Sharing information in web communities

Gabrielle DEMANGE[1]

March 21, 2007

Abstract

The paper investigates information sharing communities. The environment is characterized by the anonymity of the contributors and users, as on the Web. It is argued that a community may be worth forming because it facilitates the interpretation and understanding of the posted information. The admission within a community and the stability of multiple communities are examined when individuals differ in their tastes.

Keywords value of information, communities, anonymity

Résumé

Le papier étudie la formation de communautés partageant des informations dans un environnement qui est caractérisé par l'anonymité des utilisateurs et contributeurs à l'instar du Web. L'analyse est basée sur la valeur de l'information au sens de Blackwell (1953). Nous examinons le choix des critères d'admission dans une communauté et la stabilité des communautés dans un modèle de divergence de goûts.

[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. Quite often, information is non rivalrous and returns to scale are not decreasing. Why then do communities form with a free but restricted access ? This paper argues that 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. As an illustration, consider 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. 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 preferences of this person are known. This suggests that defining criteria on the peers who contribute to a platform facilitate the understanding and the usefulness of the conveyed information. How criteria of access are determined ? Are they too restrictive or too loose, in a sense to be made precise ? This paper builds a simple model to investigate these questions.

The analysis relies on the value of communities in providing information to its members. There are individuals who regularly look for a piece of advice on a particular topic, on movies for instance. Individuals differ in their tastes. Due to these differences, search engines may not be helpful. The reason is that, given a search, an engine provides a ranking based on various criteria such as the number of clicks and the structure of the links. Thus, a search engine can be seen as an aggregator of preferences[2] that is valuable to users who share similar tastes and know it. In particular the observed behaviors of those who have already experimented the topic reveal to others the common ranking on which they all agree. In this paper instead there is no common ranking. To take an analogy with industrial organization models, I consider a horizontal differentiation model *à* la Hotelling (1929), as opposed to a vertical one. Furthermore, individuals are assumed to be located on a circle (Salop 1979), so that any anonymous aggregation of preferences over the whole population yields a complete flat ranking. In such a

[2] Dwork and al 2001 takes this point of view and borrows tools from social choice theory in order to study aggregation over different engines under a common underlying ranking.

situation, rankings provided by search engines can only be attributed to chance or to bias. This provides a rationale to communities : a community forms with members whose preferences allow for a useful aggregation.

The value of information as defined by Blackwell (1953) is our basic tool to investigate how communities form. The anonymity of contributors is shown to play a crucial role. More specifically, pieces of information posted by peers are valuable to other peers only if all share similar tastes. Furthermore, posted information without control on the contributors may not only be useless but also detrimental by introducing some noise in the information relevant to other peers. The admission rule in a community is therefore essential in determining the value that each peer derives from the information provided by the community's members. This leads us to analyze preferences over admission rules. Community's members do not fully agree on admission criteria owing to their differences in tastes, even if all of them benefit from the community. We analyze this divergence and assume that a leader/initiator of a community has some control on the members' characteristics. We perform comparative statics on the chosen community with respect to some parameters, participation rate for instance. We then study the coexistence of several communities, called a configuration, relying on stability concepts borrowed from cooperative game theory. Conditions are displayed under which communities are too large in most stable configurations. In such situations, external effects are predominant: Whereas each peer would like its community to increase and to accept newcomers, increasing the number of communities by decreasing their size would have an overall positive effect on welfare.

This paper is related to the growing literature on the behavior of Internet users. Various algorithms have been proposed for detecting communities through a link structure, as surveyed in Newman (2001), or to analyze 'authoritative' sources from the hyperlink structure as in Kleinberg (1999), and Gibson and all (1998). Here instead, we investigate in a specific context why a community forms in the first place. Another body of recent research is concerned with "bad" behavior due to the public good aspect of Internet: free riding and excessive overload of the platform on which peers operate. The main question is whether the generated difficulties are severe enough to call for the implementation of incentives schemes (see Feldman and all 2004 and Ng, Chiu, and Liu 2005 for example). The public good aspect is present in our model through the contribution rate within a community. Not surprisingly, this rate is shown to play an important role in the analysis.

The plan of the paper is the following. Section 2 sets up the model, Section 3 studies a single community, and Section 4 analyzes configurations of communities, their stability and efficiency properties. 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 object, a movie or a restaurant for instance, is also characterized by a point on the same circle. An individual located at θ is called a θ-individual and similarly an object located at t is a t-object. An individual who buys an object derives a utility gain that is non increasing in his distance to the object. The utility gain for a θ-individual who buys a t-object is given by $u(d(\theta, t))$ where $d(\theta, t)$ is the distance on the circle between θ and t and u is non increasing, identical for all individuals. To make the problem interesting, the utility gain is neither always positive nor always negative: there is a threshold value d^* for which $u(d) > 0$ for $d < d^*$ and $u(d) < 0$ for $d > d^*$. Furthermore function u is continuous and derivable except possibly at d^*. I shall often consider a simple function called *binary* in which individuals either enjoy or not consuming the object:[3]

$$\text{for some positive } g \text{ and } b : u(d) = g, d < d^*, u(d) = -b, d > d^*. \tag{1}$$

The society is uniformly distributed on the circle. If the characteristics of a particular object is perfectly known, the set of individuals who benefit from buying it is given by those located at a distance smaller than d^*. Thus, under perfect information, whatever an object's location, the same proportion p of the people buy it, where $p = d^*/\pi$. Under imperfect information on objects' characteristics, an individual forms some assessment on the location and decides whether to buy a particular object by comparing the expected utility gain from buying it with 0. It is assumed that objects are *a priori* uniformly distributed on the circle. Thus, without further information, the expected utility gain is computed according to this prior.

We take the following assumptions. First d^* is not larger than $\pi/2$, meaning that at most half of the objects are worth to an individual. Second we assume

[3] The utility level at d^* does not matter in the sequel because the probability of an object being distant of d^* to a person is null.

either risk neutrality, $u(d) = d^* - d$, or a weak form of risk aversion : faced with the lottery of buying two objects with equal probability, the peer prefers not to buy if the sum of the distance is $2d^*$: $u(d) + u(2d^* - d) < 0$ for $d < d^*$. With $d^* = \pi/2$ for example, a risk-neutral individual is indifferent between buying or not an object, and a risk-averse one does not buy (since objects are uniformly distributed on the circle). For a binary function, weak risk aversion is satisfied if the loss b in case of a 'bad' is larger than the gain g in case of a 'good'.

2.1 Signals

Under incomplete information, 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 the case of a binary function, this is not a restriction). In opposite to some situations such as financial markets for instance, there is no benefit from sending a false signal. So 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. (A known percentage of malicious individuals could be easily incorporated).

A community as in next section is represented by an arc. By convention, an arc $[\theta, \theta']$ designates the arc from θ to θ' going clockwise. Its *size* is defined by $(\theta + \theta')/2$. Let us consider a signal[4] $\tilde{s}$ 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 of community $[\theta, \theta']$ can be computed. After learning the realized value of a 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. The ignorance of the sender's location in the community has the following consequences.

– (i) A signal $\tilde{s}$ from the whole society is useless.

