Sharing information in web communities

Gabrielle DEMANGE[1]

March 30, 2006

Preliminary version

Abstract

The paper investigates information sharing communities. There are no decreasing returns to scale and information is not rivalrous. A community may however be worth forming because it facilitates the interpretation and understanding of the posted information.

[1] EHESS, PSE, address 48 bd Jourdan, 75014 Paris, France e-mail demange@pse.ens.fr.

1 Introduction

Group structures on the Web such as peer-to-peer (P2P) systems aim at sharing various goods and disseminating information in a fully decentralized way. This paper focusses on information sharing communities. Information is non rivalrous and there are no decreasing returns to scale. Why then do communities form with a free but restricted access ? A basic rationale is related to the value of information. A tremendous quantity of information is posted on the Web, on blogs for instance. Search engines help Internet users to find pages that are relevant to their queries. In some situations however, all this information is useless. Consider for example a page in which an individual provides her opinion on movies. If she says that it is worth watching movie A, or that she prefers movie A to movie B, do I benefit from this knowledge ? If the peer is a critic, and I am pretty aware of her tastes, her judgment may be valuable to me (this is true whether her tastes are similar to mine or opposite). If instead I have no idea at all about her preferences, I learn nothing. In other words, how useful a person's statement is much depends on whether the identity/preferences of this person are known.

This suggests that a group may benefit from defining criteria on the peers who contribute to a platform in order to facilitate the understanding and the usefulness of the conveyed information. This raises several questions:

(i) How are such criteria determined ?
(ii) Can we predict whether some criteria are more apt to ensure some sort of efficiency, of stability ?
(iii) How does the value of information generated by a group depend on the way the individuals' opinions are transmitted, or aggregated ? Is it worth to weight them for instance ?

This paper builds a model that may allow to investigate these questions. I consider a situation in which search engines may not help. If there is an underlying quality ranking on which most users agree, a search engine can be seen as trying to find out this ranking through various criteria such as the number of clicks and the structure of the links (for an analysis of 'authoritative ' sources through hyperlinks structure see Kleinberg 1999). Here there is no such uniform ranking. To take an analogy with industrial organization models, I consider a horizontal differentiation model (as opposed to a vertical one) in which individuals are located on a circle (Hotelling 1929, Spence 1976). Any anonymous aggregation of preferences over the whole population gives a complete flat ranking. Thus, differences in the results of search engines should then be attributed to chance or to bias in search engines. This provides a rationale to form a community, which can be seen as a set of individuals over which aggregation is useful. Assuming that a leader/initiator of a community has some control on the members' characteristics, we perform some comparative statics on the chosen community.

There is a growing literature on the behavior of Internet users. Various algorithms have been proposed for detecting communities through the link structure (see e.g. the survey of Newman 2001). Instead this paper investigates why a community forms in the first place. Owing to the public good aspect, "bad" behavior -free riding and excessive overload of the platform on which peers operate- may generate difficulties. A large body of recent research discusses whether these difficulties are severe enough to call for the implementation of incentives schemes (see for example Feldman and all 2004 and Ng, Chiu, and Liu 2005). Not surprisingly, the contribution rate within a community plays a crucial role in our analysis.

The plan of the paper is the following. Section 2 sets up the model, Section 3 studies the formation of a community and proofs are gathered in the final section.

2 The model

I focus on situations in which individuals differ in their tastes. In particular, there is no unanimous ranking on the alternatives. The simplest way to represent such a situation is a horizontal differentiation model in which individuals are 'located' on a circle. An individual located at θ is called a θ-individual. An object, a movie or a restaurant for instance, is also characterized by a point on the same circle, denoted by t. A θ-individual who buys an object with location t derives a utility gain that is non increasing with the distance $d(\theta, t)$ between the locations of the individual and the object. The utility gain is given by $u(d(\theta, t))$ for some u identical for all individuals. Function u is non increasing and there is a unique threshold value d^* for which $u(d) > 0$ for $d < d^*$ and $u(d) < 0$ for $d < d^*$ and u is continuous, derivable except possibly at d^*. I shall often consider a simple function called *binary* in which individuals either enjoy or not consuming the object:[2]

$$\text{for some positive } g \text{ and } b, b \geq g : u(d) = g, d < d^*, u(d) = -b, d > d^*.$$

[2] The utility level at d^* does not matter in the sequel because it occurs with probability zero.

The society is uniformly distributed on the circle. If the characteristics of the objects are perfectly known, the set of individuals who benefit from buying a particular object is given by those located at a distance smaller than d^*. It will be assumed that d^* is equal to $\pi/2$. In that case, under perfect information, whatever an object's location, half of the people buy it.

I am interested in situations where information is incomplete. Objects are *a priori* uniformly distributed on the circle. Without any further information, to decide whether to buy a particular object, an individual computes the expected utility gain from buying it and compares this gain with 0. I assume risk neutrality or a weak form of risk aversion. Risk neutrality refers to the case where $u(d) = \pi/2 - d$. The weak form of risk aversion is the following : faced with the lottery of buying an object and its opposite with equal probability, the peer prefers not to buy: $u(d) + u(d - \pi) < 0$ for $d > \pi/2$. Since objects are uniformly distributed on the circle, a risk-neutral individual is indifferent between buying or not, and a risk-averse one does not buy.

2.1 Signals

There is some scope for information sharing. Individuals who have bought an object may post their opinion on it. Stating a detailed judgment is difficult. To account of this, signals are assumed to be limited. A signal s on an object takes two values, *yes* or *no*, which have the following interpretation: a peer who has bought the object recommends it or not. In opposite to some situations such as in financial markets, there is no benefit from sending a false signal. Thus, signals are assumed to be truthful : a θ-individual having bought a t-object sends *yes* if $u(d(\theta, t)) \geq 0$ and *no* if the inequality is reversed. (Introducing a known percentage of malicious individuals could be incorporated).

