D. Taraborelli and C. Roth (2010), Viable Web communities: Two case studies,
in: G. Deffuant, N. Gilbert (Eds.) *Pattern Resilience*, Kluwer, 2010.

Chapter 5

VIABLE WEB COMMUNITIES: TWO CASE STUDIES

Dario Taraborelli
Centre for Research in Social Simulation
University of Surrey
Guildford GU2 7XH, United Kingdom
d.taraborelli@surrey.ac.uk

Camille Roth
Centre d'analyse et de mathématique sociales
CNRS-EHESS
54, boulevard Raspail
75006 Paris, France
roth@ehess.fr

Abstract Online communities collaboratively generating content have been thriving since decades. The momentum produced by the so-called "Web 2.0" caused a dramatic explosion in the popularity and variety of such communities. Despite a large interest in the research community for online social networks, little has been done to characterise the viability of content-based online communities and identify factors that determine their survival. Studying how these communities evolve over time can help identify properties that drive their growth and assess their viability as a function of their internal structure and functioning. In this chapter, we discuss the problem of how to characterise the viability of online collaborative systems and how to measure it. We then present an overview of the dynamics of two distinct types of content-based online communities: communities in peer production systems and social media communities. We describe a number of macroscopic regularities in how these communities evolve and suggest how these regularities can be harnessed in order to define strategies to control community dynamics.

1. The viability of online social systems

Addressing the question of what makes an online social system "viable" requires some preliminary conceptual clarifications in order to define the scope of the present analysis. **Section 1.** of the present chapter is devoted to conceptually frame the problem: we first introduce the notion of a collaborative Web community by considering the properties that characterise it; we then discuss a number of ways in which the viability of these systems can be defined and the challenges faced by empirical research in identifying measurable indicators of viability. In **Section 2.** we present an empirical analysis of two paradigmatic cases of collaborative Web communities and discuss methodological issues emerging from the study of their dynamics from the point of view of viability. We conclude by presenting in **Section 3.** a simple model of viable online communities based on the empirical and conceptual considerations of the first two sections. Viability, we argue, is a notion that is hard to frame in the case of social systems. By discussing alternative characterisations of this notion and illustrating how to tackle it empirically, this chapter eventually aims at offering methodological insights into the study of viable online social systems. These insights, we submit, are a precondition for the application of formal definitions of viability and resilience (see Chapter 2 of this volume) to realistic models of how collaborative systems function.

1.1 Definitions

Online societies and online communities. In this chapter we will focus on a particular kind of online social systems that we can characterise as content-based, collaborative Web communities.[1] Such communities are defined by the fact that users joining these communities participate in the collaborative production of content, whether in the form of peer production (such as in the case of *wikis* or *open source development*), collaborative annotation (as in the case of *social bookmarking* and *collaborative filtering*) or media sharing (as in the case of *social media* platforms).

Web-based platforms offer particularly appealing conditions to study the nature and dynamics of collaborative groups for two main reasons. On the one hand, such systems offer the possibility of empirically studying different aspects of user behaviour at a large scale though the extraction of online datasets via dedicated tools, such as programmable interfaces (or APIs).

On the other hand, these systems offer a particularly suitable ground for the purpose of the present discussion as they are often embedded in what we may call an "online user society". Agents in an online society can be characterised as users with a unique online identity. The active user base of platforms such as *Facebook*, *Flickr*, or the *Wikipedia* can be taken as an example of an online society.[2] Users of such platforms can freely participate in discussion and content production, establish links to other members and create and maintain affiliations to the variety of communities that these platforms support.

Systems supporting an online society are ideally designed to allow the researcher to compare and assess the respective performance of communities that, although different in structure and organisation, tap into the same user base. At the microscopic level, one can observe how the social and affiliation network as well as the participative behaviour of individual users evolve over time. At a mesoscopic level, one can observe how communities evolve over time, how they recruit members and how their structure affects participation. At a macroscopic level, one can study the evolution of an online society as a whole. In this sense, communities targetting the same user base can be seen as competing with each other to recruit members, a condition that makes it possible to study aspects of group dynamics that are often inaccessible in offline communities.

A taxonomy of user-centred relations. Before discussing the functioning of online communities, we define several distinct classes of relations involving users of these systems. First, members can entertain multiple group affiliations at the same time — given the ease of joining a group, affiliation should not be taken as a straightforward indicator of active participation. Second, members can create explicit relations to other members in the form of "contact" or "friend" links. Third, members contribute *content* to a shared pool of resources maintained by the community; contributed content can consist of text, code patches, media or even, at a more meta level, of annotations which provide information about the other types of items. These classes of user-centred relations involving content, users and groups, indicate the large range of ways in which a collaborative community can evolve and the multiple forms that member participation can take. In order to tackle the question of the viability of online social systems, it becomes particularly important to be able to compare the performance of different communities on the basis of their structure and internal functioning.

