Self-organisation of interest communities: an evolutionary approach

Sylvain Dejean
Université de Rennes 1
June 3, 2008

Abstract

The purpose of this paper is to study the role of individual contributions and some aspects of online community infrastructure on the self-organizing process of interest-based community. Contrary to the classical conclusions in which the presence of free-riding behavior damages the sustainability of online community, our findings show that the heterogeneity of individual contributions favor the self-organization of interest-based communities. This result is robust to the introduction of bimodal structure of preference. We also consider two designed strategies which link the ability to provide information for the community and the perceived switching costs. The results show that both positive and negative correlations between contributions and switching costs may lead to the emergence of self-organized community of interest. If agents are (not) sensitive to the diversity of information, a positive (negative) correlation between contributions and switching costs is likely to produce self-organization.

JEL classification D23 D8 M15

1 Introduction

"Empowered by search engines, recommender systems, search and automatic filters, information technology users are spending more of their waking hours on the internet, choosing to interact with information sources customized to their individual interest. But, does the emergence of a global information infrastructure necessarily imply the emergence of a global village – a virtual community of neighbors freed of geographic constraints? Or, will the borders merely shift from those based on geography to those based on interest?"


As mentioned by Van Alstyne and Brynjolfsson, one of the main issue raised by the development of information technology is whether or not Internet is able
to change the formation and the nature of communities. In the real world, people gather over socioeconomic characteristics (family, school, office, geographic area) and in a more limited way over common interest (clubs, associations). Internet, open new perspectives for individual interactions by lowering the role of geography, generalizing asynchronous communication and allowing the storage of infinite quantity of information. The development of interest communities on the Internet can take different forms like forums, epistemic communities (Wikipedia, Debian) or sharing communities (Darknet, private Bittorrent tracker). Each of them gathers people with more or less specific preference. For example, Yahoo Groups develop discussion on very different topics while www.forum-auto.com is a French community focused on car issues, www.sharing-torrents.com is a private Bittorrent tracker where users share very different contents while xtreme wrestling torrents.net only shares wrestling-oriented contents. The creation of these online interest-based communities might appear obscure and is not well documented. However, the large number of agents interacting in a non-linear way may support the path dependency and emergence properties of complex systems (figure 1).

![Number of daily pageviews (percent) of 4 french interest communities based on cinema, informatic, woman advice and video game.](image)

To our knowledge, few studies focus on the relationship between online communities infrastructures, the behavior of users and the formation of interest communities. Nonetheless, it is an important issue for e-business practitioners and researchers who have in charge the development and the support of online interest-based communities. The model developed by Butler (2007) leads to an interesting and surprising result: infrastructures that reduce communication costs will result in less focused and stable communities. The author supports the theory that the development of ICT technologies, by lowering communication
costs, reduces the ability to form beliefs and prevents the emergence of efficient interest communities.

The content available in interest communities, being video, text or music, is generated by users. It makes the community highly dependent on individual contributions. Adar and Huberman (2000) have shown that in the P2P network "Gnutella" 66% of users download contents but never share them and 37% of files available in the community are provided by only 1% of users. Zhang and Stork (2001) show that inside a forum dedicated to Travels, two kinds of users can be distinguished: the "core members" and the "peripheral members" who respectively represent 10% and 90% of the community.

DangNguyen and Penard (2005) and Krishnan and Alli (2004) suggest that the heterogeneity of contributions can be sustainable inside P2P communities because the most important contributors are able to self-regulate their contribution, even in the presence of free-rider. Many others papers focus on the importance of individual contributions, which take the form of direct reciprocity (Axelrod, 1984), the seeking of reputation (Cabral and Hortascu, 2005, Dellarcas, 2001, Masclet and Penard, 2006, Resnick and Alli., 2006) or professional opportunities (Lerner and Tirole, 2002), with the underlying purpose of overcoming the free-rider problem to enhance the efficiency of interest communities.

The purpose of the present work is to study the role of individual contributions and online community infrastructures on the ability of the community to self-select users among theirs preferences. Our findings show that communities organized over users who share common interests are more likely to emerge when individual contributions follow an heterogeneous distribution, supporting the fact that few strong contributors and many free-rider may be favorable to the development of interest communities. We also approach the role of online community infrastructures in considering two designed strategies for the relationship between contributions and switching costs. The underlying issue is whether strategies which hold the best contributors in the community favour the emergence of interest communities. The results show that two cases must be distinguished, if agents are looking for specific information, positive correlation between contributions and switching costs lead to the emergence of organized interest communities. If agents are not sensitive to the diversity of information the association of low (high) contributors and high (low) switching costs favour the self-organizing process.