A community as defined more precisely in next section is represented by an arc. By convention, an arc $[\theta, \theta']$ designates the arc from θ to θ' going clockwise. Let us consider a signal[3] $\tilde{s}$ on an object from a member of community $[\theta, \theta']$. Anonymity is preserved, as often in P2P communities. As a consequence, the sender is considered as drawn at random from the community. In the sequel $s \in [\theta, \theta']$ refers to a signal sent by a member of community $[\theta, \theta']$.

The value of a signal can be analyzed from the viewpoint of Blackwell (1953). Signal $\tilde{s}$ is valuable to an individual if it enables him to make 'better' decisions in the sense that his expected payoff is increased. More precisely, the signal is used as follows. The joint distribution of $(\tilde{t}, \tilde{s})$ for a signal $\tilde{s}$ sent by a member

[3] A random variable is denoted by $\tilde{x}$, and its realization (when observed) by x.

of community $[\theta, \theta']$ can be computed. After learning the realized signal, peers revise their prior on the characteristic t according to Bayes' formula and decide to buy or not. Clearly a signal that does not change the prior on the object's location is useless. It is worth noting some consequences from the ignorance of the sender's location in the community.

- (i) A signal $\tilde{s}$ from the whole society is useless.
- (ii) A signal from a community smaller than the whole society may be useful : it changes the prior.
- (iii) Adding a signal may deteriorate current information

Point (i) is straightforward. A signal sent by an individual chosen at random in the whole group does not modify the prior, hence is not informative.

Point (ii) is also clear. Let us illustrate it with community $[-\pi/4, \pi/4]$. Each peer in the community sends *no* for an object located in $[3\pi/4, 5\pi/4]$. Thus the posterior density conditional on the signal being *yes* is null on that arc: the posterior clearly differs from the prior density (which is constant equal to $1/2\pi$).[4]

To show point (iii), consider a signal from a community reduced to a point, say 0. This means that the sender's preferences are known. The signal is informative: Conditional on a *yes* for example, the posterior density on $[-\pi/2, \pi/2]$ is $1/\pi$, which is the double of the prior, and null elsewhere. Add a signal stemming from $[\pi]$. The important point is that on the receipt of the two signals, it is not known which peer has sent which signal. The two signals here are always opposite to each other, hence give no information because the prior is not changed. Under the standard framework, adding a signal is never harmful, because it can simply be ignored. In this simple example, the new signal not only adds no information but even destroys the information conveyed by the first signal.

Finally note that the value of an informative signal to a person depends on his/her location. Consider community $[-\pi/2, \pi/2]$ and assume risk neutral individuals. For an individual located at one of the extreme point, the conditional gain of buying (or not buying) is zero whatever the realized signal (use the symmetry), as without signal: the individual does not benefit from information. For a peer located at the center instead, the value of the signal is positive: the conditional gain of buying (or not buying) is positive if the materialized signal is

[4] The posterior can easily be computed : given $\alpha \leq \pi/2$, the posterior $f(t|yes \in [-\alpha, \alpha])$ for $t \in [0, \pi]$ is $1/\pi$ for $t \leq -\alpha + \pi/2]$ (every peer says *yes*) 0 for $t \geq \alpha + \pi/2$ (every peer says *no*) and linear in between. The value for $t \in [-\pi, 0]$ is obtained by symmetry (see proof section).

yes, and negative if it is *no*. By following the recommendations, the peer increases his expected payoffs.

2.2 Communities

A community attracts only individuals who benefit from it. The role of contributors and users can a priori distinguished. Communities' members contribute to the content by providing information on the objects they have tested. They are also 'users' : they have access to the posted information through a fully decentralized mechanism such as Gnutella and Freenet. These mechanisms propagate queries through a P2P network without the need of a server. A query is sent to neighbors who provide an answer if they have one or otherwise pass the query to their own neighbors and so on until an answer is reached. [5] The system is anonymous and there is no distinction between members, who are recognized by an address only. Records can keep track of peers' behavior. Records on peers' contributions for instance allow the community to sustain some participation level. These records are not public and any sanction based on these records, exclusion for instance, is automatic.

In line with decentralized behavior, the signals are identically distributed across objects. Given a contribution behavior, the size of a community determines the probability of finding a recommendation for a particular object in reasonable time. Let us denote this probability by $P(\alpha)$ for community $[-\alpha, \alpha]$. P is assumed to be increasing and concave. For example, the process of successful search may follow a Poisson process with intensity proportional to the size of the community. This writes as $P(\alpha) = 1 - e^{-\lambda\alpha}$ for some positive parameter λ. This parameter reflects the contribution rate of the community members and the efficiency of the search mechanism. In addition, there is a cost of participation which includes the cost of search and the cost of contribution. Normalizing by the average number of requests, it is denoted by c.

We assume that there is no intrinsic motive to contribute such as altruism. Thus, in a viable community, as defined later, contributors are also users so as to draw some benefit. Even though there may be no direct cost (nor benefit) in allowing outsiders to be users, it may be worth restricting access to contributors simply to encourage them to contribute. A minimal amount of contributions can be asked for. Presumably, increasing contribution rate increases simultaneously

[5] See for example Kleinberg and Raghavan 2005 for a description of decentralized mechanisms and an analysis of the incentives to pass the information.

the success probability P and possibly c. Section 3.4 studies this point more closely.