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

– (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 as in figure 1 for $d^* = \pi/2$. Each peer in the community sends *no* for an object located in $[alpha + \pi/2, -\alpha - \pi/2]$. 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 $1/2\pi$)?[5]

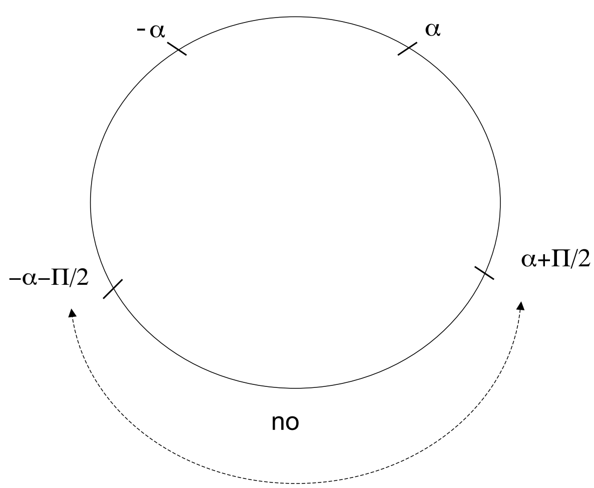


Fig. 1. $d^* = \pi/2$

To show point (iii), consider a signal from a community reduced to a point, say 0, so that the sender's preferences are known. The signal is informative:

[5] The posterior density conditional on a *yes* $\in [-\alpha, \alpha]$, $\alpha \leq \pi/2$, can easily be computed (see the proofs' section) : on $[0, \pi]$, it is equal to $1/\pi$ 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 density on $[-\pi, 0]$ is obtained by symmetry.

Conditional on a *yes* for example and taking again $d^* = \pi/2$, the posterior density on $[-\pi/2, \pi/2]$ is $1/\pi$, which is the double of the prior, and null elsewhere. Add a signal sent 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, which are always opposite to each other, give no information because the prior is not changed. In this simple example, the new signal not only adds no information but also destroys the information conveyed by the first signal. In contrast, under the standard framework, adding a signal is never harmful because it can simply be ignored. The difference is due to anonymity. in our framework, adding a signal introduces an additional source of randomness: the identity/preferences of the sender.

Finally, note that the value of an informative signal to a person depends on his/her location. Take $u(d) = \pi/2 - d$ and note that without information, an individual achieves a null expected gain whatever his decision. Consider community $[-\pi/2, \pi/2]$. For an individual located at one of the extreme points, by symmetry, the expected gain from buying conditional on the materialized signal being *yes* is null, as without signal : the individual does not benefit from information. For a peer located at the center instead, the conditional gain derived from buying is strictly positive if the materialized signal is *yes*, and negative if it is *no*. By following the recommendations, that is buying only upon receiving *yes*, the peer achieves a strictly positive expected payoff : the value of the signal is positive.

2.2 Communities

In a community, the role of contributors and users can a priori be distinguished. Contributors add to the content by providing information on the objects they have tested while users have access to the posted information. Here the set of contributors and users will be identical. This is induced by the following assumptions. First we assume that there is no intrinsic motive to contribute such as altruism. Thus, for a community to be 'viable' as defined in next section, contributors are also users so as to draw some benefit. Furthermore, even though there may be no direct cost (nor benefit) in allowing users not to contribute, it may be worth restricting access to contributors simply to encourage them to contribute. In that case users and contributors coincide. This can be implemented by an anonymous system as follows.

Users 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.[6] The system can be anonymous by recognizing members by an address only. Records, which are not public, can keep track of peers' behavior. Sanctions such as exclusion are based on these records and automatic. Records on peers' contributions for instance allow the community to sustain some participation level. In particular, they can be used to exclude users who do not contribute.

We consider a technology characterized by two data: the probability of success and an individual cost. In line with decentralized 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 a community of size α (such a community is up to a rotation $[-\alpha, \alpha]$). 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: Probability writes as $P(\alpha) = 1 - e^{-\lambda\alpha}$ for some positive parameter λ which reflects the contribution rate of the community members and the efficiency of the search mechanism.

The participation cost includes the cost for searching and contributing. Normalizing by the average number of requests, it is denoted by c. It is likely to be small and does not play an essential role in the analysis.

The probability P and the cost c are first assumed to be given. A minimal amount of contributions can be asked for to increase the success probability P. Section 3.4 investigates 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 the maximization of profit for choosing a community. Even if all community members benefit from it, they may have conflicting views about its scope.

3 Community choice

3.1 Viable community

Anybody is free not to join a community. 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.

Consider a community of size α and an individual whose distance to the center is θ. Let $U(\theta, \alpha)$ denotes his expected utility per signal conditional on having

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

an answer to a query. The criteria that determines whether he is indeed willing to participate takes the form $P(\alpha)U(\theta,\alpha) \geq c$. Thus the viability condition takes the form

$$P(\alpha)U(\theta,\alpha) \geq c \text{ for any } \theta \in [0,\alpha]. \tag{2}$$

To cover their cost, members must benefit form receiving a signal, that is U must be positive. The benefit from receiving a signal is drawn by following the recommendation, that is from buying the object in the case of a positive signal and not buying it in the opposite case. (Albeit possible, this excludes the possibility of community members who benefit from taking systematically the opposite action to the signal.) Let us determine $U(\theta,\alpha)$. Note that, whatever community, the a priori probability for a signal to take value *yes* is equal to p (which is $d^*/2\pi$), since, given an object at random, each individual says *yes* with probability p. Furthermore, on the reception of a negative signal, an individual achieves a conditional null payoff since he does not buy the object. Up to a rotation, we can consider arcs centered at zero, of the form $[-\alpha,\alpha]$. This gives

$$U(\theta,\alpha) = pE[u(d(\theta,\tilde{t}))|yes \in [-\alpha,\alpha])]. \tag{3}$$

The viability condition implies that we can limit attention to a community of size smaller than d^*. To see this, let two individuals be distant of $2d^*$ or more. Whatever the object, the sum of its distance to the two individuals is at least $2d^*$. Under risk aversion, surely one of the peers does not benefit from the community: since $u(d) + u(2d^* - d) < 0$, taking expectation over objects implies that the sum of their utility levels is negative. Thus these two individuals distant of $2d^*$ or more, cannot belong to the same viable community (even for a null cos c). Viable sizes are smaller than d^* (possibly equal to d^* under risk neutrality and null cost).

Some simple properties of U are useful to analyze further viability. As we have just seen, a signal sent by a community on an object changes the assessment on its location. This change is more or less beneficial to a peer depending on his position in the community. Proposition 1 below states how the expected value per signal varies with the position θ and the size α.

Proposition 1.

– *(i) Given*[7] *While we assume that a user of a community is also a member α, $\alpha \leq d^*$, utility $U(\theta,\alpha)$ decreases with the distance θ to the center, $\theta \leq d^*$.*

[7] The utility of an individual outside the community who follows the recommendation can be derived, so we do not restrict θ to be lower than α.

– *(ii) Given θ* $\theta \leq d^*$, *utility* $U(\theta, \alpha)$ *decreases with* α *on* $[0, d^*]$.

Point (i) is natural given the symmetry. It says that the expected benefits derived from following a 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 whose distance to the center is more than d^* are more likely to be recommended and those closer to the center 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. This result obtains because, as α increases, the distribution of the distance to a peer of the recommended objects becomes riskier in the sense of first order stochastic dominance. An implication of property (ii) is the superiority of an expert 'everything equal'. More precisely, a system in which an expert sends as many signals as the communities' members at the same total cost makes every peer better off.

Proposition 1 is helpful to analyze the size of a community. First, it simplifies the viability condition since, according to point (i), the peers who achieve the lowest benefit are located at the extreme points of the community. Thus all peers in a community are willing to participate if those at the extreme are willing to do so. Let us denote by $V(\alpha)$ the expected utility per signal for a peer located at an extreme point of a community of size α:

$$V(\alpha) = U(\alpha, \alpha) \tag{4}$$

The viability condition can be written simply as

$$P(\alpha)V(\alpha) \geq c. \tag{5}$$

Note that V is positive for α small enough. Hence, under a small enough cost, the set of viable sizes is non empty.

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.