To get this result, we build a model based on evolutionary methodology, which follows the previous work of Curien, Fauchard, Laffond, Lesourne, moreau and Lainé (2000, 2001) in their attempt to modelize the self-organizing properties of online communities. Agents of the model gather in communities to share information, each of them has a preference for a particular category of information and is looking for people who have the same preference. The agent-based model considers an initial distribution of agents in communities chosen at random, then the dynamic of the system is based on properties of variation and selection. At each stage of the game, agents evaluate their community (selection process) and choose to stay in it only if there are sufficient users who share the same interest, if not, the agent randomly chooses another community.
(the variation process). The system is at the equilibrium when no agent takes
the decision to leave a community, up to this point, we measure the quality of
the segmentation allowed by the emergence of interest-based communities. The
decision of agents leads (or not) to the formation of interest-based communities
and the quality of segmentation in the community influences the decision of
agents by a feedback mechanism. Macro-structure (self-organized communities)
emerges from the interaction of micro-structure (agents).

Agents differ on two dimensions. first their ability to produce information for
the community and second, the switching cost they perceive. Contributions and
switching costs follow different statistical distributions to study the impact of
agents heterogeneity on the formation of interest-based communities. Following
Klemperer (1987), switching costs can take the form of transaction and learning
costs which in the case of online interest community, relate to the loss of social
capital and reputation or the cost of searching on a new community.

The remainder of the paper proceeds as follows. Section 2 describe the
model. Section 3 discusses characteristics and properties of the agent-based
model. Section 4 deals with the simulation analysis. Section 5 conclude and
opens new lines of research.

2 The model

We consider $N$ agents $i$ with $i \in \{1, 2, ..., N\}$, each one is characterized by a
predefined preference for a particular category of information that we note $s$ with
$s \in \{1, 2, ..., z\}$. Each individual provides information of his favorite category, let
$\phi_i$ be the amount of information produced by the agent $i$. As a consequence, two
agents who belong to the same community and have a preference for different
categories can never exchange valuable information. Individuals are already
defined by switching costs that we note $r_i$ for agent $i$. These costs play an
important role in the model because they make the individuals able to change
the community.

Agents of the model gather in virtual spaces for exchanging information of
common interests. We call these spaces indifferently by the names "forum"or
"community" and we denote by $f$ the number of forums initially available in
the model. At each stage of the process, an agent belongs to a community $j$
with $j \in \{1, 2, ..., f\}$. We can now characterize each individual by a vector of
information that we note:

$$q_{i,t} = \{j, s, \phi_i, r_i\}$$

Components of the vector of agent $i$ are respectively the forum he belongs
to, his favorite information category, the amount of information provided and
the switching cost he supports, at time $t$. 


2.1 The process of selection

At each stage $t$ of the process, with $t \in \{1, 2, ..., T\}$, individuals experience and evaluate the forum they take part in. Then, they decide to stay or to leave the community. We note

$$\forall i \neq i, \quad \Omega_{i,s,j,t} = \sum_{i'=1}^{n} \phi_{i',s,j,t}$$

as the total amount of information of category $s$ available for agent $i$ in the forum $j$ at time $t$. $\Omega_{i,s,j,t}$ is composed of the sum of information produced by agents of category $s$ who belong to forum $j$ at time $t$ minus the information provided by agent $i$. We define a second element which will be part of the individual decision process and is expressed as follows:

$$\forall i \neq i, \quad C_{i,s,j,t} = \frac{\sum_{i'=1}^{n} \sum_{s'=1}^{s} \phi_{i',s',j,t}}{\sum_{i'=1}^{n} \phi_{i',s,j,t} + \varepsilon}$$

$C_{i,s,j,t}$ is the desutility supported by agent $i$ from receiving a type of information which is different from $s$, that’s to say a non-valuable information. The sum of information brought by agents whose preference is not $s$ but who take part in the forum $f$ at time $t$ is located at the numerator and is divided by the sum of information of category $s$ available in the community.

The denominator is increased by a constant $\varepsilon$ to ensure that it could not equal zero if no information of type $s$ was available in the community. A sufficiently low value of $\varepsilon$ permits to be sure that an agent who comes in a forum where no valuable information (i.e. of his favorite type) is proposed won’t be incited to stay in this community ($C_{i,s,j,t} \rightarrow \infty$, and the desutility is maximal).