Our purpose is to analyze how a community forms. In opposite to a set up in which a firm organizes the community, there is no clear criteria such as profit for choosing a community. A question is whether there are conflicting interests within the community members.

3 Community choice

3.1 Viable community

Anybody is free not to participate in a community. Thus a community is said to be *viable* if each of its members benefits from it, accounting for the failure of search and the participation cost. To cover their cost, members must benefit from following the recommendations stemming from the community. Here, following the recommendation means buying in the case of a positive signal and not buying in the opposite case. (Albeit possible, we exclude the possibility of community members who benefit from taking systematically the opposite action to the signal.)

Up to a rotation, we may focus on arcs centered at zero, of the form $[-\alpha, \alpha]$. The value α is called the size of the community. Let $U(\theta, [-\alpha, \alpha])$ denote the expected utility of a θ-individual who follows the recommendation of $[-\alpha, \alpha]$ assuming he has found an answer to his query. The overall ex ante utility level that determines if indeed a θ-member is willing to participate includes the probability of finding a recommendation and the participation cost: it is given by $P(\alpha)U(\theta, [-\alpha, \alpha]) - c$. Thus, community $[-\alpha, \alpha]$ is viable if

$$P(\alpha)U(\theta, [-\alpha, \alpha]) \geq c \text{ for any } \theta \in [-\alpha, \alpha]. \tag{1}$$

To determine U, note that an individual does not buy an object for which the signal is negative. Furthermore, a priori, the probability for a signal to be *no* (or *yes*) is one half, whatever the community (since, given an object at random, each individual says *no* with probability one half). This gives

$$U(\theta, [-\alpha, \alpha]) = \frac{1}{2}E[u(d(\theta, \tilde{t}))|yes \in [-\alpha, \alpha])] \tag{2}$$

We first determine conditions under which an individual indeed benefits from following the recommendation, that is conditions under which U is non negative.

We can limit attention to a community of size smaller than $\pi/2$. Indeed if two individuals are located at two opposite points, whatever the object, the sum of the distance of the object to the two points is π. Assume risk aversion. Since by assumption $u(d)+u(d-\pi) < 0$, taking expectation over objects implies that the sum of the expected utilities of the two opposite peers is negative : since surely one of the peers does not benefit from the community, the community is not viable. Under risk neutrality, direct computation shows that the community is at most of size $\pi/2$.

As we have just seen, a signal on an object from a community changes the assessment on its location. This change is more or less beneficial depending on the position of the peer in the community.

Lemma 1.

- *(i) Given* α*, utility* $U(\theta, [-\alpha, \alpha])$ *increases with* θ *on* $[-\pi, 0]$ *and decreases on* $[0, \pi]$*.*
- *(ii) Given* θ *in* $[-\pi/2, \pi/2]$ *utility* $U(\theta, [-\alpha, \alpha])$ *decreases with* α *on* $[0, \pi/2]$

Point (i) is rather natural given the symmetry. It says that the benefits derived from following the signal decrease with the distance to the center. In particular they are are the largest for individuals at the center.

According to point (ii), the expected value per signal is greater the smaller the community, that is the less uncertain the sender. This is easy to understand for the center. As α increases, objects that are at the distance larger than $\pi/2$ from the center are more likely to be recommended and those that are closer get less recommended. Thus increasing α is clearly harmful for the center since the objects that he dislikes are more recommended and the ones he likes are less recommended. For an individual who is not at the center, the distribution of signals is 'biased' with respect to his own preferences and as α increases some objects that he likes get more recommended. According to (ii) this benefit is however small enough so that increasing α is still harmful.

Another way to interpret property (ii) is in terms of an expert. A system in which an expert can send as many signals as the communities' members at the same cost per query (that is same $P(\alpha)$ and same c) is more efficient.

From point (i), the peers who get the lowest benefit are located at the extreme points $-\alpha$ or α in $[-\alpha, \alpha]$. Thus all individuals are willing to follow the signal provided this is the case for those at one of the extreme points. The viability condition can be written simply as

$$P(\alpha)U(\alpha, [-\alpha, \alpha]) \geq c$$

Assuming that the cost is not too high, there is surely a non empty interval $[\underline{\alpha}, \overline{\alpha}]$ of possible values for the viable size. Furthermore, in well behaved cases, the set of viable sizes is exactly an interval. This is the case for example with u binary.[6]

Property (ii) points out a trade-off faced by peers : increasing the size increases the probability of getting an answer but decreases the value of an answer. Next section analyzes the choice of the size.

3.2 Peers' choices

In practice, a 'leader' initiates a community and possibly defines criteria for accepting peers. The leader's optimal community size is given by the value α^0 that maximizes the payoff $P(\alpha)U(0, [-\alpha, \alpha])$. This will be the leader's choice provided it is viable. Similarly let α^θ denote the value that maximizes the payoff $P(\alpha)U(\theta, [-\alpha, \alpha])$, that is the preferred size of a θ-peer in a community centered at zero.

Proposition 1. *Assume a binary function. The leader's optimal size is less than the peers' optimal one :* $\alpha^0 \leq \alpha^\theta$*. There are two cases:*

1. *The leader's choice is his optimum* α^0 *if it is viable; in that case the community is closed to outsiders.*
2. *Otherwise, surely* α^0 *is too large to be viable, and the leader chooses the maximal viable size* $\overline{\alpha}$*.*

In case *1* some outsiders would benefit from joining the community (since individuals located at the extreme points achieve a strictly positive gain). In case *2* peers at the extreme of the community just cover their cost and there is no need to close the community. Instead all peers would benefit from an increase in the community size up to α^0 but no outsiders want to join.