1.2 Defining and securing viability

Collaborative communities built on top of an online user society face a number of risks that potentially threaten their survival. Peer production systems, for one, typically die of inactivity for an insufficient number of valuable contributors or, conversely, whenever quality assessment becomes unmanageable due to content explosion or ineffective measures against vandalism. The governance of such communities has been based so far on best practices and recommendations, as empirical evidence on the impact of specific policies on how these communities evolve in time is still relatively scarce.[3]

Addressing the problem of the viability of these systems and their governance, therefore requires understanding: (1) what characterises these communities as "viable" and (2) once a viable state is explicitly defined—what policies can be devised in order to achieve or maintain this state. To answer the first question we need to make a number of conceptual distinctions.

The goal-directedness of viability. As opposed to biological systems (e.g. such as an ecosystem or a colony of microorganisms) whose viability can be defined in a straightforward way as a function of what prevents the system from extinguishing some clearly identifiable resources, understanding viability in social systems (as characterised by the complex web of relations described in the previous section) is significantly more challenging. The main reason why this is the case is the fact that the concept of viability for a social system is intrinsically *goal-dependent*. To the external observer a community designed to promote the systematic destruction of its member base (for instance by virtue of rules and policies that automatically cause the removal of older members and enforce a compulsory member turnover) may look as an example of an ultimately non-viable social system. However, insofar as policies are effective in fulfilling the community's goal over time and in controlling the community dynamics in a way that is functional to the achievement of that goal, one may say that such a community is (at least in a technical sense) viable. An extremely relativist view may then conclude that any attempt to "measure" the viability of a collaborative community (or to compare the respective viability of a set of communities targetting the same user base) is set to fail unless the performance of a community is pitted against its specifically stated goals. This approach, however, may conceal some useful characterisations that one can make about communities being viable, especially on quantitative grounds, regardless of what their goal is.

Dimensions of viability.

1) Viability as membership growth. The first typical way to assess the performance of a collaborative community consists in looking at the *growth* of its members base over time. Unsurprisingly, sheer growth is one of the main aspects on which the quantitative literature on collaborative online communities has focussed so far (see Almeida et al., 2007; Voss, 2005; Godfrey and Tu, 2000; Godfrey and Tu, 2001; Lam and Riedl, 2009; Mislove et al., 2007; Roth et al., 2008; Taraborelli and Roth, 2010). Studying the viability of a collaborative community by taking into account its population growth (and the speed thereof) is a valuable approach as long as growth *per se* is a desirable feature for the sustainability of the system. In some cases, however, an uncontrolled growth of participants is likely to lead to the breakdown of collaboration. This is particularly sensitive when growth in population and the growth in the content that a community produces start to diverge: we will address this issue later by referring to the notion of the attentional span of the members of a community. Even if we take the absolute growth in members of a community as an indicator of its performance in securing a solid user base, we need to consider how its growth compares to that of other communities, how it relates to the actual turnover of its members and how it is affected by different processes through which new members are recruited.

- In cases in which membership does not imply exclusive affiliation, the growth of a given community should not (at least not in principle) be directly affected by the **comparative growth rate** of other communities built on the same user base. However, this is unlikely to be the case as, even in an online society with a constantly growing number of users, communities are *de facto* competing for members. What is crucial is then to understand the nature and effectiveness of processes by which communities manage to secure their membership.
- Two communities may perform equally well in growing their overall membership over time, however they may differ significantly in how good they are at maintaining existing members, or controlling **member turnover**. The same net growth in population can be the result of (*i*) a slow, steady growth in members without a significant drop-off of existing members or (*ii*) the result of a high turnover, whereby the number of new recruits outweighs the number of losses. Considering the role of turnover beside sheer population growth is reasonable if we assume that, while turnover is essential to secure renewal within a community, a high rate of turnover will disrupt the continuity within the community, i.e. a condition that might be needed in order to preserve the possibility of norm transmission within the community (see Forte et al., 2009).
- A final important aspect related to membership growth is the **variety of recruitment processes** that may result in the same observable increase in the population of a given community. The way in which communities recruit members—whether by tapping into the individual social networks of their current members (*social recruitment*) or by focussing on the interests of the new recruits (e.g. *recruitment by homophily*)—has crucial consequences. As we will see in the second part of this chapter, social recruitment processes are likely to reach a point after which the community becomes too cohesive to allow a significant turnover in its membership.