To analyze the choice of community's size, we shall take throughout the following assumptions.

A0 (concavity assumption) the functions $U(\theta, \alpha)$ and $V(\alpha)$ are logconcave with respect to α.

A1 (elasticity conditions)

$$\frac{-V'}{V}(\alpha) \geq \frac{-U'_\alpha}{U}(0,\alpha) \geq \frac{-U'_\alpha}{U}(\theta,\alpha). \tag{6}$$

Under A0, the set of viable sizes is a nonempty interval $[\underline{\alpha}, \overline{\alpha}]$ under a low enough cost c (because PV is log concave as a product of logconcave functions).

Let us interpret the elasticity assumption A1. The second inequality of (6) says that the relative loss incurred by a peer due to an increase in the size is larger for the center. The first inequality says that these relative losses are all smaller than the relative decrease in the utility of an individual located at an extreme. This decrease includes not only the variation due to the size but also the variation due to the position (because $V'(\alpha) = [U'_\alpha + U'_\theta](\alpha,\alpha)$. Since the latter is negative (U decreases with the distance to the center) inequalities (6) are compatible. For example, assumptions A0 and A1 hold for a binary function, as shown in the appendix.

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 $P(\alpha)U(0,\alpha)$. This will be the leader's choice provided it is viable. Similarly let α^θ denote the value that maximizes the payoff $P(\alpha)U(\theta,\alpha)$, that is the preferred size of a θ-peer in a community centered at zero.

Proposition 2.

1. *The leader's optimal size is less than the peers' optimal one :* $\alpha^0 \leq \alpha^\theta$.
2. *The leader's choice is*
   (a) *either the leader's optimum* α^0 *if* $P(\alpha^0)V(\alpha^0) > c$; *in that case some outsiders would achieve a positive payoff by joining but the community is closed to them.*
   (b) *or the maximal viable size* $\overline{\alpha}$.

Thus, the leader's choice is the minimum of α^0 *and* $\overline{\alpha}$.

Point *1* makes precise the direction of possible disagreements with the leader: peers all prefer a larger size than the leader. Thus, disagreement occurs in case (a) where the leader can choose his preferred size because it is viable. Furthermore, since the viability condition is strict, $P(\alpha^0)V(\alpha^0) > c$, individuals at the extreme achieve a positive payoff. Hence by continuity of the payoffs, close enough outsiders would achieve a positive payoff by joining. Even if they wanted to join (which may depend on the other opportunities they face, as we shall see in next section) they are not allowed to do so. the community is closed. According[8] to point (b), it is never the case that the leader's optimum is not viable because it is too small. Instead all peers, including the leader, would benefit from an increase in the community size up to α^0. However no outsiders want to join and peers at the extreme of the community just cover their cost.

We shall perform some comparative exercises and illustrate them with a binary function as given by (1). In that case one has (see appendix)

$$U(\theta, \alpha) = pg[1 - k(\alpha + \theta^2/\alpha)] \text{ with } k = \frac{(1 + b/g)}{4d^*} \tag{7}$$

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 or the number of Internet users for instance, result in an exogenous change of the probability of success. The impact of such a change 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 2 depicts the maximal viable size $\overline{\alpha}$ (the increasing line) and the leader's optimum α^0 (the decreasing line) as a function of λ for a binary function. Since the leader's choice is the minimum of these two values, the following configurations obtain 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.

This can be explained as follows. 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. Instead, when the community is closed, 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. Thus, increasing the population has different effects on the leader's choice depending on whether this choice is constrained or not. As the precision of information within a community is negatively related to its size, the impact of the contribution rate on the size directly translates into an impact on the precision of information : as the contribution rate increases, information is first made less precise (but the higher chance of getting one compensates the loss) and then more and more precise.

[8] In the boundary case, the leader's optimum α^0 is just viable, equal to $\overline{\alpha}$.

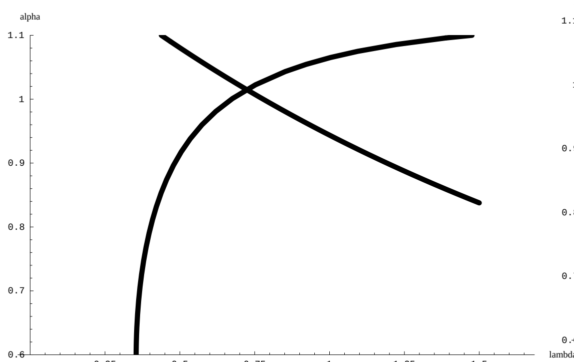


Fig. 2. $c = 0.1$ and $b/g = 1.5$

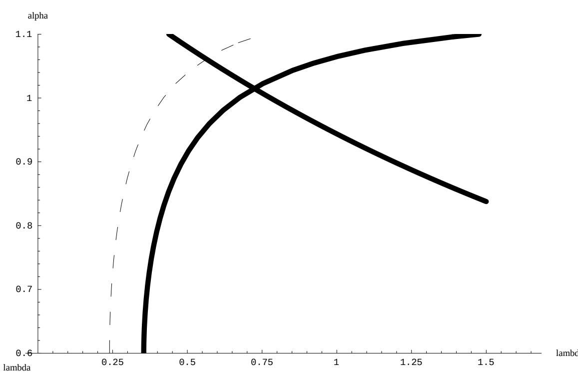


Fig. 3. $d^* = \pi/2$, $c = 0.1$ and $c = 0.07$

Advertising Ads provide revenues that may change the leader's choice criteria. To simplify, assume that peers do not mind ads and that ads do not influence their preferences on the object on which they are searching information. Let the revenues generated by ads be proportional to the number of peers and consider two alternative ways of distributing them.

First, the leader captures all 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, α^0, he chooses a larger size (between α^0 and $\overline{\alpha}$). The more he cares about revenues, the closer his choice to the maximal viable size. As a result, information is less precise. The effect can be substantial for large λ because the maximal viable size $\overline{\alpha}$ is large and α^0 is small. Whereas a community could be tailored to his specific tastes, the leader may choose a loose criteria so as to capture ads revenues.

Second, ads revenues are distributed equally among peers, which amounts to diminish cost c. This results in an increase in the maximal viable size and leaves the optimal leader's size unchanged. Hence, the leader's choice is closer and more often equal to his optimal value. 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).

3.3 Voting

Consider a community with size α^0, the leader's optimum. Peers who are not located at the center would all like to increase the community (since $\alpha^\theta > \alpha^0$). Thus, a community with the same center and slightly larger than α^0 is 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 $[-\alpha^0, \alpha^0]$ by the required majority). This suggests some instability. A different and 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.

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 $[-\overline{\alpha}+d\alpha, \overline{\alpha}+d\alpha]$. If accepted, peers in $[-\overline{\alpha}, d\alpha/4[$ either leave or are further away from the center (with no change in size) and hence are made worse off (this follows from point (i) of Proposition 1) : A strict majority of incumbents vote against the proposal. 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 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 next Proposition, the loss outweighs the benefit under the following assumption A2, which is satisfied for a binary function.

A2 $[U_\alpha + U_\theta](\theta, \alpha^0)$ decreases with θ in $[0, \alpha^0]$

Proposition 3. *Assume A2. At the leader's choice, there is no strict majority for changing only one side of the community.*

3.4 Enticing contribution

The success probability partly determines the viability and the choice of the size of a community. Whereas in previous section, it was taken as exogenous, we consider here that it is influenced by the peers' contribution rates. As said previously, a minimum rate can be asked for, implemented through records on peers' contributions. This section analyzes this choice.