Proposition holds more generally under elasticity conditions on the gain with respect to α. Let $V(\alpha) = U(\alpha, [-\alpha, \alpha])$ be the utility of an individual located at the extreme of the community. Proposition holds if

$$\frac{-V'}{V}(\alpha) \geq \frac{-U'_\alpha}{U}(0, [-\alpha, \alpha]) \geq \frac{-U'_\alpha}{U}(\theta, [-\alpha, \alpha])$$

According to the right inequality, the relative loss incurred by an individual due to an increase in the size is larger for the center. The left ratio is the relative

[6] It suffices that $P(\alpha)U(\alpha, [-\alpha, \alpha])$ be concave in α. Since P is assumed to be concave, it is satisfied if $U(\alpha, [-\alpha, \alpha])$ is concave. This holds with a binary function: as computed in the proofs $U(\alpha, [-\alpha, \alpha]) = \frac{1}{2}[g - (g+b)\frac{\alpha}{\pi}]$.

loss incurred by the extreme individual. Thus, the derivative of V with respect to α includes not only the variation of utility with respect to the size but also the variation due to the point of evaluation (that is $V' = U'_\theta + U'_\alpha(\alpha, [-\alpha, \alpha])$). This explains why the above nequalities can be compatible.

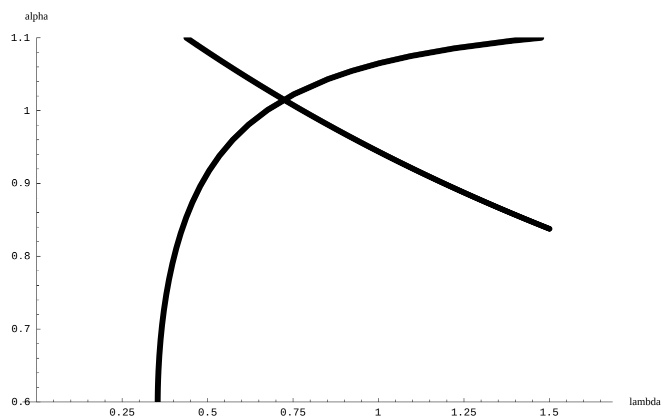

Fig. 1. $c = 0.1, b/g = 1.5$

Comparative statics Various policies can influence the probability of successful search. Next section studies policies that entice the peers within a community to contribute more (possibly at some cost for them). Other factors, the efficiency of the technology, the number of Internet users for instance, result in an exogenous change of the probability of success. This is easily illustrated with a Poisson process $P(\alpha) = 1 - e^{-\lambda\alpha}$. An increase in the population of Internet users other things being equal is represented by an increase in the parameter λ. Figure 1 depicts the maximal viable size $\overline{\alpha}$ (the increasing line) and the leader's optimum as a function of λ for $c = 0.1$ (the decreasing line). Recall that the leader's choice is the minimum of these two values. Hence the following configurations are obtained as λ increases: first there is no viable community for λ low enough, second the leader's choice is constrained equal to the maximal viable size, and third the leader can choose his optimum value.

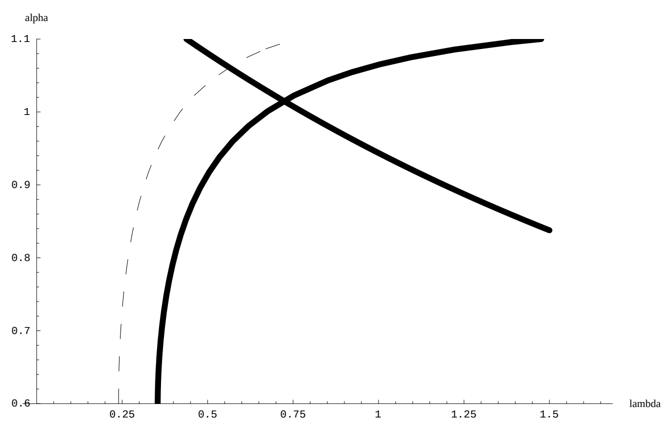

Fig. 2. $c = 0.1$ and $c = 0.07$, $b/g = 1.5$

Thus, increasing the population has different effects on the size and the information. Increasing the population within a community makes it more attractive to outsiders. When the community is constrained by viability, for intermediate values of λ, these outsiders are welcome. As a result the size is increased and information is more dispersed. When the community is closed instead, for a large enough λ, increasing the population allows the leader to choose a community restricted to peers whose tastes are more and more similar to his owns: the size decreases and information becomes more precise.

Advertising Assume that ads provide revenues that are proportional to the number of peers. Peers do not mind ads, which furthermore do not influence their preferences on the object on which they are searching information. Consider two alternative ways of distributing the revenues generated by ads.

First, the leader captures all the ads revenues. In that case he sets up a community that maximizes a combination of his own interests and the revenues. His choice is unchanged if the viability constraint binds. Otherwise, instead of choosing his own optimum, he chooses a larger size (between α^0 and $\overline{\alpha}$). (In particular, if he is greedy and mostly cares about revenues, his choice is always very close to the maximal viable size.) Thus, ads has the effect of making information less precise in the situation where it could be tailored to specific tastes (for large λ, $\overline{\alpha}$ is much larger than α^0).

A second possibility is to distribute revenues equally among peers. This amounts to diminish c and results in an increase in the maximal size: the leader's choice is more often equal to his optimal value, and in any case closer to it. In Figure 2, the maximal viable size is drawn for two distinct values of the cost: $c = 0.1$ (the plain increasing line) and $c = 0.07$ (the dashed line). The optimal leader's size is is independent of c (provided benefits cover the cost).