If a forum is exclusively composed of one category of information $C_{i,s,j,t} = 0$, agents who have particular interests in this type of information bear no costs for taking part in this community. The desutility of agent $i$ rises as information of his favorite type is mixed with non-valuable information. We can now characterize the utility function of agent $i$

$$U_{i,s,j,t} = \Omega_{i,s,j,t} \cdot r_i - C_{i,s,j,t} \cdot k$$

Individual decision-making is based on a cost-benefit analysis. Agents of the model are searching for a community where everyone shares the same interest. On the one hand, the more information of your favorite category is available in a given forum ($\Omega_{i,s,j,t}$) the higher is the probability that agent $i$ stays in this community. On the other hand, individual utilities decrease as information become more heterogeneous ($C_{i,s,j,t}$).

Switching costs act as a force of inertia enhancing the desutility of moving to a new forum. Finally, parameter $k$, which is common to every agent of the model, is chosen to control the role played by the diversity of information. In
other words, it ponders the relative weight of both terms in the right hand side of (1).

Everything equal, a high value of \( k \) means that all members of communities attach importance to the diversity of information, so agents are more willing to change communities. If \( k \) is sufficiently low, the evolutive process which leads to self-organization could never begin or at least be delayed (because agents are accommodated with the diversity of information). It could occur in weak specific interest communities where members are searching for broader advice or randomized information.

We can now define the rules of the decision-making.

\[
\begin{align*}
U_{i,s,j,t} > 0 \quad & \text{Agent } q_i \text{ stay at time } t + 1 \text{ where he was at } t \\
U_{i,s,j,t} \leq 0 \quad & \text{Agent } q_i \text{ move at } t + 1 \text{ in a forum randomly chosen in } f
\end{align*}
\]

This trial and errors process follow a win-stay/loose-shift-like strategy. Nowak and Sigmund (2004) have shown that this is an evolutionary stable strategy when agents are able to interrupt an interaction. The equilibrium is reached when no agent wishes to move during two consecutive periods and the self-organization is measured at this equilibrium.

### 2.2 Measure of self-organisation

The self-organization of interest community should follow two conditions. First, every agent who shares a common interest has to be gathered in the same community. Then forums have to be specialized in one particular category of information. Complying with these two conditions is a guarantee that \( U_{i,s,j,t} \) is the maximum for all agents of the model. The Herfindahl-Hirschman Index (HHI) is a simple and efficient way to measure this double concentration. We build an index \( I^G \) defined by:

\[
I^G = \frac{\sum_{j=1}^{f} \left( \sum_{s=1}^{z} \left( \frac{n_{s,j}}{\sum_{s=1}^{z} n_{s,j}} \right)^2 \right)}{2} + \frac{\sum_{s=1}^{z} \left( \sum_{j=1}^{f} \left( \frac{n_{s,j}}{\sum_{j=1}^{f} n_{s,j}} \right)^2 \right)}{2}
\]

\( n_{s,j} \) is the number of agents of type \( s \) in the forum \( j \), consequently, \( \sum_{s=1}^{z} \left( \frac{n_{s,j}}{\sum_{s=1}^{z} n_{s,j}} \right)^2 \) is the HHI index which measures the specialization of the community \( j \) and \( \sum_{j=1}^{f} \left( \frac{n_{s,j}}{\sum_{j=1}^{f} n_{s,j}} \right)^2 \)
is the HHI index of the concentration of preference $s$. These two indexes vary between 0 and 1, their highest values mean that each community $j$ is specialized in one category $s$ and each category of information $s$ is only shared in one forum $j$. The two terms on the numerator are, respectively, the mean of HHI index for all communities and all types of preferences. We can note that $I^G$ assigns the same weight to the specialization of communities and the concentration of preferences in the measurement of self-organization.

$I^G$ has to deal with the possibility of observing agents with $U_{i,s,j,t} \leq 0$ at the end of the simulation. These agents are members of a randomly chosen community at the last stage of the game which introduces disruption in the self-organization estimation. For example if every agent is randomly distributed in communities, $I^G$ will not equal 0 even if self-organization don’t working. In the same way, if we don’t take into account agents for which $U_{i,s,j,t} \leq 0$, results where 100% or 80% of agents are perfectly self-organized will give the same value of $I^G$. These two phenomenons could lead to a misunderstanding in the process of self-organization, that is why we choose to ponder $I^G$ by the number of agents who don’t find any community when simulation ended. That is $I^A = I^G \cdot \frac{N - R}{N}$, with $R$ the number of agents for which $U_{i,s,j,t} > 0$ at the last stage of the simulation.