We assume that the peer's participation cost c is an increasing function of his contributions. As a result, no peer will contribute more than the minimum required rate: his cost would increase with a null benefit since the impact of a single individual on the success probability is negligible. This is a standard effect in public good provision. Thus, given the minimum required rate λ, let $P(\lambda, \alpha)$ and $c(\lambda)$ denote respectively the probability of success when *each* peer contributes λ and the incurred individual cost. P is non decreasing and concave and c is non decreasing and convex. The maximal viable size now depends on λ; it 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 of (λ, α) that maximizes $P(\lambda, \alpha)U(0, \alpha) - c(\lambda)$.

Proposition 4. *The leader's choice is*

(a) 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.
(b) a community with maximal viable size $\overline{\alpha}(\lambda)$ for the chosen rate; the choice of λ trades off the benefits from increasing contribution and the loss due to a smaller community size (i.e. $\overline{\alpha}(\lambda)$ decreases at the chosen value of λ.)

We find the two regimes in which, given the chosen participation rate, the choice of the size is dictated by the same considerations as in the previous section. As for the contribution rate, note that the marginal benefit from increasing the contribution rate is decreasing with the distance to the center : it is given by $P_\lambda U(\theta, \alpha) - c'$, which decreases as U with respect to the distance θ. Hence, surely, at the chosen contribution rate, the leader's marginal benefit is nonnegative: otherwise a Pareto improvement within the community would be found by reducing the rate. Thus, in case (a), where the leader is not constrained by viability, all peers would prefer to increase the size and to decrease the contribution rate. In case (b), the leader would benefit from an increase in the contribution rate and from an increase in size. He faces a trade-off because increasing contribution incites some peers at the extreme to leave thereby decreasing the size.

4 Configurations of communities

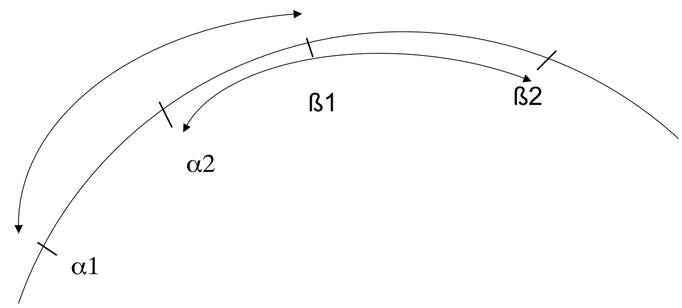
We have so far considered a single community. Our aim in this section is to analyze the coexistence of several communities. We shall assume that an individual is willing to join one communiy at most. Individuals are free to join whatever community that is open to them, or not to join any. Thus, faced with a set of proposed communities, they pick up their preferred choice. We look at equilibrium situations under which individuals correctly expect the value they derive from a community. As explained below, without coordination failure, it follows that communities can be described by non overlapping arcs, meaning that the arcs do not intersect except possibly at their boundaries. In that case the probability of success is still given by function P.[9]

[9] As a simple example of coordination failure let a single viable community be proposed. Viability is computed under an assumption on the participation embodied in the function P. If nobody joins the proposed community, the probability of success drops down to zero and it is indeed rational not to join.

Justifying non overlapping commuities Let two proposed communities with intersecting sets of characteristics: (α_1, β_1) and (α_2, β_2) with $\alpha_2 < \beta_1$. We have to distinguish the members in the community, denoted by C_1 and C_2, from the sets of characteristics. Individuals whose characteristic belongs to the intersection (α_2, β_1) can choose either of the two communities. They base their choice on the comparison between the distributions of the signals and the expected success probability (In the previous section, P is the success probability under the participation of all individuals in the community. If communities overlap, P gives an upper bound to the success probability). These distributions in turn depend on who joins each community, and cannot be directly derived from the admissible set of characteristics. If for example all individuals in the intersection choose C_1, the distributions of the signals for C_1 and C_2 are as if the proposed communities were (α_1, β_1) and (β_1, β_2). Let us consider the distribution of objects conditional on a recommendation from C_1. The distribution depends on who joins the community. Assume its density has a single peak which is in (α_1, α_2). Using a similar argument as in the proof of Proposition 1 gives that the utility for a signal stemming from C_1 decreases with the distance to the peak. Taking the similar assumption for C_2 gives that if a θ-peer weakly prefers C_1 to C_2 then all peers in the intersection who are closer than θ to C_1's center strictly prefers C_1 to C_2 (and similarly exchanging C_1 and C_2). It follows that there is a threshold value θ^* for which all individuals in (α_2, θ^*) choose C_1 and all in (θ^*, β_1) choose C_2. This means that the two communities C_1 and C_2 can be replaced by $(\alpha_1, \theta*)$ and $(\theta*, \beta_2)$. Furthermore this behavior ensures that the hypothesis on the monotonicity of the densities is fulfilled.

Remark Due to coordination problems, we cannot derive from rational behavior this hypothesis on single peaked densities. Assume for instance that for some θ^* all individuals in the intersection (α_2, β_1) choose C_2 if their θ is smaller than θ^*, and choose C_1 if it is larger than θ^*. Thus the characteristics of members of C_1 are in two disjoint intervals, (α_1, α_2) and (θ^*, β_1), and those in C_2 are in (α_2, θ^*) and $(\beta_1, \beta_2,)$. Because peers contribute to the signals of their community, we cannot exclude non monotone densities. This can be qualified as a coordination failure, because everybody would be better off if there was an exchange of peers of equal measure between C_1 and C_2 : this would keep the probability of success constant for both communities but would improve the quality of the information to everybody.

From now on, we consider configurations that are composed with arcs that do not overlap (i.e. they intersect at most at one of their extreme points). A configuration is said to be *full* if every individual belongs to a community : it



is composed of consecutive arcs that fill the entire circle (individuals who are located at a common boundary of two communities are negligible so that there is no need to specify to which community they belong if they are indifferent between both). I focus on symmetric configurations. A symmetric configuration is given by n non overlapping communities of equal size α, with n at most equal to π/α. For n exactly equal to π/α the configuration is full, and otherwise, there are gaps between any two neighbors.

Our aim is to determine which configurations may last, and which ones are optimal in a sense to be made precise. various stability tests can be contemplated. These tests depend on the proposals to deviate that are possible and the criteria according to which such proposals are successful.

4.1 Stability

The stability of a configuration requires that no proposal of change is accepted. Here, a proposal is said to be *accepted* if all members in the *new* community are better off than under the standing configuration. This notion is basically the blocking condition of cooperative game theory. Observe that insiders who are not included in the proposal have no say. It is appropriate in a setting where there are no 'property rights'.

Let us describe formally stability notions. Denote by $\bar{u}(\theta)$ the utility level achieved by a θ-individual at the standing configuration :

$\bar{u}(\theta) = P(\alpha)U(d, \alpha)$ if the individual belongs to a community and d is his distance to the center

$\bar{u}(\theta) = 0$ if he does not belong to any community.

I consider here proposals to form a new community that contains a community's leader, who can be assumed to be located at zero. Let $U(\theta, [-\alpha', \alpha''])$ denote the utility of a θ-peer in community $[-\alpha', \alpha'']$ (which is equal to $U(\theta - m, \alpha)$ where m is the middle of the community and α its size). A proposal to form community $[-\alpha', \alpha'']$ is accepted if

$$P([\alpha' + \alpha'']/2)U(\theta, [-\alpha', \alpha'']) > \bar{u}(\theta) \text{ for each } \theta \text{ in } [-\alpha', \alpha''].$$

Acceptance obtains if all insiders who are included in the proposal benefit from it *and* outsiders, if any, are all willing to join. Observe that benefits are evaluated under the assumption that everybody joins, which is indeed rational if the proposal is accepted.