3.3 Voting

Under the leader's optimum, other peers would all like to increase the community (since $\alpha^\theta > \alpha^0$). This suggests some instability. In particular, communities with the same center and slightly larger than $[-\alpha^0, \alpha^0]$ are preferred by a majority (more precisely, for any qualified majority, one can find α larger than α^0 for which community $[-\alpha, \alpha]$ is preferred to $[-\alpha^0, \alpha^0]$). A different and probably more sensible way of changing the community is that some peers propose to accept newcomers who are close to their own tastes. (In view of the preceding discussion, only an increase of the community may be worth considering.) This amounts to change only one boundary at a time. Consider the voting game in which members vote on accepting new members on one side, that is on changing one of the boundaries, say increasing α keeping $-\alpha$ fixed.

Proposition 2. *Assume a binary function. At the leader's choice, there is no strict majority for changing only one side of the community.*

The impact of an accepted unilateral change depends on whether the community size is at the leader's optimum α^0 or at the upper bound $\overline{\alpha}$. Consider for example a proposal to accept individuals with characteristics at the right boundary. If the size is $\overline{\alpha}$, increasing the community on one side implies that some individuals on the other side will leave: accepting the proposal can only result in a rotation of the community which becomes $[-\overline{\alpha}+d\alpha, \overline{\alpha}+d\alpha]$. A strict majority of incumbents vote against the proposal : all peers in $[-\overline{\alpha}, d\alpha/4[$ either leave or are further away from the center without size effect and hence are made worse off (this follows from point (i) of lemma 1). If instead the community size is α^0, accepting the proposal results in $[-\alpha^0, \alpha^0 + d\alpha]$ (assuming $d\alpha$ small enough). Not only the community is enlarged but also the center is modified. Now the impact is unclear even for individuals on the negative side because there are two opposite effects: a possible benefit from an increase in the size and a loss from being further away from the center. According to Proposition 2 the loss outweighs the benefit in the binary case.

3.4 Enticing contribution

Participation/contribution is a crucial parameter of the model that in part determines the viability and the choice of the size of a community. The preceding section assumes a fixed participation. Here instead, the contribution rate within the community is a choice variable, in addition to the size. Let a minimum rate be asked for, rate denoted by λ. As said previously, a threshold can be implemented through records on peers' contributions. The individual's cost c is an increasing function $c(\lambda)$ of the peer's contribution. As a result, no peer will contribute more than the required threshold: his cost would increase with a null benefit since the impact of a single individual on the success probability is negligible. Let $P(\lambda, \alpha)$ denote the probability of success when *each* peer contributes λ. P is non decreasing and concave and c is non decreasing and convex. The maximal viable size depends on λ and is denoted by $\overline{\alpha}(\lambda)$.

Consider the situation in which the leader chooses both the size and the minimal contribution rate. Without constraint on viability, the leader's optimum is the value (λ, α) that maximizes $P(\lambda, \alpha)U(0, [-\alpha, \alpha]) - c(\lambda)$.

Proposition 3. *Assume a binary function. The leader's choice is*

1. *the leader's optimum values if the community is viable; the community is closed to outsiders. Other peers' would prefer to increase the size and decrease the participation.*

2. *a community with maximal viable size* $(\lambda, \overline{\alpha}(\lambda))$ *in which the choice of* λ *trades off the benefits from increasing contribution and the loss due to a smaller community size (surely* $\overline{\alpha}(\lambda)$ *decreases at the chosen value of* λ*.)*

Again there are two regimes. In case *1*, the leader is not constrained by viability conditions and his best choice is to close the community to some outsiders who would like to join. All peers would prefer to increase the size and to decrease contribution (but not by the same amount). In case *2*, the leader would benefit from an increase in contribution or from an increase in size. He faces however a trade-off because increasing contribution incites some peers at the extreme to leave thereby decreasing the size.

In conclusion, this paper considers a community as a cluster of individuals with similar preferences. The possible improvement in the value (understanding) of information determines the scope of a community. The analysis is clearly very preliminary. In particular, it has been restricted to a single community. A natural development is to understand the coexistence of several communities, possibly with overlap and with different objectives.

References

Blackwell, D. (1953), 'Equivalent Comparison of Experiments', *Annals of Mathematics and Statistics*, 24, 265-272.

Feldman, Michal, Christos Papadimitriou, John Chuang, Ion Stoica (2004) 'Free-Riding and Whitewashing in Peer-to-Peer Systems' 3rd Annual Workshop on Economics and Information Security (WEIS04).

Hotelling, Harold (1929) 'Stability in Competition'. Economic Journal, 39, 41-57.

Kleinberg J. (1999) 'Authoritative sources in a hyperlinked environment' Proc. 9th ACM-SIAM Symposium on Discrete Algorithms, 1998. Extended version in Journal of the ACM 46.

Kleinberg J. and P. Raghavan (2005) 'Query Incentive Networks'. Proc. 46th IEEE Symposium on Foundations of Computer Science.

Newman (2001) M. 'Detecting community structures in networks'. Physical Review E 69, 066133.

Ng W.-Y, D.M. Chiu, and W.K. Liu (2005) 'Club formation by rational sharing: Content, viability and community structure' *Springer's LNCS, Lecture Notes in Computer Science.*

Spence, A. Michael (1976) 'Product Selection, Fixed Costs and Monopolistic Competition'. Review of Economic Studies, 43, 217-235.

4 Proofs

To simplify notation denote $U(\theta,\alpha) = U(\theta, [-\alpha,\alpha])$ and $V(\alpha) = U(\alpha, [-\alpha,\alpha])$. Also because u may be discontinuous at $d = \pi/2$, we have to integrate functions discontinuous at some points. In that case the integral is computed as the sum of the integrals over the intervals of continuity.