### 2.3 Distribution of Initial random sequences

As described in the introduction, the goal of this paper is to determine the role of probability distribution for individual contributions ($\phi_i$) and switching costs ($r_i$) on the ability of interest community to be self-organized. That is why the model generates pseudo-random numbers following different distributions. To ensure that the same amount of information is exchanged in the model, each distribution has the same mean. The table below describes the different distributions used in the simulations, each of them represents a sequence of 10000 random numbers.

<table>
<thead>
<tr>
<th>$s$</th>
<th>Distribution for contributions and switching costs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Uniform (0,2)</td>
</tr>
<tr>
<td>$x$ min</td>
<td>0</td>
</tr>
<tr>
<td>$x$ max</td>
<td>1.99</td>
</tr>
<tr>
<td>$\mu$</td>
<td>0.99</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.33</td>
</tr>
<tr>
<td>1$^{st}$ quartile</td>
<td>0.49</td>
</tr>
<tr>
<td>2$^{nd}$ quartile</td>
<td>0.99</td>
</tr>
<tr>
<td>3$^{rd}$ quartile</td>
<td>1.48</td>
</tr>
<tr>
<td>4$^{th}$ quartile</td>
<td>1.99</td>
</tr>
</tbody>
</table>

The choice of these three distributions permits to balance the effect of the heterogeneity of individual contributions and switching costs on the self-organizing process. The Normal distribution is the more homogeneous which means that every agent produces almost the same amount of information and
have similar switching costs. The exponential distribution is consistent with the empirical evidence that few people contribute to the majority of information in interest-based communities and represent the more heterogeneous distribution of our sample.

3 General features of the model

Before turning to the core of our analysis, we have to focus on the role played by the key parameter $k$ during the evolutive process as well as the emerging property of self-organizing communities of interest.

As the choice of an agent to stay or leave a community depends on the amount and the relevance of information, the parameter $k$ allows to distinguish different levels of taste for the diversity of information. It could also be understood as the ability of the system to favour the mobility of forum members. Considering a high value of $k$ (or an important desutility of processing diverse information) the utility function (1) will be positive only if the community is specialized in information of type $s$ (in this case $C_{i,s,j,t} \rightarrow 0$ to compensate for a high value of $k$) and/or the switching costs of the agent $i$ are high and/or the amount of information of type $s$ is important. The necessity for agent $i$ to take part in a very specialized community of interest decreases with $k$.

3.1 The 2x2 case

To illustrate the role played by $k$ and describe the different states of the system we analyse the case with two categories of information and two forums with homogeneous agents. The number of each type of agents in the two forums is given by

<table>
<thead>
<tr>
<th>$j = 1$</th>
<th>$j = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s = 1$</td>
<td>$n_{1,1}$</td>
</tr>
<tr>
<td>$s = 2$</td>
<td>$n_{2,1}$</td>
</tr>
</tbody>
</table>

Under assumption of homogeneous agents let us consider that $\phi_i = 1$ and $r_i = 1$. According to (1) the utility $U_{i,s,j,t}$ of the $N$ agents is positive if

$$\forall j = 1, 2, s = 1, 2, st \neq s \quad k \ n_{s,j} < (n_{s,j} - 1)^2$$

(C1)

this condition implies the existence of a stable state in the system. A low value of $k$ favors the achievement of C1 which mean that a stable equilibrium is more able to emerge when agents don't pay too much attention to the diversity of information. If $k$ is high, the utility of agent $i$ will be positive only if the forum $j$ gathers few people of preference $s'$ and an important number of agents with preference $s$.

A state $l \in \chi$ is an absorbing state if $p_{ll} = 1$, once reached the system is stuck in this state and the probability to converge to another state $l' \neq l$ is $p_{ll'} = 0$. The condition C1 don't create incentives for agents to move to another forum and don't allow the system to self-select users according to their preferences. The self-organizing process emerges if the following condition is satisfied:
\[ \forall j = 1, 2, \ s = 1, 2, \ s \neq s, \ j' \neq j, \ \frac{(n_{s,j'} - 1)^2}{n_{s',j'}} < k < \frac{(n_{s,j} - 1)^2}{n_{s',j}} \quad (C2) \]

The transition matrix \( T = (P(X_{i,s,t} = j))_{j=1,2} \) of this discrete-time Markov chain, with \( P(X_{i,s,t} = j) \) the probability for individual \( i \) with preference \( s \) to be in the forum \( j \) at time \( t \), take the form \( T_1 = \left( \frac{1}{2} \ 0 \right) \) with \( \frac{(n_{1,2} - 1)^2}{n_{2,2}} < k < \frac{(n_{1,1} - 1)^2}{n_{2,1}} \). As a consequence, agent \( i \) with preference \( s \) who is present in the community 1 stay in it at the next period. If the same agent is in the forum 2 his utility function will be negative and he will move to forum 1 with probability 1/2. Following \( C2 \), \( n_{s,j} \) increases while \( n_{s,j'} \) decreases until \( n_{s,j'} = 0 \). At this stage condition \( C1 \) ensures the stability of the system. The reaching of condition \( C2 \) guarantees the availability of the self-organizing process which make the system converging to an absorbing state where all agents supporting the same preference take part to the same forum.