We shall distinguish stability with respect to proposals that are larger in size than the existing community from those that are smaller. A symmetric configuration with communities' size α is *stable against enlargement* if there is no proposal $[-\alpha', \alpha'']$ with size $[\alpha' + \alpha'']/2$ larger than α that is accepted. Similarly it is *stable against reduction* if there is no proposal $[-\alpha', \alpha'']$ with size smaller than α that is accepted.

Stability against enlargement. Let α^{ext} be the maximum of PV, the payoff to an extreme individual. Observe that α^{ext} is less than α^0, thanks to assumptions A0 and A1 (the function $\log(PV)$ is concave and is decreasing at α^0 since $\frac{P'}{P}(\alpha^0) + \frac{V'}{V}(\alpha^0) \leq \frac{P'}{P}(\alpha^0) + \frac{U'_\alpha}{U}(0, \alpha^0) = 0$).

Proposition 5.

A symmetric configuration with gaps is stable against enlargement if and only the communities size is larger than the leader's choice.

A full symmetric configuration with communities size larger than α^{ext} (the maximizer of PV) is stable against enlargement.

For a configuration with gaps, it is easy to understand why stability requires communities size not to be smaller than the leader's choice. Recall that the leader's choice is $min(\alpha^0, \overline{\alpha})$. If $\alpha < min(\alpha^0, \overline{\alpha})$, consider a slight enlargement with the same leader. The new members, who do not belong to any community, are willing to join (because $\alpha < \overline{\alpha}$) and furthermore all insiders benefit from such an enlargement (because $\alpha < \alpha^0$ and using Proposition 2).

For a full configuration, the external stability condition is much weaker. Since communities are adjacent, outsiders, whatever close, achieve a positive payoff at the standing configuration. Thus outsiders are more difficult to attract in a full configuration, which explains why smaller sizes than the leader's choice are compatible with external stability.

Stability against reduction.

Proposition 6. *A symmetric configuration is stable against reduction if the communities' size is smaller than some value α^{int}, where $\alpha^{int} > \alpha^0$.*

The distinction between configuration with or without gaps is not relevant as far as reduction is concerned. The reason is that stability can be checked by considering only proposals to reduction that are included in the community, as shown in the proof. Such a proposal consists of excluding insiders too distant from the center and outsiders play no role.

Whereas stability to reduction requires communities not to be too large, stability to enlargement requires them to be large enough. From the previous results, both stability conditions are compatible.

Corollary. *A symmetric configuration with communities' size equal to the leader's choice is stable both against enlargement and reduction.*

4.2 Efficient configurations

Our objective here is to assess the efficiency properties of a stable configuration in particular of the one determined by leader's choices.

To simplify the analysis, individual cost is taken to be nil. In a setup as here in which utility is not transferable, there is not a unique welfare criteria to evaluate the efficiency properties of configurations. In line with our framework, we restrict ourselves to *anonymous* criteria. Such a criteria is characterized by an increasing scalar function Φ, that weights a utility level v by $\Phi(v)$: the total welfare reached by a community of size α is given by

$$\int_{-\alpha}^{\alpha} \Phi[P(\alpha)U(\theta,\alpha)]d\theta \tag{8}$$

where function Φ is assumed to be continuous. A concave function Φ represents an aversion to inequality, and at the opposite a convex one puts relative more weight on large utility levels. The average welfare value per member (that is the value given by (8) divided by 2α) lies between the minimum and the maximum of the range of $\Phi(v)$ over the community. The minimum is reached for a Rawlsian criteria, obtained as a limit of increasingly concave functions Φ, and the maximum is reached by taking a limit of increasingly convex functions.

The total welfare at a configuration is the sum of the welfare within each community. Consider a symmetric configuration with community size α. We may restrict to viable sizes, and to the maximal number communities. We shall neglect integer problems and take that there are $n = \pi/\alpha$ communities. Thus total welfare is given by

$$W(\alpha) = \frac{\pi}{\alpha} \int_{-\alpha}^{\alpha} \Phi[P(\alpha)U(\theta, \alpha)] d\theta. \tag{9}$$

An optimal configuration is given by the value α^{Φ} that maximizes W. To understand its determinants, observe that there are two effects as the size of each community increases: a pure size effect within a community as if everybody could stay at the same distance to the leader, and a position effect, due to rearrangement of the communities following a reduction in their number. The position effect increases the distance to the leader on average, which is detrimental. The size effect may or may not be beneficial to peers depending on their positions and the community's size. An optimal size balances the two effects.

The position and size effects just described are summarized by applying the intermediate value theorem to (20): welfare is proportional to the welfare of a 'representative' individual, more precisely it is equal to $2\pi\Phi[P(\alpha)U(\hat{\theta}(\alpha), \alpha)]$ for some $\hat{\theta}(\alpha)$. At α^0, the 'representative' individual would benefit from an increase in size if his distance to the leader was unchanged but he will become at the same time more distant from the center. Depending on which effect is stronger, the optimal size may be larger or smaller than α^0.

In the binary case for example with Φ linear, $\Phi(v) = v$, welfare is equal to $2\pi P(\alpha)U(\alpha/\sqrt{2}, \alpha)$, that is the 'representative' individual is located at $\alpha/\sqrt{2}$. One can show that the position effect is stronger : α^{Φ} is smaller than α^0. This is more generally true under the following assumption.

A3 $[U_\alpha + \frac{\theta}{\alpha^0} U_\theta](\theta, \alpha^0)$ decreases with θ in $[0, \alpha^0]$

Proposition 7. *Under A3, whatever criteria Φ*

$$\alpha^{ext} < \alpha^{\Phi} < \alpha^0$$

In particular, whatever criteria Φ the optimal size of communities is smaller than the leader's optimum.

An implication of the proposition is that communities can be too large at a stable configuration. Recall that the preferred size of any peer is larger than the leader's one (since $\alpha^0 \leq \alpha^\theta$ from Proposition 2). Thus, if the communities' size is between α^Φ and α^0, whereas everybody expects to benefit from an increase in his community, such an expansion has a negative impact on welfare. This is due to negative external effects due to the changes in position that are not taken into account by peers when they assess the size of their own community : some communities have to disappear, leading to an increase in the distance to leaders.

Concluding remarks. This paper considers a community as a cluster of individuals with similar preferences. The possible improvement in the value of information determines the scope of a community. The analysis is conducted under strong assumptions. A peer belongs to a single community at most and uses the first signal to which he has access. Also, signals are not aggregated within a community. A natural development is to investigate different behaviors for individuals or communities. Many questions are raised. Let us mention a few:

(i) How does the value of information generated by a group depend on the way the individuals' opinions are transmitted, or aggregated ?
(ii) Can we predict whether some aggregation criteria are more apt to ensure some sort of efficiency, of stability ?
(iii) How is the analysis modified if individuals benefit from joining several communities ?

Another direction of investigation is to introduce firms with profit criteria and to study their coexistence with communities.

References

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

Dwork C., R. Kumar , M. Naor, and D.Sivakumar (2001) 'Rank Aggregation Methods for the Web' *Proceedings of the 10th international conference on World Wide Web* 613 - 622.

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).

D. Gibson, J. Kleinberg, P. Raghavan. (1998) 'Inferring Web communities from link topology'. Proc. 9th ACM Conference on Hypertext and Hypermedia.

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*.

Salop S.C. (1979) ‘Monopolistic Competition with Outside Goods’, The Bell Journal of Economics, Vol. 10, No. 1, pp. 141-156.

