

The impact of founders and early participants on the attainment of critical mass in online political communities

Wybo Wiersma

Oxford Internet Institute

mail@wybowiersma.net

1 Introduction

Online communities can greatly impact society, as the recent events in the Middle East, and the rise of global online social movements such as Occupy Wallstreet, suggest ([aouragh_egyptian_2011](#) [ellis_riots_2011](#)). But in order to thrive, and become a community at all, online communities have to attain a critical mass of initial users. The central problem of which is that, until a certain number of participants are present, joining the community is not going to affect outcomes, or be socially rewarding to newcomers. The exact factors that determine growth in the early stages of online communities and movements, are still ill understood. Even internet giant Google has failed to attain critical mass for some of its platforms, such as *Google Wave*.

In the research proposed here, budding online political communities will be studied, and the ones that turn out to become successful (grow or continue to exist) will be contrasted with those that fail. More specifically, it will be examined whether the *starters* (founders and early participants, [schelling_micromotives_2006](#)) of successful online political communities, differ, either individually or in their position in social networks, or whether it are the budding communities as a whole that differ, for example in community-level dynamics. Contrary to the traditional belief in the impact of leaders, the answer to this question is not a given, as studies of ‘leaderless’ online movements, and smart-mobs, suggest ([howard_smart_2002](#) [colomer_science_2011](#)

schelling_micromotives_2006 oneill_cyber_2009 colomer_science_2011 ganz_leading_2008

). Also, and more importantly, several theories suggest the possibility that *starters* of successful communities could be socially indistinguishable from those of communities that fail.

In Granovetter's threshold theory, for example, people vary in their threshold k , requiring k others to become mobilized. A low k makes *starters* different from late adopters. But all *starters* will have a low k regardless of the early community they are in (and thus are similar in this respect). In addition, his theory proposes that the success of movements is determined by whether people with low enough k s can be reached: leading to a riot if k s 0 to 100 are present, and only a single protester ($k = 0$) amidst solid citizens, if instead of $k = 1$ an additional $k = 2$ is present. It is therefore possible that *starters* at successful communities do not, or only barely, differ from those at failing ones (**granovetter_threshold_1978 granovetter_threshold_1988**).

In order to gain a greater understanding of the relative impact of *starters*, their social positioning, and community-level dynamics, the comparison between successful and failing communities will be made for relatively wide selection of factors. The factors that will be examined were suggested by theory or found relevant in earlier studies. For individual *starters*, they include cultural- and social capital (language-use, and Ego-network analysis). For communities the contrast will be made by analysing the frequency of interactions, community atmosphere (sentiment analysis), and network-properties (network analysis).

The remainder of this proposal is structured as follows. The theoretical background will be sketched first. Then the hypotheses and the methods that will be used to research them, are introduced. Following this, the data collection strategy, and precise definition of success that will be used, are set out. Finally, in the conclusion, a brief analysis will be given of the limits of the proposed study, and potential impact of the expected findings, among which enabling the early identification of successful communities.

2 Theory

As noted, in Granovetter's threshold theory, social connectedness and serendipity in reaching those with low k s, are important for success. Others, such as Watts have built

upon Granovetters work by overlaying a social network structure on the population, and limiting those that count towards the k of potential joiners, to their neighbours in the network (**watts_simple_2002 watts_threshold_2009**). This addition not only makes Granovetters model more realistic, but also introduces a plausible mechanism (social capital: being well-connected) that determines whether *starters* can reach susceptible potential joiners.

In addition to threshold theories, there are two more branches of theory that question whether success depends on the properties of *starters*. The first is grounded on social learning- and information cascades theory, and introduces the requirement of reaching people in the right order. In this theoretical model, people make adoption decisions about (social) behaviour based on two types of information: a private signal about whether adoption will likely provide a benefit (say 1 or -1), and the actions (not private signals) of others, faced with same the decision before, with each data-point weighted equally.

The first person decides based on her private signal. If she adopts, and the second person receives a positive signal, he will surely adopt. If he receives a negative private signal ([1, -1]), he will decide by flipping a coin. The third person, observing 2 adoptions ([1, 1]), will surely adopt as well, even if she receives a negative signal herself. If she sees one adoption and one rejection, she will decide based on her own private signal (similar to the first person). Following this logic, a cascade will appear very soon, in 99.6% of cases after 8 turns. Two positive private signals in a row are enough to start a cascade of adoption, even if all following signals are negative. This makes the order in which potential joiners are approached (often determined by chance), very important (private signals [1, 1, -1, -1] make all four adopt, while [-1, -1, 1, 1] leads to the opposite outcome) (**bikhchandani_theory_1992 bikhchandani_learning_1998**).