In sum, different aspects in the growth of a community's membership can be relevant to its viability: the rate of population growth over time, the relation between growth in content and growth in population, the rate of member turnover, the mechanisms underlying recruitment and population growth.

2) Viability as participation. The second challenge in defining the viability of collaborative online communities consists in understanding what diversifies community membership in terms of individual participation and how this, in turn, affects how a community thrives at a macroscopic level. A community may be said viable if it manages to secure a minimum number of participants committed to perform specific tasks within the community that are essential to the achievement of its goals. As self-allocation of effort is a typical feature of peer production systems (see Benkler, 2002), individual incentives to participation play a central role in specifying the conditions that make a community viable. A lack of balance between regular members and administrators, a redundancy of effort by community members, a lack of participants devoted to quality control and norm enforcement are examples of ways in which patterns of participation can be disruptive for a community striving to achieve a specific goal. A detailed analysis of the drivers of participation in peer production is beyond the scope of the present work, but we should mention three key aspects, that are relevant to the problem of the viability of these systems:

- the relation between individual **motivation and participation**. What actually drives users to participate in online peer production is an issue that has been addressed by several authors (Hars and Ou, 2002, Benkler, 2002, Benkler, 2006). Status recognition and expertise within a community are key factors in strengthening membership, however to what extent these internal factors are effective in terms of recruitment and task allocation is an issue that still has to be explored. A community is viable if it manages to channel individual motivation in a way that is functional to achieving a proper division of labour.

- the distinction between **passive membership and participation**. Several communities thrive despite a relative low number of active members. This suggests that communities may have a high potential of recruitment even though only a small proportion of their membership is responsible for actual content production. This is actually a known property of peer production systems such as wiki-based communities where a majority of participants only marginally contributes to actual content production as opposed to quality control (Kittur et al., 2007, Roth, 2007) and in all those systems where the existence of a community of "lurkers" is a vital condition for the performance of a community (Nonnecke et al., 2006). In this respect, the performance of a mature peer production system may actually depend on more subtle factors than the sheer proportion of actively contributing members: fighting vandalism, for one, seems to depend more on the number of passive watchers regularly monitoring content than on a small proportion of active contributors.

- the relation between **competitiveness and participation**. One final issue that is key to defining the viability of collaborative systems is understanding to what extent these systems are in mutual competition. Depending on the underlying design, two communities competing on the same topical or social "niche" (McPherson, 1983) may well independently thrive without affecting each other's performance; conversely, one may observe a migration of users from one community to the other (which will affect the viability of each community if measured by membership growth), or a change in participation rate not involving an actual migration or termination of membership (which will only affect the viability of a community as measured through participation metrics).

Understanding drivers of participation and types of participation is crucial to characterise those communities who may be effective at recruiting members but unable to secure a proper division of labour, such as content production versus content maintenance by active users. In this respect, both role diversification and proportion of active participants represent critical variables to measure the maturity of a community and its potential viability.

Achieving viability. Once we have identified a specific standard of viability as a function of what aspects of an online community we wish to focus on, the next question is what policies are available to effectively achieve and maintain that standard. The question bears on the delicate issue of control policies in peer production systems, or how to devise an appropriate governance model for systems where individual effort is typically self-allocated and in which traditional organisational structure is not applicable (see Benkler, 2002; Forte et al., 2009. We review in this section three classes of factors that are instrumental in *controlling* the dynamics of a community towards reaching a viable state.

1) Viability and quality control. The content dynamics of a collaborative community deserve as much attention as membership dynamics from the perspective of its viability. As we suggested, a divergence in the respective growth of content and population may easily lead to a breakdown of collaboration. This makes quality control policies one of the main factors behind the successful performance of collaborative communities, a problem that is particularly sensitive in communities where content is collectively curated in order to meet some shared quality standards (Forte et al., 2009). The most extensively studied case of collaborative quality control in online peer production is by far the Wikipedia (see Halim et al., 2009, Stvilia et al., 2008, Suh et al., 2008, Wöhner and Peters, 2009, Kittur et al., 2007, Kittur et al., 2009)—a case in which the effectiveness of quality control policies, the social processes involved in their enforcement and the general distribution of labour among contributors have been empirically analysed. Factors that may drive a community to achieve a viable state from this perspective include, first of all, the balance between inclusiveness and quality control: too strict quality control policies may drive away potentially valuable contributors, but the same effect can actually result from a demographic explosion or by loose or poorly effective quality control mechanisms.