Proof of Lemma 1 It is first useful to compute the posterior distribution on the receipt of a single signal from community $[-\alpha,\alpha]$. Let $f^\alpha(t)$ denote the density of t conditional on a *yes* from $[-\alpha,\alpha]$. By Bayes formula

$$f^\alpha(t) = P(yes \in [-\alpha,\alpha]|t)\frac{f(t)}{P(yes \in [-\alpha,\alpha])}$$

where f is the prior distribution of t $f(t) = 1/(2\pi)$ and $P(yes \in [-\alpha,\alpha]) = 1/2$. Given t, individuals with a type θ in $[t-\pi/2, t+\pi/2]$ say *yes* and others say no. Hence

$$P(yes \in [-\alpha,\alpha]|t) = \frac{1}{2\alpha}\int_{-\alpha}^{\alpha} 1_{[t-\pi/2,t+\pi/2]}(\theta)d(\theta)$$

Assume $\alpha \leq \pi/2$. In community $[-\alpha,\alpha]$, each member sends a signal *yes* for $t \in [\alpha-\pi/2, -\alpha+\pi/2]$, which we call the acquiescence zone, and each one sends no on $[\alpha+\pi/2, -\alpha+3\pi/2]$, which we call the refusal zone. Outside these two zones there is disagreement. By symmetry $f^\alpha(t) = f^\alpha(-t)$. Restricting to $t \geq 0$ one has :

$$f^\alpha(t) = \begin{cases} 1/\pi & \text{for } t \in [0, -\alpha+\pi/2] \text{ acquiescence} \\ \dfrac{\alpha+\pi/2-t}{2\pi\alpha} & \text{for } t \in [-\alpha+\pi/2, \alpha+\pi/2] \text{ disagreement zone} \\ 0 & \text{for } t \in [\alpha+\pi/2, \pi] \text{ refusal} \end{cases}$$

Hence the posterior distribution is doubled on the acquiescence zone, is null on the refusal zone, and linear in between. In the case of $\alpha > \pi/2$, there are no agreement zones and the expression for the disagrement zone holds over $[0,\pi]$.

From (2) the expected utility derived by a θ-individual in $[-\alpha,\alpha]$ assuming he follows recommendations from the community is

$$U(\theta,\alpha) = \frac{1}{2}\int_{-\pi}^{\pi} f^\alpha(t)u(d(\theta,t))dt \tag{3}$$

(i) We show that U is decreasing with θ for $\theta \geq 0$. The result extends by symmetry for $\theta \leq 0$. Note that $d(\theta,t)$ is equal to $\theta - t$ for $t \in [\theta-\pi,\theta]$ and $t-\theta$ for $t \in [\theta,\theta+\pi]$. Since all expressions are taken modulo 2π write $2U$ as

$$2U(\theta,\alpha) = \int_{\theta-\pi}^{\theta} f^\alpha(t)u(\theta-t))dt + \int_{\theta}^{\theta+\pi} f^\alpha(t)u(t-\theta))dt$$

The function is derivable with respect to θ (the only problem could be if a bound of integration corresponded to the discontinuity u, but this is not the case).

The derivative is $\int_{\theta-\pi}^{\theta} f^{\alpha}(t)u(\theta-t))dt + \int_{\theta}^{\theta+\pi} f^{\alpha}(t)u(t-\theta))dt$ Operating a change of variable $d=\theta-t$ in the first integral and $d=t-\theta$ gives the following expression for the derivative :

$$2\frac{\partial U(\theta,\alpha)}{\partial\theta} = \int_0^{\pi} [f^{\alpha}(\theta-d) - f^{\alpha}(\theta+d)]u'(d))dd$$

The term in brackets is always positive because the point $\theta-d$ is closer to the center than $\theta+d$ for θ positive. Since u is decreasing the derivative of U with respect to θ is negative, which gives the desired result.

(ii) We show that $U(\theta,\alpha)$ decreases with respect to α for θ in $[-\pi/2,\pi/2]$. By symmetry, the expectation for a θ-individual over the set of negative t is equal to that of a (-θ)-individual over the set of positive t. Thus it is enough to show that the utility derived from objects $t \geq 0$ for an individual located in $[-\pi/2,\pi/2]$ decreases with α. This utility writes

$$\int_0^{\pi} f^{\alpha}(t)u(d(\theta,t))dt.$$

Function $f^{\alpha}(t)$ is derivable with respect to α. Hence (separating into the two intervals of continuity of $u(d(\theta,t))$), the derivative exists and is given by :

$$\int_{-\alpha+\pi/2}^{\alpha+\pi/2} -\frac{\pi/2-t}{2\pi\alpha^2}u(d(\theta,t))dt.$$

It can be rewritten as

$$\int_{-\alpha+\pi/2}^{\pi/2} -\frac{\pi/2-t}{2\pi\alpha^2}[u(d(\theta,t)) - u(d(\theta,\pi-t)]dt$$

by splitting the interval of integration into two parts and operating a change of variable t into $\pi-t$ in the second part. Now it suffices to check that the term in square bracket is always non negative, or equivalently the distance between θ and t is smaller than the distance between θ and $\pi-t$ for any θ in $[-\pi/2,\pi/2]$ and t in $[-\alpha+\pi/2,\pi/2]$. For $\theta>0$, $d(\theta,t)=t-\theta$ or $\theta-t$ and $d(\theta,\pi-t)=\pi-t-\theta$. The difference $d(\theta,\pi-t)-d(\theta,t)$ is $\pi-2t$ or $\pi-2\theta$ which are both positive. For $-\theta<0$, $d(-\theta,t)=t+\theta$ and $d(-\theta,\pi-t)=\pi-t+\theta$ if $t\geq\theta$ or $\pi+t-\theta$ if $t\leq\theta$. The difference $d(-\theta,\pi-t)-d(-\theta,t)$ is again either $\pi-2t$ or $\pi-2\theta$, which are both positive. Thus the integral is negative, which proves point (ii). ■