Finally if

\[ \forall j = 1, 2, \ s = 1, 2, \ s \neq s, \ j' \neq j, \ k \geq \max \left\{ \frac{(n_{s,j'} - 1)^2}{n_{s',j'}}, \frac{(n_{s,j} - 1)^2}{n_{s',j}} \right\} \quad (C3) \]

the transition matrix takes the form \( T_2 = \left( \frac{1}{2} \ rac{1}{2} \right) \) and utility function of agent \( i \) is negative whatever his community. Contrary to condition \( C1 \), the system is not necessarily stuck in this state and the condition \( C2 \) can be randomly reached. The probability for the system to move from \( C3 \) to \( C2 \) depends on the value of \( k \) as well as the number of forums and type of preferences. The higher the number of forums and preferences the lower the probability that a random distribution leads to condition \( C2 \).

To illustrate the different states and the dynamic describe for the 2x2 case, we consider \( n = 100, \ z = 2, \ f = 2, \ \phi_i = 1 \) and \( r_i = 1 \). The simulation is composed of 100 agents, 2 types of preferences and 2 forums\(^1\). The initial distribution is symmetric over the two forums and the two types of preferences. The matrix \( A_t = a_{s,j} \) represents the number of agents with preference \( s \) in the forum \( j \) at time \( t \). At the first stage of the process distribution is uniform \( A_1 = \left\{ \begin{array}{cc} 25 & 25 \\ 25 & 25 \end{array} \right\} \), then the simulation runs for different values of \( k \). Note that only the random variation in the choice of the community creates the necessary disequilibrium for allowing the self-organizing process to begin, the simulation stops after 150 periods. The graphics below show the number of agents in the two forums at the each stage of the game. The number of agents in each forum is on the ordinates, stages of the process are on the abscissa and the plain and dashes lines represent respectively the forum 1 and 2.

\(^1\)All the simulations of this paper are computed using \texttt{mathematica 5.0}. Files are available on request at sylvain.dejean@univ-rennes1.fr.
The graphic 1 is drawn for $k = 5$ and condition $(C1)$ holds for each agent $i$. As the utility function of each agent is positive the system stays in its initial state. For graphics 2, 3 and 4, $k > \max\left\{ \frac{(n_{1,1} - 1)^2}{n_{1,2}}, \frac{(n_{1,2} - 1)^2}{n_{2,2}} \right\}$ at the first stage of the process which mean that all the agents have a negative utility function. The random distribution at the second stage makes the system switches to condition $C2$ for $k = 25$. The number of each type of agent in the two forums are $A_2 = \begin{pmatrix} 23 & 27 \\ 22 & 28 \end{pmatrix}$, which creates an advantage of being in the forum 2 for the agents of type 1 and in the forum 1 for the agents of type 2. The absorbing state is reached after 8 periods and $C2$ has led to a self-organized system where all agents of the same preference are brought together in the same forum. The same process occurs later for $k = 50$ because to reach condition $C2$ the system needs higher deviation from the initial distribution. When $k = 60$, the emergence of self-organized interest communities is more unlikely to happen for a high value of $k$, the random distribution provides by the transition matrix $T_2$ have a
lower probability to create the condition $C2$. In this type of Markov chain it's theoretically possible to reach a stable state which is close to a nash equilibrium even if the system don't favour it (Young, 1993). In this article we focus on the relative ability to reach a self-organized equilibrium when some aspects of the system varies, by considering a finite period of time for simulations.

3.2 number of communities and emerging property

Following the method described in the section 2.2 to measure the self-organization of communities of interest, the model is only able to reach $I^{G} = 1$ (i.e. to be perfectly self-organized) if $f \geq z^2$. The number of communities plays an important role in the evolutive process and the speed of convergence to the equilibrium. To illustrate this point, the model is parameterized using $n = 1000$, $z = 5$, $k = 25$, and $r_i = 1$, $\phi_i$ follows an uniform distribution on $[0, 2]$. The distribution of agents in the first period is also randomly generated and only the number of forums varies. The number of agents in each forum is on the ordinate, stages of the process are on the abscissa and each line represent a forum.