2) Viability and governance. Despite the fact that collaborative communities and peer production systems are often referred to as systems that accept unconstrained contributions from their members, they often implement forms of "soft governance" and hierarchical organisation that help maintain a community active and content production focussed. Governance-related factors include solutions controlling *content production* (i.e. what kind of content is allowed within the community's product) as well as systems controlling *member affiliation* (i.e. who can join or who can perform specific tasks). In both cases governance solutions can be enforced *a priori*, by imposing limits onto the production of participation or content production, and/or *a posteriori*, by removing inappropriate content or removing existing members not complying with community norms.

3) Viability and sociability. A third class of factors relates to social interactions, especially in collaborative communities built on top of online networking services. Even though governance measures can essentially exert indirect control on the shape of social interactions, a number of properties of the social network of the members of a group have proved to show an important role in controlling the dynamics of collaborative communities — as a driver of recruitment of new affiliates, or, on the contrary, as an obstacle towards further growth, whenever high social cohesiveness hinders the affiliation of users not belonging to the social neighbourhood of a group. In terms of peer production, we may also expect that groups whose members maintain at the same time a too large number of social ties will start showing symptoms of breakdown in collaboration or in the ability to effectively monitor content production, which in turn will threaten the viability of the system.

2. Two case studies

In this section we illustrate how the above conceptual distinctions can be put to work in the empirical study of the evolution of collaborative systems. We focus in particular on the relation between growth-related viability and control factors by looking at properties that spur or regulate growth in two paradigmatic cases of collaborative systems: peer production systems and communities in social media.

2.1 Peer production systems

Wikis are, in a broad sense, websites whose content can be contributed and modified by any user in a collective and collaborative fashion. As such, they represent one of the most prominent examples of Web-based peer production systems. The most famous and possibly the most successful of these websites, *Wikipedia*, has attracted a substantial interest in the research community in recent years (Lih, 2004; Anthony et al., 2005; Bryant et al., 2005). The Web has seen, however, several thousands of other wikis thrive and proliferate, with varying degrees of success: some recruit many users, achieving sustainability with established role distributions, frequent updates and efficient measures against vandalism, while others fight to attract contributors. Wiki-based communities can have distinct policies or scope but be equally sustainable, or have identical policies but die for a variety of reasons; all endeavoring to survive within what may be called the "wikisphere".

This first case study consists of an exploratory investigation of some factors likely to account for diverse wiki destiny and viability, in terms of technical, social and structural features. In this context, we understand "viability" as dynamic sustainability of both population and quality content: in other words, a viable wiki should be able to grow in terms of articles and users in such a way that the whole content can be maintained by a sufficient number of users. Our aim is however not to provide a formal definition of an appropriate notion of viability for wiki-based communities, but rather to present a detailed descriptive analysis of the demographic and structural dynamics of a large sample of wikis as an empirical basis for further research. In particular, we discuss these results in light of the role played by governance measure in affecting the viability of these communities, moving beyond the *Wikipedia* case.[4]

Wiki dynamics. Various governance systems and software parameters, i.e. technical and social constraints, define a landscape wherein each online community is settled, grows and lives. How can the growth and evolution of such communities be assessed? As content-based online communities, wikis mainly evolve in two dimensions: (i) contributors, who may or may not constitute an active community; as discussed e.g. by Bryant et al., 2005; and (ii) pages, which may or may not amount to authoritative or useful content; as demonstrated for example by Giles, 2005.

Users and pages are likely to obey a dual dynamic: while more users may contribute to more pages, content proliferation seems to require more attention from users. As a first approximation, it may thus seem judicious to assess the healthiness of a wiki through these variables, taken as demographic indicators for its actual growth and activity.[5]

To our knowledge, the present case study represents an original longitudinal analysis of the content and population dynamics of a large set of wikis. As well as almost always focusing on *Wikipedia*, previous quantitative wiki research has mainly examined the topological structure of underlying interaction or hyperlink networks (Capocci et al., 2006; Zlatic et al., 2006) or article-level features (Brandes and Lerner, 2008; Wilkinson and Huberman, 2007), with little interest in the specific dynamics of the demographic determinants themselves (with the exception of Kittur et al., 2007 who investigates *Wikipedia*'s demographics of casual *vs*. committed contributors).