Binary function. We show that the expected utility derived by a θ-individual, $\theta\in[-\pi/2,\pi/2]$, who follows a recommendation from community $[-\alpha,\alpha]$ is

$$U(\theta,\alpha) = \frac{1}{2}\big[g-(g+b)\frac{\alpha}{2\pi}(1+\frac{\theta^2}{\alpha^2})\big] \text{ for } \theta\in[-\alpha,\alpha] \qquad (4)$$

$$U(\theta,\alpha) = \frac{1}{2}\big[g - (g+b)\frac{\theta}{\pi}\big] \text{ for } \theta \notin [-\alpha,\alpha] \quad (5)$$

As explained above, we can compute the utility derived from objects $t \geq 0$ for an individual located in $[-\pi/2, \pi/2]$ who follows the recommendation of $[-\alpha, \alpha]$. The overall utility of a θ-individual is derived by using the symmetry, namely that the utility for a θ-individual on negative t is equal to that of a (-θ)-individual on positive t.

We shall use repeatedly that for t_1, t_2 in the (positive) disagreement zone one has

$$\int_{t_1}^{t_2} f^{\alpha}(t)dt = \frac{-1}{4\pi\alpha}(\alpha + \pi/2 - t)^2|_{t_1}^{t_2} = \frac{1}{4\pi\alpha}(t_2 - t_1)(2\alpha + \pi - (t_2 + t_1)) \quad (6)$$

Also a *yes* never comes from the refusal zone, which has not to be considered.

It is useful to distinguish between members and non members of the community. Consider first a θ-individual in the community. He achieves :

in the acquiescence zone $[0, -\alpha + \pi/2]$: $g\frac{\pi/2-\alpha}{\pi}$,

in the disagreement zone $[-\alpha + \pi/2, \alpha + \pi/2]$: benefit g for any object in $[-\alpha + \pi/2, \theta + \pi/2]$, and $-b$ on $[\theta + \pi/2, \alpha + \pi/2]$, which gives using (6)

$$\frac{1}{4\pi\alpha}[g(\alpha+\theta)(3\alpha-\theta) - b(\alpha-\theta)^2].$$

Using the symmetry property and collecting terms, the overall utility of a θ-peer is

$$g\frac{\pi - 2\alpha}{\pi} + \frac{1}{4\pi\alpha}[g(\alpha-\theta)(3\alpha+\theta) - b(\alpha+\theta)^2] + \frac{1}{4\pi\alpha}[g(\alpha+\theta)(3\alpha-\theta) - b(\alpha-\theta)^2]$$

or

$$g\frac{\pi - 2\alpha}{\pi} + \frac{1}{2\pi\alpha}[g(3\alpha^2 - \theta^2) - b(\alpha^2 + \theta^2)]$$

which gives (4).

Consider now an individual outside the community who follows the signal, and distinguish between a positive or negative location.

An individual located at $-\theta$, $\alpha < \theta < \pi/2$, achieves in the acquiescence zone g for objects with $0 \leq t \leq -\theta + \pi/2$ and $-b$ for $-\theta + \pi/2 \leq t \leq -\alpha + \pi/2$, which gives $g\frac{\pi/2-\theta}{\pi} - b\frac{\theta-\alpha}{\pi}$ In the disagreement zone he dislikes all objects hence achieves $-b\frac{\alpha}{\pi}$.

A individual located at θ with $\alpha < \theta < \pi/2$, gets the benefit g for all objects that he buys (since a t-object with $t > 0$ may be recommended only if $t < \alpha + \pi/2$ which is smaller than $\theta + \pi/2$). Thus he achieves $g\frac{\pi/2-\alpha}{\pi}$ from the acquiescence zone and $g\frac{\alpha}{\pi}$ from the disagreement zone.

Using the symmetry property and collecting terms, the overall utility of a θ-individual outside the community but distant from less than $\pi/2$ to the center is

$$g\frac{\pi/2-\theta}{\pi}-b\frac{\theta-\alpha}{\pi}-b\frac{\alpha}{\pi}+g\frac{\pi/2-\alpha}{\pi}+g\frac{\alpha}{\pi}$$

which gives (5).

Proof of proposition 1 Let α^{max} be the size for which $V(\alpha^{max})=0$. The size α^{θ} preferred by a θ-individual maximizes $P(\alpha)U(\theta,\alpha)$ with respect to α. We may restrict by symmetry to positive θ, and also to values for which U is positive, that is $0\leq\theta\leq\alpha\leq\alpha^{max}$. With u binary, the function U is concave with respect to α hence also PU. The derivative

$$P(\alpha)U(\theta,\alpha)\left[\frac{P_{\alpha}}{P}(\alpha)+\frac{U_{\alpha}}{U}(\theta,\alpha)\right]$$

is decreasing with α. At α^0, we have $\frac{P_\alpha}{P}(\alpha)=\frac{-U_\alpha}{U}(0,\alpha)$. Since PU is positive, $\alpha^0<\alpha^\theta$ holds if $\frac{U'_\alpha}{U}$ is increasing with θ positive. Set $k=(1+\rho)/2\pi$, Then

$$\frac{U_\alpha}{U}=-k(1-\frac{\theta^2}{\alpha^2})/[1-k(\alpha+\frac{\theta^2}{\alpha})]$$

The function is increasing in θ^2 if $\frac{k}{\alpha}<\frac{1}{\alpha^2}[1-k\alpha]$ or equivalently if $1-2k\alpha>0$. This condition is satisfied because $1-2k\alpha$ is proportional to $U(\alpha,\alpha)=V(\alpha)$, which is positive for $\alpha\leq\alpha^{max}$.