5 Proofs

Consider a θ-individual in $[-\alpha, \alpha]$ who follows the recommendations from the community. To simplify notation let us consider $Y(\theta, \alpha)$ the expected utility conditional on receiving a *yes*. From (3),

$$Y(\theta, \alpha) = \int_{-\pi}^{\pi} f^{\alpha}(t) u(d(\theta, t)) dt \tag{10}$$

where $f^{\alpha}(t)$ is the density of an object conditional on the receipt of a *yes* from $[-\alpha, \alpha]$. We have $U(\theta, \alpha]) = pY(\theta, \alpha)$ where $P(yes \in [-\alpha, \alpha]) = p = d^*/\pi$. Hence all properties can be shown on Y. Proof of Proposition 1 It is first useful to compute $f^{\alpha}(t)$. By viability, we restrict to $\alpha \leq d^*$. Bayes formula gives

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

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

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

In community $[-\alpha, \alpha]$, each member sends a signal *yes* for $t \in [\alpha - d^*, -\alpha + d^*]$, which we call the *acquiescence zone*, and each one sends *no* on $[\alpha + d^*, -\alpha - d^*]$, which we call the *refusal zone*. Outside these two zones there is *disagreement*. Since $\alpha \leq d^* \leq \pi/2$, all zones are non empty. By symmetry $f^{\alpha}(t) = f^{\alpha}(-t)$.

Restricting to $t \geq 0$ one has :

$$f^{\alpha}(t) = \begin{cases} \dfrac{1}{2d^*} & \text{for } t \in [0, -\alpha + d^*] \text{ acquiescence zone} \\ \dfrac{\alpha + d^* - t}{4d^*\alpha} & \text{for } t \in [-\alpha + d^*, \alpha + d^*] \text{ disagreement zone} \\ 0 & \text{for } t \in [\alpha + d^*, \pi] \text{ refusal zone} \end{cases}$$

Hence the density conditional on a *yes* is twice the unconditional one on the acquiescence zone, is null on the refusal zone, and linear in between.

Observe from expression (10) that $Y(\theta, \alpha)$ is the expectation of $u(d)$ under a distribution of the distance to θ that depends of α. Since u is decreasing, it suffices to show that these distributions are ordered by first order stochastic dominance as θ or α varies. Given δ, denote

$$F(\theta, \alpha; \delta) = Proba(d(\theta, t) \leq \delta | yes \in [-\alpha, \alpha]).$$

$F(\theta, \alpha; .)$ is the cumulative distribution of the distance to θ of the objects that are recommended by community $[-\alpha, \alpha]$ We have

$$F(\theta, \alpha; \delta) = \int_{-\pi}^{\pi} f^{\alpha}(t) 1_{[\theta-\delta, \theta+\delta]}(t) dt. \tag{11}$$

and

$$Y(\theta, \alpha) = \int_{-\pi}^{\pi} u(\delta) dF(\theta, \alpha; \delta)$$

(i) Y decreases with positive θ if for any $\delta, 0 \leq \delta \leq \pi$, any θ', θ with $\theta' \leq \theta$

$$F(\theta, \alpha; \delta) \leq F(\theta', \alpha; \delta)$$

Expression (11) shows that F is derivable with respect to θ with a derivative equal to $f^{\alpha}(\theta + \delta) - f^{\alpha}(\theta - \delta)$. This term is always non positive[10] for θ positive because the point $\theta - \delta$ is closer to the center than $\theta + \delta$.

(ii) Similarly, $Y(\theta, \alpha)$ decreases with respect to α for θ in $[-d^*, d^*]$ if for any $\delta, 0 \leq \delta \leq \pi$, any α', α, $0 \leq \alpha' \leq \alpha$

$$F(\theta, \alpha; \delta) \leq F(\theta, \alpha'; \delta)$$

that is $F(\theta, \alpha)$ decreases with respect to α. F is derivable since function $f^{\alpha}(t)$ is derivable almost everywhere with respect to α. Note that f^{α} has a null derivative

[10] The proof shows the more general property claimed in section 4 : if the density of the recommended objects by a community has a single peak, the utility for a signal stemming from that community decreases with the distance to the peak.

except on the disagreement zones. Hence taking the derivative under the integral, one has $\partial F/\partial\alpha = I + J$ where

$$I = \int_{-\alpha+d^*}^{\alpha+d^*} f_\alpha^\alpha(t) 1_{[\theta-x,\theta+x]} dt \text{ where } f_\alpha^\alpha(t) = -\frac{d^* - t}{8d^*\alpha^2}.$$

and J is defined similarly over the negative disagreement zone. We show $I \leq 0$. Note that f_α^α is negative on $[-\alpha + d^*, d^*]$ and antisymmetric around d^*. The middle of $[\theta - \delta, \theta + \delta]$, θ, is by assumption less than d^* the middle of $[-\alpha + d^*, \alpha + d^*]$. Hence I is surely non positive. The same argument applies to show $J \leq 0$, hence $\partial F/\partial\alpha \leq 0$, as desired. ∎

Proof of proposition 2 The size α^θ preferred by a θ-individual maximizes $P(\alpha)U(\theta, \alpha)$ with respect to α. Let α^{max} be the size for which $V(\alpha^{max}) = 0$. We may restrict to size smaller than α^{max}, and by symmetry, to positive θ, that is $0 \leq \theta \leq \alpha \leq \alpha^{max}$. Since U is positive, we consider instead the maximization of $logPU$. Under A0, the function is concave with respect to α, with a derivative given by

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

To prove that $\alpha^0 \leq \alpha^\theta$ it suffices to show that this derivative is nonnegative at α^0. At α^0, the derivative is null for $\theta = 0$: $\frac{P_\alpha}{P}(\alpha^0) + \frac{U_\alpha}{U}(0, \alpha^0) = 0$. Hence, assumption A1 ensures that $\frac{P_\alpha}{P}(\alpha^0) + \frac{U_\alpha}{U}(\theta, \alpha^0) \geq 0$, the desired result.

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

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

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

$$P_\alpha U(0, \alpha) + PU_\alpha(0, \alpha) + \mu[P_\alpha V + PV'](\alpha) = 0 \tag{12}$$

If μ is null, the optimal choice is α^0 as expected. If μ is positive, the constraint binds : α^0 is not viable, i.e. outside the interval $[\underline{\alpha}, \overline{\alpha}]$. Furthermore the solution solves $P(\alpha)V(\alpha) = c$, hence is either $\underline{\alpha}$ or $\overline{\alpha}$. Note that PV increases at $\underline{\alpha}$ and decreases at $\overline{\alpha}$. From (12), the derivatives of PU and PV are of opposite sign. If the latter derivative is positive, PV increases : the solution is $\underline{\alpha}$. Since under A1, $P_\alpha V + PV' > 0$ implies $P_\alpha U(0, \alpha) + PU_\alpha(0, \alpha) > 0$, it must be that the derivative of PV is non positive: the solution is $\overline{\alpha}$.

Proof of proposition 3

Let community $[-\alpha, \alpha]$ be changed into $[-\alpha - d\alpha, \alpha]$. The utility of a θ-individual in $[-\alpha - d\alpha, \alpha]$. 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

$$P(\alpha + d\alpha/2)Y(\theta + d\alpha/2, \alpha + d\alpha/2) > P(\alpha)Y(\theta, \alpha).$$

By logconcavity of Y, this inequality is equivalent to $P_\alpha U + PU_\alpha + PU_\theta(\theta, \alpha) \geq 0$. We show that it is not satisfied at any θ.

