Comparing online community management policies: an agent-based approach

Alberto Cottica

March 17, 2017

1 Introduction

Online communities are used to aggregate and process information dispersed across many individuals. Pioneered in the 1980s, they have become more widespread with mass adoption of the Internet, and are now used across many different contexts in business [?, ?], politics and public decision making [?, ?, ?], expertise sharing [?, ?, ?], and education [?]. At the same time as they spread across domains, they did so geographically: for example, they have attracted large numbers of users and large venture capital investments in China [?]. Most online communities lack a central command structure; despite this, many display remarkably coherent behaviour, and have proven effective at large tasks like writing the largest encyclopedia in human history (Wikipedia), providing an always-on free helpline for software engineering problems (StackOverflow), or building, and continuously updating, a detailed map of planet Earth (OpenStreetMap) [?].

Organizations running online communities typically employ community managers, tasked with encouraging participation and resolving conflict: this practice is almost as old as online communities themselves and predates the Internet [?], although it has become much more widespread as Internet access became a mass phenomenon. Though most participants to online communities are unpaid and answer to no one, a small number of them (only one or two in the smaller communities, many more in the larger ones) report to a central command, and carry out its directives. Following the convention of practitioners themselves, we shall henceforth call such directives *policies*.

Putting in place policies for online communities is costly. Professional community managers need to be recruited, trained and paid; software tools to monitor communities and make their work possible need to be developed and maintained. This raises the question of what benefits organisations running online

communities expect from policies; and why they choose certain policies, and not others.

A full investigation of this matter is outside the scope of this paper; however, in what follows we outline and briefly discuss the set of assumptions that underpin our investigation.

1. In line with the network science approach to online communities, we model online communities as social networks of interactions across participants.

2. We assume that organisations can be modelled as economic agents maximising some objective function. The target variable being maximised can be profit (for online communities run by commercial companies); or welfare (for online communities run by governments or other nonprofit entities); or some combination of the two.

3. We assume that the topology of the interaction network characteristic of online communities affects their ability to contribute to the maximisation of the target variable.

4. We assume that such organisations choose their policies as follows:

   - Solve their maximisation problem over network topology. This yields a vector of desired network characteristics, where "desired" means that those characteristics define a maximum of the objective function. These solutions will be statements with the form "In order to best meet our ultimate [profit or welfare] goals, the interaction network in our online community should be in state Θ_D, where Θ is a vector of topology-related parameters".

   - Derive a course of action that community managers could take to change the network away from its present state Θ_0 to the desired state Θ_D.

   - Encode such course of action in a set of simple instructions for community managers to execute. They call them policies; computer scientists might think of such instructions as algorithms; economists call them mechanisms.

This paper is a first step at modelling the choice of an online community management policy based on the goals of the principal organisation. To simplify the problem, we forego directly modelling the organisation as it maximises its objective function. Instead, we think of policies as heuristics. The principal organisation considers a set of indicators that (a) can be argued to correlate to either profit or welfare and (b) are simple to compute. It chooses from a set of policies, which have beneficial effects on those indicators. Different policies influence indicators in different ways: for example, one can be more effective to

achieve a high level of activity, whereas another can be better at promoting inclusivity. We assume capacity constraints, so this has the potential to engender trade-offs: one indicator can be improved only at the expense of another. Given constraints, the organisation's actual choice of policies will of course depend on its relative preferences for the different indicators.

To compare policies, we first build an agent-based model of online communities, in such a way that some meaningful indicators are simple to extract from it. We then simulate the impact of different policies on those indicators.

Section **??** reviews the literature on this topic. Section **??**describes the model, offers a rationale for the policies therein, and specifies the simulation protocol. Section **??** presents the results. Section **??** discusses them.

2 Literature review

3 Materials and methods

3.1 Model

The overall phenomenon I want to model is online community management practices. Online communities predate the Internet, and are widespread in business, academia, civil society activism. Many organisations running online communities employ professionals ("community managers") to enact certain practices on the communities themselves, like content moderation, conflict management or orientation of newcomers.It follows that such professionals somehow earn their keep by furthering the goals of their employers.

This subject fascinates me because it is a very simple setting that captures the idea of influencing emergent social dynamics. The organisations running online communities do so in pursuit of their own goals, but they cannot control what online community members do. They can, and do, attempt to influence them ? sometimes with success. I believe a deep understanding of the contact surface between design and emergence might be relevant in many areas of business and public policy.

I model a subset of this phenomenon which lends itself to being formalized in simple algorithms, which I call "policies" after online community management lingo. Specifically, I would like to model two policies, both of them in use in many real-life online communities:

1. **Onboarding**: "when a user posts something for the first time, leave her

a comment." This practice is used to welcome newcomers and reinforce their first experience of participating in the online community.

2. **Engagement**: "if someone posts something, leave her her a comment even if this is not their first time posting. If she posts more than once during the same period, only leave her one comment."

3.2   Agent types

The model represents a stylised online community. There are two types of agents:

1. **Ordinary members**. These represent people who spontaneously joined the online community, in pursuit of their own interests.

2. **Community managers**. These represent people who are in the online community to do a job. The job is that of enacting the policies that the community's parent organization want done. We will assume there is only one such agent, without loss of generality.