The third branch of theories is concerned with the diffusion of innovations (not social behaviour per se). It divides the population of potential adopters in 5 groups: innovators, early adopters, early majority, late majority and laggards, which each have their own preferences and psychometric profiles (**rogers_diffusion_1995**). Geoffrey Moore, building on Rogers work, focuses on why innovations fail to spread from early adopters to the early majority. He ascribes this to the early majority generally being risk-averse pragmatists, who want proven technology, and prefer to do business with market-

leaders only, as opposed to innovators who are more comfortable with risk. In his theory, success is a matter of adapting the message to match the needs and language of successive adopter-groups (**moore_crossing_2002**): thus addressing potential joiners with the right message. Rogers also provides a listing of 26 dimensions along which adopters in each category differ. Some of which are likely to differ between the *starters* of successful and failing communities as well.

3 Hypotheses and Methods

3.1 Properties of starters

The main hypothesis, is that (*H0*) *starters* either do not differ significantly between early successful and failing communities, or (*H1*) *starters* behind successful communities have more social and cultural capital, but are similar in other *k*-factor affecting respects, such as high risk-tolerance.

Hypothesis 1 is based on earlier studies, in which two types of early adopters seem to have been found: Initial studies of early adopters (of new farm technology, in the 1930's) found them to be less connected, and less well-off than later adopters, while in later studies, early adopters were generally found to be of higher socio-economic status (**daberkow_socioeconomic_1998 diederer_innovation_2003 donnelly_social_1970**). A similar division seems plausible for online community founders: Those with little social capital might be more likely to be the first to seek out equal minded people on-line, while the well-connected would be in a better position to be aware of new ideas, have leadership skills, and be successful at drawing in followers. In a recent study of protest-messages on Twitter, some evidence was indeed found that *starters* are not always well-connected, but span the whole range of social connectedness (**gonzalez-bailon_dynamics_2011**). Thus what is hypothesised here, is that though both types of *starter* exist, better connected *starters* are likely to be at the root of more successful online political communities.

Social connectedness can be assessed through surveys, but more effectively through automated means. Doing automated Google searches, and scraping the number of *starters*' LinkedIn, Twitter- and Facebook-followers (or even whole ego networks, with their permission), could give a rough indication of the online presence of *starters*. Be-

cause of issues with the uniqueness of names, and other inaccuracies, this method cannot be used as the sole approach, but when checked against survey-data about a subset of *starters*, it can increase the scale of the study. Also, when ego networks are obtained, it may be possible to find out whether existing friend-networks formed the basis of communities.

Whether *starters* of successful communities have more cultural capital, and match the early adopter type to a larger extent, is the second likely difference that will be examined. This can be linked to certain demographics, such as education and income, and several other of Rogers 26 factors. Information on about five such factors (to be determined by a pilot study) will be elicited using surveys. A complementary way to gauge cultural capital, is through computational linguistic analysis. Mean Length of Utterance (sentence-length), for example, can be used as a proxy for sophistication of language use (**bernstein_linguistic_1962**). Diversity of word-usage, is another proxy. The advantage of these computational methods is that they can be automated, and applied to community posts on a large scale, at the very least allowing for corroboration of evidence (**wiersma_automatically_2011**).

3.2 Community-level factors

If hypothesis 0 is true, and *starters* do not differ between successful and failing communities, then it is hypothesized that the (most important) differences might lie in community-level factors (*H2*). For these, the expectation is that interaction patterns, and community-atmosphere, will be important for success, in addition to various background factors affecting the community.

First of all, patterns of interaction may differ. Lampe and Kraut, for example, found that newcomers are 12% points more likely to come back if they received a reply to their first post (**joyce_predicting_2006**). In addition, a lot of interaction can make a community seem larger than it is, thus increasing its perceived *k* for new arrivals. And perceptions can make a difference, at least with regard to global information. For example it was found that people are much more likely to sign a petition if a million others have allegedly signed it, than when this number is smaller (for the same petition) (**margetts_how_2009 margetts_social_2011**). As for methods, a quantitative analysis of interactions, such as frequencies per time-period, is one way to dis-

sect this dimension. Another is network analysis, which makes it possible to examine variations in social connectedness within the communities, and to see whether various community-level roles are fulfilled, such as that of answering-person, and discussion-person (**fisher_you_2006**).

Another potentially important community-factor, is atmosphere, or the prevailing sentiment within the community. Tentative research findings suggest that changes in sentiment may forebode the fall of communities (forthcoming, González-Bailón). While a positive atmosphere can be welcoming to newcomers, and will stimulate lurkers to contribute, animosity on the other hand, might stimulate heated discussions, and thus increase interactions (**chen_online_2006**). Other important factors are the extent to which *starters* and early joiners are committed to the community (identity-based commitment), to other individual members (bond-based commitment) or to community-values (value-based commitment). Surveying *starters* can shed light on these factors, while computational methods (sentiment-analysis based on keywords and expressions) will allow for analysis at a larger scale.