Let $f(\theta)$ be this function for $\alpha = \alpha^0$. For $\theta = 0$, by definition $P_\alpha U + PU_\alpha(0, \alpha^0) = 0$ and by symmetry of U $U_\theta = 0$, hence $f(0) = 0$. For $\theta = \alpha^0$, $f(\alpha^0)$ is equal to $V'(\alpha^0)$, which is negative by A2. Hence f is negative for each $\theta > 0$ if it decreases with θ. Since U decreases with θ, this is ensured by the assumption that $U_\alpha + U_\theta$ decreases with θ. ∎

Proof of proposition 4 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

$$P_\alpha U(0, \alpha) + PU_\alpha(0, \alpha) + \mu[P_\alpha V + PV'](\alpha) = 0 \tag{13}$$

$$P_\lambda U(0, \alpha) - c'(\lambda) + \mu[P_\lambda V(\alpha) - c'(\lambda)] = 0 \tag{14}$$

When μ is null, the viability constraint does not bind, and the leader can choose its optimal value. The same analysis as in proposition 2 yields that for the chosen value of λ other peers would like the size to increase. As for the contribution rate, since peers' utility level $U(\theta, \alpha)$ is less than that of the center, (14) yields that aθ-peer would prefer a smaller contribution rate, that is $P_\lambda U(\theta, \alpha) - c'(\lambda) \leq 0$.

When μ is positive, we know that (13) and A1 implies that α is set at the maximal viable size associated to the chosen λ, $\overline{\alpha}(\lambda)$. From the first order condition on λ (14), $P_\lambda U(0, \alpha) - c'(\lambda)$ and $P_\lambda V(\alpha) - c'(\lambda)$ are of opposite sign. Since $U(0, \alpha) > V(\alpha)$ it must be that the former is positive : the leader would prefer to increase the contribution rate.

Proof of Proposition 5:

Consider a community with gaps. We proved in the text that a necessary condition for stability against enlargement is that the community size α be larger than the leader's choice, that is larger than $min(\alpha^0, \overline{\alpha})$. Conversely let α be between $min(\alpha^0, \overline{\alpha})$ and $\overline{\alpha}$. If α is the maximal viable size, an enlargement is not viable so no outsider wants to join. If the size α is between α^0 and $\overline{\alpha}$, then it is stable for another reason: the individuals at or close to the leader do not benefit from an increase in size and if the enlargement is not centered, half of them are further away from the leader, also diminishing their utility (from Proposition 2).

Consider now a full community with size α larger than α^V and a proposal to an enlargement. At the standing configuration, the minimum utility level is achieved by an individual located at an extreme point of a community, hence $\bar{u}(\theta) \geq (PV)(\alpha)$ for each θ. Let us consider an individual located at an extreme point of the proposal of size β larger than α. If the proposal is accepted, his expected payoff is $(PV)(\beta)$. This is strictly lower than $(PV)(\alpha)$ since PV decreases for size larger than α^{ext}. Thus individuals at or close to the extreme of the new community would be strictly worse off by accepting: the proposal is rejected. ■

Proof of Proposition 6 The proof is divided in three steps.

Step 1. We first show that stability can be checked by considering proposals that are centered at the leader's position. Let $[-m-\beta, -m+\beta]$ be a reduction proposal, $\beta < \alpha$, centered at $-m$, $0 < m$. We have $m < \beta$ because the leader belongs to the proposal. We show that if $[-m-\beta, -m+\beta]$ is accepted, then $[-\beta, \beta]$ is also accepted. Assuming

$$P(\beta)U(\theta, [-m-\beta, -m+\beta]) > \bar{u}(\theta), \theta \in [-m-\beta, -m+\beta]. \qquad (15)$$

we have to show (using notation $U(\theta, [-\beta, \beta]) = U(\theta, \beta)$)

$$P(\beta)U(\theta, \beta) > \bar{u}(\theta), \theta \in [-\beta, \beta] \qquad (16)$$

Up to a rotation $U(\theta, [-m-\beta, -m+\beta]) = U(\theta+m, [-\beta, \beta]) = U(\theta+m, \beta)$ therefore (15) can be written as

$$P(\beta)U(\theta+m, \beta) > \bar{u}(\theta), \theta \in [-\beta, \beta]. \qquad (17)$$

Applying (17) to $\theta = 0$ gives $P(\beta)U(m, \beta) > \bar{u}(0)$. Hence, from the properties of U (Proposition 1): $P(\beta)U(\theta, \beta) > \bar{u}(0)$ for $0 \leq \theta \leq m$.

Now take θ', with $\beta > \theta' > m$. Let $\theta = \theta' - m$. Note that both θ and θ' are in community $[0, \alpha]$ in the standing configuration and θ is closer to the center than θ'. Therefore $\bar{u}(\theta) \geq \bar{u}(\theta')$. Thus (17) applied to θ' gives

$P(\beta)U(\theta', \beta) > \bar{u}(\theta) > \bar{u}(\theta')$, for $\theta' \in [m, \beta]$.

Thus inequality (16) is met for any positive θ smaller than β, which gives the result for any $\theta \in [-\beta, \beta]$ by symmetry of U.

Step 2. From Step 1, stability against reduction is satisfied if no centered proposal is accepted, that is inequality (16) is not satisfied at any β, $\beta < \alpha$. Since $[-\beta, \beta]$ is included in $[\alpha, \alpha]$, we have $\bar{u}(\theta) = P(\alpha)U(\theta, \alpha)$ for each $\theta \in [-\beta, \beta]$.

So stability against reduction is satisfied if for no β, $\beta < \alpha$

$$P(\beta)U(\theta, \beta) > P(\alpha)U(\theta, \alpha), \theta \in [-\beta, \beta] \qquad (18)$$

We show that under A1, this is equivalent to :

$$P(\alpha)U(\beta,\alpha) > P(\beta)V(\beta) \text{ for each } 0 < \beta < \alpha \tag{19}$$

Inequalities (19) are sufficient : for no β (18) is satisfied at θ close to β. Conversely, let inequality $P(\alpha)U(\alpha^*,\alpha) \le P(\alpha^*)V(\alpha^*)$ be met for some α^* smaller than α: Proposal $[-\alpha^*,\alpha^*]$ makes α^*- or $-\alpha^*$-individuals at least as well off. The elasticity conditions A1 then imply that all members in $]-\alpha^*,\alpha^*[$ are strictly better off. To see this, observe that the ratio[11] $\frac{U(\theta,\alpha)}{U(\theta',\alpha)}$ decreases with α for $\theta < \theta'$. For $\theta < \alpha^* < \alpha$, taking $\theta' = \alpha^*$ gives

$$U(\theta,\alpha) \le U(\theta,\alpha^*)\frac{U(\alpha^*,\alpha)}{U(\alpha^*,\alpha^*)}.$$

Using inequality $P(\alpha^*)U(\alpha^*,\alpha^*) > P(\alpha)U(\alpha^*,\alpha)$ gives that for each $\theta \in]-\alpha^*,\alpha^*[$

$$P(\alpha^*)U(\theta,\alpha^*) \ge P(\alpha)U(\theta,\alpha^*)\frac{U(\alpha^*,\alpha)}{U(\alpha^*,\alpha^*)} > P(\alpha)U(\theta,\alpha)$$

Thus (18) is satisfied at $\beta = \alpha^*$: proposal $[-\alpha^*,\alpha^*]$ is accepted.

Step 3. Let A be the set of sizes of configurations stable against reduction. From Step 2. A is the set of viable sizes for which inequalities (19) hold. We show that A is an interval of the form $]0,\alpha^{int}]$.