By using the command $\text{SeedRandom}$ on mathematica 5.0, the same pseudo random numbers are generated during the simulation. It ensures that the result is not caused by a random event.

---

2 If the number of forum is less that the number of type of preference, a community can’t be specialized in only one type of information except if some agents leave the game. In both case self-organization isn’t maximal.

3 By using the command $\text{SeedRandom}$ on mathematica 5.0, the same pseudo random numbers are generated during the simulation. It ensures that the result is not caused by a random event.
Keeping everything equal, a low number of forums reduces the number of potential moves and accelerates the convergence to the equilibrium. On the contrary, a high number of forums delays the moment when a community reaches the critical mass of information which lead to the emergence of an organized structure.

Graphics 6 and 7 illustrate the way organized communities arise in the model. The emergence of a self-organization in a community is characterized by a sudden increase in the number of agents who have the same interest. There is a threshold value when the critical mass of a particular information is reached.
Then, the agents who don’t share the same interest leave the forum and those who share the same interest but are in another community are led to join the specializing community. As described by Gilbert and Conte (1995) the question of emergence in computer simulation of social life, is linked to the nature of the complex system. Interaction between a large number of agents (at the micro level) produces some observable patterns of the global state (the macro level). In the following sections, we will observe and measure regularities formed by the behaviour of agents and try to highlight under which conditions these forms emerge.

4 Heterogeneity of contributions

To capture the impact of the diversity and the heterogeneity of contributions on the organization of interest communities, \( \phi_i \) is generated following the three distributions described in section 2.3. \( r_i \) remained constant to focus on the role of \( \phi_i \). Parameters are defined as follows: \( n = 1000, f = 5, z = 5, r_i = 1 \).

Generating 1000 pseudo random numbers should not allow the law of large numbers to apply, then the strong path dependency associated with the evolutive process of the model should play a role on the measurement of self-organization. To avoid this problem, series of 20 simulations involving the same parameters and only differentiated by \( \phi_i \) are conducted. Graphic data are the average of the results of the 20 series which guarantee that \( I_G \) don’t depend on the particularity of extreme values.

```

<table>
<thead>
<tr>
<th>( \phi_i )</th>
<th>mean of ( I_A ) for ( k \in [1, 36] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>0.35</td>
</tr>
<tr>
<td>Uniform</td>
<td>0.29</td>
</tr>
<tr>
<td>Normal</td>
<td>0.22</td>
</tr>
</tbody>
</table>
```

Graphic 8: heterogeneity of contributions and self-organization.

The results of simulations confirm the previous remark that self-organization occurs for an intermediate value of \( k \). For \( k = 1 \), the value of \( I_G \) is the result of the initial random distribution because no agents are willing to move to another community. When \( k = 36 \), whatever the distribution of \( \phi_i \), agents
attach too much importance to the diversity of information to be able to stay in a community.

The main result of this simulation is the superiority of the exponential distribution in the self-organizing process. The normal distribution appears to be less appropriate for the emergence of organized communities of interests, especially for a high value of $k$. The heterogeneity of contributions seems to play an important role in the ability of the model to allow the emergence of organization when agents are searching for specific information. Let us take an example to illustrate this fact. Consider 4 agents who share the same interest, each of them produces 5 units of information. There are 4 forums where agents can exchange information. The probability for an agent to be in a particular forum at time $t$ is 0.25 and he decides to stay in a community if 15 units of information are exchanged, which occurs when 3 agents are in the same forum. This event is associated with a probability of $4.68\% \left(\binom{4}{3}(0.25)^3(0.75)^1\right)$. Consider now that the distribution of contributions is heterogeneous, 3 agents produce each 2 units of information each and the fourth agent provides 14 units. For a community to emerge, the fourth agent has to be in the same forum as another agent, which happens with a probability of $10\% \left(\binom{4}{2}(0.25)^2(0.75)^2\right)$.

The more people are willing to exchange specific information, the more a population has to be composed of agents who provide a large amount of information. These individuals allow the forum to reach a critical mass of information, from which a self-organized community emerges.

