to a score that is highly dispersed within the same group. The intra-class correlation is 0.055, meaning that only 5.5% of the variance of the score is between the groups. The low intra-class correlation indicates a low degree of reliability of the principal component scores as a measurement for the group embeddedness. As a consequence, these items are not used as a direct measurement for the degree of embeddedness. The intra-class correlation for the Mudfold scale scores is 0.12, which indicates a higher degree of reliability of this index for the measurement of the group embeddedness. So the used indictor of the degree of embeddedness is a proxy for the density of the informal network.

[6] Additional analyses (Matzat 2001) also include other factors of influence, such as being a question sender, attachment to the group and self efficacy. Since the participation in the online discussion may affect attachment, and vice versa, such variables may control 'too much' of what ought to be explained. As a consequence, these variables are not included in the models presented in the tables. The conclusions, however, do not change when these additional factors of influence are included.

[7] Matzat (2004c) shows how a high degree of embeddedness facilitates the emergence of norms in online groups.