The final factors for which successful communities might differ, are offline- and qualitative factors, such as institutional backing. For example, NGOs and political groups with a significant offline component were found to have more incoming web-links (**gonzalez-bailon_opening_2009 gonzalez-bailon_traps_2009**). Also, institutional backing might make critical mass attainment easier by instilling the expectation that critical mass will be attained, making a return on investment seem more likely to potential contributors. Other background-factors that might benefit communities, are their age, them being ‘first to market’, appearance in the media, and other external factors. As previous research on community-factors is more limited than that on properties of founders and early adopters of innovations, additional community factors might be considered as the research proceeds.

4 Data collection

Data will be collected for two types of online political communities: discussion-groups which do not champion for explicit causes, aims or events (such as rallies or protests), and those which do. Both are included because on-line-only groups and those with prac-

tical aims are likely to have different critical mass dynamics. In addition, communities on two different platforms will be studied: Facebook-groups, and mailing-lists. These platforms were chosen because Facebook IDs and e-mail addresses provide a good (and unique) handle on *starters* online presence, and allow the researcher to contact *starters* for surveys. Finally, both successful, and failing communities will be analysed.

Critical mass attainment (success at early growth) is hard to define. Attempts have been made to put a number on the community-level critical mass threshold (30 and up). But this will not be necessary here, because critical mass attainment will be operationalised indirectly, as relative success at early growth. Data will be collected at two points in time from early communities (less than a year old) with 20 or fewer active participants. The separation of those that failed, and succeeded will happen only after the second collection. Success will be defined relatively, as compared to the other communities that were part of the study. Growth relative to initial size will be calculated, and used to rank them. The top 1/3rd will then be considered successful. The advantage of this approach is that the study does not depend on an a-priori definition of critical mass, nor on picking sufficient communities that turn out to match such a definition (at the modest price of defining success in relative terms).

The first collection will happen in spring of the first year, and the second in spring of the second year (leaving the last year to finalize the study). At least 12 communities will be included for each of the 8 fields (2 types * 2 platforms * 2 success-states * 3/2th), for a total of 144 (see table 1). Surveys will be solicited from least three *starters* per community (432 total), to ensure that for moderate effect-sizes statistically significant results can be obtained across each division. The computational analyses will be applied to all (and possibly more communities). Examples of communities that might be included are listed in the table. Sampling of communities will happen in a semi-random fashion, with a sampling frame based on Facebooks graph API, and lists of public mailing-lists, such as Google- and Yahoo-groups. Communities will be drawn from these, and selected if they meet the size and age-criteria, and are explicitly political in nature.

Table 1: *Types of communities and platforms (success-states not included)*

Divisions	Discussion	Action
Facebook	Liberalism vs. Conservatism	UK Uncut
	Global Forum	Occupy Wallstreet
Lists	Politics Articles	Philly Against War
	Democracy Village	OccupySF WorkingGroups

5 Conclusion

There is still much to be understood about the crucial early stages of online communities in which critical mass is attained. The research proposed here is limited in that it is specific to two platforms, takes into account a limited number of factors, and cannot use an ideal sampling frame. Nevertheless, as the first systematic study of the differences between early successful, and failing communities (as opposed to a comparison between various adopter-groups, or the spread of a particular movement), it has the potential to provide groundbreaking results.

If a difference is found between *starters*, it is expected that those at successful communities will be socially better connected, and have more cultural capital. They might also be a better fit with the archetype early adopter. Such findings might allow us to extend Granovetters theoretical framework by introducing factors orthogonal to adopter thresholds k , which affect the size of *starters* mobilisation-impact on others. It can do this thanks to its comparative design, which isolates factors related to success from general low- k -factors. And although Bikhchandanis cascades theory already discusses the potential impact of thought-leaders on critical mass attainment, the proposed work will add to this by, shedding light on their distinctive properties.

Expected differences at the community-level are, that successful communities have more intense interactions, and have a more welcoming, or more polarized atmosphere. Even if only differences at the community-level are found, this will be highly interesting, as it can contribute to our understanding of online ‘leaderless movements’ and the impact of community-wide processes. Gaining a more detailed understanding of the early phases of online communities will have great practical applications as well. Not

only might findings inform the creation of online political communities, but they could also help political actors identify rising communities and potential allies early on. Improving our understanding of the ignition of online communities will thus also likely spark off into benefits for our increasingly wired society.