4.1 bimodal preference

In the previous simulations the preferences of agents were unimodal. To consider how the results depend on this particular assumption, we modify the model to allow for bimodal preferences. Now agents are defined over two categories of information: the utility function of agent $i$ takes this form

$$\forall s' \neq s \quad U_{i,s,s',j,t} = (\Omega_{i,s,j,t}r_i - C_{i,s,j,t}k)\beta + (\Omega_{i,s',j,t}r_i - C_{i,s',j,t}k)(1 - \beta) \quad (2)$$

For convenience sake, equation (2) is linearly defined and parameter $\beta$ balances the intensity of preferences. For $\beta = 0$ (or symmetrically $\beta = 1$) preferences are unimodal and for $\beta = 0.5$ the agent $i$ is indifferent to his two favorite categories of information. If $\beta \in [0, 0.5]$ (symmetrically $\beta \in [0, 0.5]$), the agent $i$ has a strong and a weak preference for information of category $s(s')$ and $s'(s)$. The quantity of information $\phi_i$ provided by agent $i$ is also balanced by parameter $\beta$, this assumption permits to link contributory capacities of agent $i$ to his preference.

To compare with the unimodal preference case, parameters are defined as in the previous section: $n = 1000, f = 5, z = 5, r_i = 1$. For the tree functions of distribution used in the simulation we give $\beta$ three values $0.6, 0.8$ and $1^4$ which correspond to stronger preferences, over a particular category of information, as $\beta$ becomes closer to $1$.

---

4 As (2) is symmetric over $i$ and $j$, it would make no difference to run the simulation for $\beta = \{0, 0.2, 0.4\}$. 14
A system in which individuals have more balanced preferences among two categories of information leads to a lower level of self-organization. The more agents are looking for specific information, the more a strong preference over one type of information favours the self organizing process. The reason is straightforward: let us consider a perfect balanced preference ($\beta = 0.5$) and a system with only two different types of information; under these assumptions an agent is indifferent to information of category $i$ and $j$. Thus, a forum is not able to make the difference between identical agents and this community gathers information of type $i$ and $j$. If the assumption of a perfect balanced preference is relaxed, agents can discriminate between the communities which share their strong and weak preferences and favour the former. In this situation, the selec-

---

<table>
<thead>
<tr>
<th>Mean of $I^A$ for $k \in [1, 36]$</th>
<th>$\beta = 0.6$</th>
<th>$\beta = 0.8$</th>
<th>$\beta = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>0.22</td>
<td>0.27</td>
<td>0.62</td>
</tr>
<tr>
<td>Uniform</td>
<td>0.17</td>
<td>0.25</td>
<td>0.43</td>
</tr>
<tr>
<td>Normal</td>
<td>0.13</td>
<td>0.18</td>
<td>0.38</td>
</tr>
</tbody>
</table>

---

5Everything equal, the utility function derived from the presence in a community which
tion process can occur and a self-organized equilibrium emerges. Even if under the assumption of bimodal preferences individuals have twice as much chance to be in a forum where at least one of his preference is shared, it makes the specialization of a forum over a particular interest more difficult.

The result of the previous section is confirmed whatever the value of $\beta$. It ensures that the structure of preference is not responsible for the superiority of heterogeneous distributions in term of self-organization.

If we take a closer look at the results of simulations, we observe that $I_A$ never reaches 1 for $\beta = 0.6$ and $\beta = 0.8$. The higher level of organization obtained is $I_A = 0.83$ and takes, for example$^6$, the form of

$$A_{80} = \begin{pmatrix} 0 & 0 & 0 & 0 & 200 \\ 0 & 0 & 200 & 0 & 0 \\ 200 & 0 & 0 & 0 & 0 \\ 0 & 0 & 200 & 0 & 0 \\ 0 & 0 & 0 & 0 & 200 \end{pmatrix}$$

In this configuration, agents who share a common strong interest are gathered in one forum, but every forum is not specialized in one particular type of information. The intuition for that imperfect segmentation is that some important contributors, in the early stage of the game, induce the double specialization of that community by locating in a community and by producing the amount of information $i$ and $j$, which lead to the emergence of virtual communities. This result suggests that large interest communities which gather people with diverse interests, like Amazon or Yahoo group, could be the consequence of a multimodal structure of preference.

5 On the relationship between contributions and switching costs

There is a growing literature in economic and management science which focuses on the role of trust, reputation and social capital on the organization of online communities. Management strategy can influence the building of trust on online auction system (Zhou, Dresner and Windle, 2007; Dellarocas, 2001 and 2003) and seeking status or reputation may influence the ability to contribute to virtual communities (Wasko and Faraj, 2005). As a result, rewarding best contributors by enhancing their social capital or create incentives for individuals who participate less in production of contents, can change individual switching costs by increasing or lowering the cost of leaving a community.