The optimal choice of the leader is the value of α that solves :

maximize $P(\alpha)U(0,\alpha)$ under the viability constraint $P(\alpha)V(\alpha)\geq 0$.

Let μ be the multiplier associated with the constraint. The first order condition is

$$P_\alpha U(0,\alpha)+PU_\alpha(0,\alpha)+\mu[P_\alpha V+PV'](\alpha)=0 \qquad (7)$$

If μ is null, the optimal choice is α^0 as expected. If μ is positive, the constraint binds and α^0 is not viable, i.e. outside the interval $[\underline{\alpha},\overline{\alpha}]$. Since by definition $\underline{\alpha}$ and $\overline{\alpha}$ are respectively the smallest and largest solution to $P(\alpha)V(\alpha)=c$, PV increases at $\underline{\alpha}$ and decreases at $\overline{\alpha}$. From (7), the derivatives of PU and PV are of opposite sign. If the latter derivative is positive, PV increases : the solution is $\underline{\alpha}$. We show that this is not possible in the binary case. For that it suffices to show the inequality[7]

$$\frac{-V'}{V}(\alpha)\geq\frac{-U'_\alpha}{U}(0,\alpha)$$

[7] Note that the derivative of $V(\alpha)=U(\alpha,\alpha)$ with respect to α includes not only the variation with respect to the size but also the variation due to the point of evaluation (the argument θ). This is why the inequalities $\frac{-V'}{V}(\alpha)\geq\frac{-U'_\alpha}{U}(0,\alpha)\geq\frac{-U'_\alpha}{U}(\alpha,\alpha)$ are compatible.

From (4) $2V(\alpha) = g(1-2k\alpha)$ which gives that

$$-\frac{V'}{V}(\alpha) = \frac{2k}{1-2k\alpha} > \frac{k}{1-k\alpha}$$

The term on the right is equal to $\frac{-U'_\alpha}{U}(0,\alpha)$, which gives the desired result. ■

Proof of proposition 2

Let community $[-\alpha,\alpha]$ be changed into $[-\alpha,\alpha+d\alpha]$. The utility of a θ-individual in $[-\alpha,\alpha+d\alpha]$ is the same as that of a $(\theta-d\alpha/2)$-individual in the community centered at zero $[-\alpha-d\alpha/2,\alpha+d\alpha/2]$. Hence a θ-individual will approve the change if $U(\theta-d\alpha/2,\alpha+d\alpha/2)$ is larger than $U(\theta,\alpha)$. We show that in the binary case it is not true for a majority of individuals if the community size is α^0, smaller than $\overline{\alpha}$ (the case where the leader's choice is $\overline{\alpha}$ is analyzed just after the proposition).

From the just above expression, the marginal impact of $d\alpha$ on the utility of a θ-individual is half $P_\alpha U + PU_\alpha - PU_\theta$ evaluated at (θ,α^0). In the binary case setting as before $k=(1+\rho)/2\pi$, one has:

$$U(\theta,\alpha) = U(0,\alpha) - \frac{g}{2}k\frac{\theta^2}{\alpha},\ U_\alpha(\theta,\alpha) = U_\alpha(0,\alpha) + \frac{g}{2}k\frac{\theta^2}{\alpha^2},\ U_\theta(\theta,\alpha) = -gk\frac{\theta}{\alpha}.$$

By definition of α^0 $(P_\alpha U+PU_\alpha)(0,\alpha^0)$ is null. Hence, collecting terms, $P_\alpha U + PU_\alpha - PU_\theta$ at (θ,α^0) is proportional to

$$\frac{\theta}{\alpha}\left[P[\frac{\theta}{\alpha}+2] - P_\alpha\theta\right]$$

For θ negative in $[-\alpha,\alpha]$, this expression is negative. By the concavity of U all individuals are worse off. This is also true for the center: no strict majority approves the change. ■

Proof of proposition 3 The optimal choice of the leader solves

$P(\lambda,\alpha)U(0,\alpha) - c(\lambda)$ under the viability constraint $P(\lambda,\alpha)V(\alpha) - c(\lambda) \geq 0$

Let μ be the multiplier associated with the constraint. The first order conditions are

$$P_\alpha U(0,\alpha) + PU_\alpha(0,\alpha) + \mu[P_\alpha V + PV'](\alpha) = 0 \qquad (8)$$

$$P_\lambda U(0,\alpha) - c'(\lambda) + \mu[P_\lambda V(\alpha) - c'(\lambda)] = 0 \qquad (9)$$

When μ is null, the viability constraint does not bind, and the leader can choose its optimal value. The same analysis as in proposition 1 yields that for the chosen value of λ other peers would like the size to increase. As for the contribution, since their utility level is smaller than the center, they would prefer a smaller contribution rate, that is $P_\lambda U(\theta, \alpha) - c'(\lambda) \leq 0$.

When μ is positive, we know that (8) implies in the binary case that α is set at the maximal viable size, $\overline{\alpha}(\lambda)$. As for the choice of λ, the first order condition (9) , $P_\lambda U(\alpha) - c'(\lambda)$ and $P_\lambda V(\alpha) - c'(\lambda)$ are of opposite sign. Since $U(\alpha) > V(\alpha)$ this implies that the former is positive : the leader would prefer to increase the contribution rate.