3.3   Agent properties

The properties of ordinary community members:

1. **Chattiness**, the propensity to participate in the online conversation. It can be modelled as a baseline probability to post in any given period, independent of anything else.

2. **Responsiveness**, the propensity to engage in online conversation when the member has received one or more comments in the period. It can be modelled as an additional probability, which is a function of the number of comments received.

3. **Join date**, a property that encodes the period in which the agent became a member of the online community.

4. **Number of comments written**. A property that keeps track of the number of comments written by the member.

5. **Number of comments received**. A property that keeps track of the number of comments received by the member.

6. **Membership strength**. A property that encodes the history of comments received. The idea is to capture the intensity with which a member feels part of the community: if this drops too low, the member might decide to leave the community altogether, like in [**?**].

The behaviour of community managers is driven by the policy they are enacting rather than by parameters like chattiness and responsiveness. Membership strength is meaningless ? the community manager does not need to feel involved, she is being paid for what she does. The model will keep track of the number of comments they write and receive for tidiness.

3.4   Agent behaviour

Ordinary community members can:

- Join the community. We assume this is exogenous: at each time step a certain number of agents are created.

- Engage in communication (by posting a comment). If they do so, they need to choose another agent as the target of that communication.

- Leave the community.

Community managers can only engage in communication. They do not join, nor leave.

4   Environment

This is a network model. The environment is assumed not to influence it. In NetLogo terms, patches do not play any role, and the model is limited to turtles and links.

However, there are two types of links:

- Links that represent communication events (user A comments user B at time t). These links are directed and unweighted.

- Links the represent pairwise intimacy. These links are directed and weighted. The weight is, at each point in time, the sum of the time-discounted values of previous communication events. The idea is that a recent communication from user A to user B has a greater influence on how close B feels to

A than a less recent one. This idea was introduced in [**?**]; it is useful because it allows us to model "exclusionary" communities, where newcomers cannot get a foothold because members prefer to talk to the friends they already have.

4.1 Model mechanics

Upon initialization, we create one community manager, and a small number of community members (at least 2). We assume the latter to all be connected to each other by intimacy links of weight greater than zero.

At each time step, the following happens.

1. Zero or more members join the community

2. Members decide whether to engage in communication. We assume this choice depends on chattiness, responsiveness and the number of comments received in the previous period. I have empirical results that can both confirm this is a plausible assumption and provide an estimate of the marginal effect of receiving comments on the probability to engage ([**?**]).

3. Members who have decided to engage in communication now must choose one other member to communicate with. We assume that their choice depends on pairwise intimacy; the closer A feels towards B, the more likely she will choose B as the target of their communication. This can be modelled by a function $\phi_{i,j}(t_n)$ that takes value 1 if member i leaves a comment to member j at time t_n, and zero if she does not.

$$P[\phi_{ij}(t_n) = 1] = \frac{1 + \beta W_{ij}(t_{n-1})W_{ji}(t_{n-1})}{\sum_j [1 + \beta W_{ij}(t_{n-1})W_{ji}(t_{n-1})]} \qquad (1)$$

This assumes that the average number of comments left by each member in each period in which they decide to engage in communication is 1. β is a "cultural" parameter that measures how important previous intimacy is in deciding who to interact with. When it is zero, members leave a comment to a randomly chosen member; when it is infinity, they only ever interact with members that they have interacted before.

4. The community manager engages in her own communication events, depending on whether she is enacting onboarding, engagement or both.

5. Communication events are encoded in communication links.

6. The value of intimacy links are updated as follows. Let $W(t_n)_{i,j}$ be the intimacy felt by person i for person j at time t_n. Let τ be the time discount rate. Then:

$$W_{ij}(t_n) = W_{ij}(t_{n-1})e^{-\frac{\Delta t}{\tau}} + \Phi_{ij}(t_n) \qquad (2)$$

6

7. Members decide whether to leave the community. This is done by computing the membership strength S, which is simply the sum of the weights of the incoming intimacy links (in other words, the weighted in-degree of the member in the intimacy network), and comparing it with a threshold $\bar{S}$.

4.2 Outputs and purpose of the model

I aim to explore how different initial conditions, and different policies and combinations thereof, affect certain features of the online community that might be seen as desirable by the organizations running it. This model lends itself to computing four of them.

1. **Activity**. Active, lively online communities are likely to be worth more than static, slow-going ones. Activity can be measured by the number of communication events that take place over a period of time.

2. **Diversity of contribution**. For communities aimed at knowledge work, like Wikipedia, a higher interactivity between members can lead to a more reliable output, as people watch out for each other's errors. All other things being equal, a community where contributions are more evenly distributed is therefore more desirable than one where only a few individuals contribute a lion's share of the content. We can measure diversity of contributions as the Gini coefficient of the distribution in the number of comments by agent.

3. **Social capital**. An online community is better at motivating its members to participate if they feel strongly that they are part of it. We can measure social capital in this sense by a sum of the membership strength S_i across all is.

4. **Inclusivity**. An online community is likely to be more valuable if most newcomers feel welcome and included, and find friends quickly. Inclusivity can be measured by the Gini coefficient of the distribution of S_i. As it approaches 1, fewer members enjoy strong membership, whereas many are mostly ignored (recall that S_i encodes the time-discounted comments that were left to user i). This part of the analysis is a reconsideration of [**?**].

5 Results

6 Discussion