The framework of the model leads to two different designs on the relationship between contributions and switching costs. First, the model considers a positive correlation, which means that the best contributors have the highest switching cost. It seems to be a plausible assumption that agents who take time and

$^6$ this example is the result of a simulation with normal distribution of contributions, $k = 11$ and $\beta = 0.8$. 

shares his strongest interest will be higher than the one from the forum where his weaker interest is exchanged.
attention for the community are probably less willing to leave it. Then we will consider a negative correlation between contributions and switching costs. It could be the consequence of a management strategy which favours the mobility of best contributors, by allowing to keep their status or export their reputation out of the community.

graphic 12: variation of correlation with normal distribution

graphic 13: variation of correlation with uniform distribution
The positive correlation leads to higher level of self-organization, whatever
the distribution of contributions and switching costs. The more heterogeneous
the distribution is the more advantageous is the positive correlation. That con-
irms the importance of big contributors in the self-organizing process. This
result suggests that a strategy which favours the mobility (immobility) of the
least(most) important contributors favours the self-segmentation of the commu-
nity. As shown in graphics 12, 13 and 14, the benefit of the positive correlation
rises with the importance of the diversity of information (parameter $k$). As
agents are more adverse to the diversity of information, they need a higher level
of valuable information to stay in a particular community. In this case, best
contributors associated with strong switching costs act as an "anchor point" for
the rest of the community. If the model considers a negative correlation be-
tween $\phi_i$ and $r_i$, the system converge to the maximum entropy (as in graphic 4)
because the less important contributors never represent a sufficiently important
mass of information to attract the best contributors and allow the emergence of
self-organized communities.

An other point can be derived from this result. If we look at the graphic 12,
13 and 14, we observe that self-organized communities emerge for a low value
of $k$ when a negative correlation associates contributions and switching costs
and for a high value of $k$ when we consider a positive correlation. This result
suggests important suggestions for the management of community of interest.
Do we have to apply the same strategy to favour self-organized communities
when agents are looking for specific information (high $k$) and when they’re not
sensitive to the diversity of information (low $k$)?
For low values of \( k \), agents have a positive utility even if a community is not specialized in a particular interest. Following this condition, it is difficult to create the necessary disequilibrium which leads the evolutive process to begin and allow the emergence of segmented communities. That’s what could be done if switching costs of the most important contributors decreased. The informational disequilibrium produced by the migration of these agents create incentives to move for the less important contributors despite their high switching costs.

6 Conclusion

One of the main changes observed during the rise of the information and communication technologies, is the emergence of numerous web forums, discussion groups, sharing communities or online collaborative experiences. The efficiency and the success of these new social interactions is based on their capacity of gathering people who share a common interest. The purpose of this paper was to study how the diversity of contributions among individuals have an impact on the development of interest-based communities. Using an agent-based model, individuals, who differ by their preference for a type of information, interact by sharing information. These micro-interactions shape macro-structure which in turn creates the condition for the migration of agents and the segmentation of preference. Considering different statistical distributions of individual contributions, it has been shown that the most heterogeneous distribution, which follows an exponential law, leads to higher levels of self-organization. This assumption about the distribution of individual contribution is consistent with the empirical evidence that there is an important number of people who provide few information to the community and a small number of others who are in charge of the majority of information exchanged.

The assumption about multi-modal preferences leads to an imperfect specialization of the communities. Some of them attract more than one interest-based group and are consistent with the existence of communities which gather different types of information. This paper also studies two designed strategies for practitioners of interest communities, the first deals with the possibility to increase the switching costs of the most important contributors and the second does the opposite, combining high production of information and low mobility impediments. Two cases must be distinguished: when agents are not sensitive to the diversity of information (low value of \( k \)) a negative correlation between contributions and switching costs favours the emergence of communities where individuals share common interests. Conversely, when agents are looking for specific information (high value of \( k \)) the relative immobility of the best contributors is required to reach a high level of self-organization. This conclusion supports the fact that more attention has to be paid to the most important contributors when individuals can’t bear the information of their "non-preferred" category.

This paper represents a first step to understand the underlying process of the emergence and evolution of interest-based communities. Further studies have
to be performed in the empirical field. It would validate our theoretical findings that interest communities which share high specific contents are more able to advantage their core members than generalist communities and secondly that heterogeneous distribution of contributions may lead a community to share more homogeneous contents. Even if our theoretical framework is voluntarily simple it could be interesting to propose improvements in the cognitive capacities of our agents by allowing them, for example, to form belief about communities.

References

Adar E, Huberman BA (2000), Free Riding on Gnutella, First Monday, Vol 5, 10


Dang Nguyen G, Pénard T, (2005), Network Cooperation and Incentives within Online Communities, in Brousseau E and Curien N (Eds.), Internet and Digital Economics, Cambridge University Press.