Note first that the set A contains $]0,\alpha^V]$, hence is non empty. To see this, recall that PV increases on $[0,\alpha^V]$. Hence for α smaller than α^V we have

$$P(\alpha)U(\beta,\alpha) > P(\alpha)U(\alpha,\alpha) = P(\alpha)V(\alpha) > P(\beta)V(\beta) \text{ for each } \beta, 0 < \beta < \alpha$$

(the first inequality holds because $U(\beta,\alpha)$ decreases with α). The set A is bounded by $\overline{\alpha}$. If $A =]0,\overline{\alpha}]$, take $\alpha^{int} = \overline{\alpha}$. Otherwise, consider the smallest value, α^{int} for which (19) does not hold. We have

$P(\alpha)U(\beta,\alpha) > P(\beta)V(\beta)$ for each $\alpha < \alpha^{int}$ and $\beta < \alpha$. Furthermore, since (19) does not hold at α^{int}, there is some $\alpha^* < \alpha^{int}$ with $P(\alpha^{int})U(\alpha^*,\alpha^{int}) \le P(\alpha^*)V(\alpha^*)$. Therefore, by continuity, $P(\alpha^{int})U(\alpha^*,\alpha^{int}) = P(\alpha^*)V(\alpha^*)$ and the function $\alpha \to P(\alpha)U(\alpha^*,\alpha)$ decreases at α^{int}. Since it is a concave function, it decreases for larger values than α^{int}, hence

$$P(\alpha)U(\alpha^*,\alpha) \le P(\alpha^{int})U(\alpha^*,\alpha^{int}) = P(\alpha^*)V(\alpha^*).$$

Inequality (19) does not hold for a size larger than α^{int}: the set A is the interval $]0,\alpha^{int}[$. ∎

[11] The derivative with respect to α of the log of the ratio is $\frac{U_\alpha(\theta,\alpha)}{U(\theta,\alpha)} - \frac{U_\alpha(\theta',\alpha)}{U(\theta',\alpha)}$ which is negative under A'1.

PROOF OF PROPOSITION 7. Total welfare W (20) writes:

$$W(\alpha) = \frac{2\pi}{\alpha} \int_0^\alpha \Phi[P(\alpha)U(\theta,\alpha)]d\theta. \tag{20}$$

Take the derivative with respect to α:

$$W'(\alpha) = 2\frac{\pi}{\alpha}\Big\{ \int_0^\alpha \Phi'[P(\alpha)U(\theta,\alpha)][PU_\alpha + P_\alpha U](\theta,\alpha)d\theta + \Phi[P(\alpha)U(\alpha,\alpha)] - \frac{1}{\alpha}\int_0^\alpha \Phi[P(\alpha)U(\theta,\alpha)]d\theta\Big\} \tag{21}$$

Integration by parts of $\theta \to \theta\Phi[P(\alpha)U(\theta,\alpha)]$ gives

$$\alpha\Phi[P(\alpha)U(\alpha,\alpha)] - \int_0^\alpha \Phi[P(\alpha)U(\theta,\alpha)]d\theta = \int_0^\alpha \theta\Phi'[P(\alpha)U(\theta,\alpha)]P(\alpha)U_\theta(\theta,\alpha)d\theta$$

thus

$$W'(\alpha) = 2\frac{\pi}{\alpha}\int_0^\alpha \Phi'[P(\alpha)U(\theta,\alpha)][PU_\alpha + P_\alpha U + \frac{\theta}{\alpha}PU_\theta](\theta,\alpha)]d\theta.$$

The term $[PU_\alpha + P_\alpha U]$ represents the size effect. For a θ-peer, it depends on the community size being lower or larger than the preferred size α^θ. However, if the community size is less than α^0, the size effect due to a small increase is beneficial to all peers. The term $[\frac{\theta}{\alpha}PU_\theta]$, which is always negative, reflects the position effect. Since Φ' is positive, let us focus on $[PU_\alpha + P_\alpha U + \frac{\theta}{\alpha}PU_\theta]$

It is non positive at the two extreme points of integration : at $\theta = 0$ because both the size and position effects are null; at $\theta = \alpha^0$ because the integrand is exactly equal $(PV)'(\alpha^0)$, the variation of utility for an individual located at the extreme of the community, which is negative from proposition 2. Under assumption A3, $[PU_\alpha + P_\alpha U + \frac{\theta}{\alpha}PU_\theta] + PU$ is decreasing with θ, hence stays nonnegative on $[0, \alpha^0]$ which gives the result.

∎

APPENDIX : BINARY FUNCTION. Given a binary function, set $k = \frac{(1+b/g)}{4d^*}$. We compute below that for $\theta \in [-d^*, d^*]$ and $\alpha \le d^*$ U is given by

$$U(\theta,\alpha) = pg[1 - k\alpha(1 + \frac{\theta^2}{\alpha^2})] \text{ for } \theta \in [-\alpha,\alpha] \tag{22}$$

Let us check that assumptions A0, A1, A2 and A3 are satisfied.

Equation (22) gives $V(\alpha) = pg(1 - 2k\alpha)$ and the maximum viable size (if c is null) is $\alpha^{max} = 1/2k$. and

$$-\frac{V'}{V}(\alpha) = \frac{2k}{1-2k\alpha} \text{ and } -\frac{U_\alpha}{U}(\theta,\alpha) = k(1 - \frac{\theta^2}{\alpha^2})/[1 - k(\alpha + \frac{\theta^2}{\alpha})].$$

A0 is met since U and V are concave in α.

A1. Since $2V(\alpha) = g(1 - 2k\alpha)$, we have

$$-\frac{V'}{V}(\alpha) = \frac{2k}{1-2k\alpha} > \frac{k}{1-k\alpha} = -\frac{U_\alpha}{U}(0,\alpha)$$

Also, the function $-\frac{U_\alpha}{U}(\theta,\alpha)$ is decreasing 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 for $\alpha \leq \alpha^{max} = 1/2k$.

A2 and A3. One has:

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

This gives

$[U_\alpha + U_\theta] = U_\alpha(0,\alpha) - pgk\frac{\theta}{\alpha}[2 - \frac{\theta}{\alpha}]$ and $[U_\alpha + \frac{\theta}{\alpha^0}U_\theta] = U_\alpha(0,\alpha) - pgk\frac{\theta^2}{\alpha^2}$.

Both are decreasing with θ in $[0,\alpha]$.

Computation of (22). We shall use repeatedly that for t_1, t_2 in the (positive) disagreement zone one has

$$\int_{t_1}^{t_2} f^\alpha(t)dt = \frac{-1}{8d^*\alpha}(\alpha + d^* - t)^2|_{t_1}^{t_2} = \frac{1}{8d^*\alpha}(t_2 - t_1)(2\alpha + 2d^* - (t_2 + t_1)) \quad (23)$$

Let us first compute Y, the utility derived from buying conditional on a *yes*, for a θ-individual in the community. For objects $t \geq 0$, he achieves :

in the acquiescence zone $[0, -\alpha + d^*]$: g,

in the disagreement zone $[-\alpha+d^*, \alpha+d^*]$: g for any object in $[-\alpha+d^*, \theta+d^*]$, and $-b$ on $[\theta + d^*, \alpha + d^*]$.

The refusal zone has not to be considered since a *yes* never comes from it. This gives using (23)

$$g\frac{d^* - \alpha}{2d^*} + \frac{1}{8d^*\alpha}[g(\alpha+\theta)(3\alpha-\theta) - b(\alpha-\theta)^2].$$

By symmetry, the utility for a θ-individual on negative t is equal to that of a (-θ)-individual on positive t. Thus the utility for t-objects with $t \leq 0$ is equal to:

$$g\frac{d^* - \alpha}{2d^*} + \frac{1}{8d^*\alpha}[g(\alpha-\theta)(3\alpha+\theta) - b(\alpha+\theta)^2]$$

Collecting terms gives Y, the overall utility of a θ-peer receiving a *yes* :

$$Y(\theta,\alpha) = g(\frac{d^* - \alpha}{d^*}) + \frac{1}{4d^*\alpha}[g(3\alpha^2 - \theta^2) - b(\alpha^2 + \theta^2)]$$

Rearranging and multiplying by p gives (22).