Dataset. We constructed a dataset made of simple statistics gathered for a large number of MediaWiki-based wikis,[6] which enabled us to consider the same set of variables across all wikis and make sure these variables were generally available. The data was collected over the period August 2007–April 2008 from a publicly-available database[7] totalling 11 500+ wikis. We applied further restrictions on this dataset, as described in Roth et al., 2008.

To sum up, the final, "clean" dataset that we considered for this study is made up of about 360 wikis, all of which have an initial population between 400 and 20 000 users, are *not* hosted at some specific 'wiki farms' that do not report useful data, and which do not have major discontinuities in the daily change of their population or content. As such, we assume this subset to be representative of a homogeneous sample of wikis having a relatively sizeable yet not exceptionally large base of registered users—the latter being a hallmark of typical outliers (such as the English Wikipedia) in the wikisphere. Besides, the exclusion of discontinuously-growing wikis ensures that the observed dynamics are due to genuinely bottom-up user-driven behaviour rather than top-down administrative intervention or external attacks. In this sense, and from a viability theory perspective, we are thereby focusing on the *autonomous dynamics* of such systems.

Variables. We considered a set of four raw quantitative variables: *population size* (U), measured by the number of *registered* users; *content size* (P), measured by the number of so-called "good" pages (i.e. actual *content* pages excluding default pages created by the wiki engine), hereafter indifferently called "pages", "good pages" or "articles" ; *administrator population* (A), the number of users who are granted "administrator status", i.e. special rights to modify sensitive content and perform maintenance activity; and *editing activity* (E), measured by the total number of edits. We also included one qualitative variable indicating the presence of an access control mechanism: *editing permission* (R), i.e. the possibility of creating a page for unregistered/anonymous users. R is either 1, "anonymous editing allowed", or 0, "registered users only". However simplistic these variables may be, they provide key indicators of the global dynamics of a wiki, and shed light on diverse aspects of its structure and evolution. We collected the values of these variables for each wiki every day and over a period of 250 days, i.e. approximatively 8 months.

Structural metrics. Wiki dynamics were studied as a function of a number of structural metrics based on the above variables, and that we can broadly categorise in two broad types of independent variables. On the one hand, **descriptive indicators**, i.e. variables on which wiki administrators have no direct control: *user activity* (i.e. the proportion of edits per user E/U), *user density* (i.e. the proportion of users per page U/P), and *edit density* (i.e. the proportion of edits per page E/P). On the other hand, **governance factors**, variables that wiki administrators can directly control: *administrator ratio* (i.e. the proportion of users who are granted administrator status A/U), *administrator density* (i.e. the proportion of administrators per page A/P), *editing permission* (R). See Tab. 5.1 for a summary.

Table 5.1: Wiki metrics used as independent variables

	U/P	user densitiy
descriptive metrics	E/U	user activity
	E/P	edit density
	A/U	administrator ratio
governance metrics	A/P	administrator density
	R	edition permission

We subsequently assessed wiki dynamics by comparing their diverse growth paths with respect to a set of independent variables. 'Growth' is defined in terms of population and content size variation (see Tab. 5.2): user growth G_U (resp. page growth G_P) is the ratio between final and initial populations (resp. content sizes): $G_U = U_{\text{last}}/U_{\text{first}}$ (resp. $G_P = P_{\text{last}}/P_{\text{first}}$). For each continuous variable, instead of carrying out a delicate analysis by dealing with clouds of points, we adopted a more insightful approach by dividing wikis into five quantiles, each including exactly 20% of all wikis in the clean dataset. We then computed and compared growth rate means over all wikis for each quantile. This representation was applied to all the above-mentioned variables, except for R where there are only two "quantiles" (0 or 1), enabling us to distinguish population quantiles on a unique graph.

Determinants of wiki dynamics. The results suggest that different structural and governance-related factors have significant correlation with — and

Table 5.2: Wiki growth indicators

Wiki growth	G_U	population growth ($U_{\text{last}}/U_{\text{first}}$)
	G_P	content growth ($P_{\text{last}}/P_{\text{first}}$)

plausibly, in some cases, effect on — the content and population dynamics of a wiki:

- **Significant descriptive indicators.** Figure 5.1 shows the effect of **user activity** (measured as the proportion of *edits per user*) on growth rates. The results suggest that user activity correlates very strongly with wiki growth, not only in terms of content production (which is to a certain extent unsurprising) but also new member recruitment. The effect becomes stronger with initially more populated wikis: the more users are actively editing, the more a wiki grows in content *and* population.

- **Significant governance factors.** Turning to governance features, we first analysed the effects of the **administrator density** on wiki dynamics by examining the impact of the overall proportion of administrators per page. Figure 5.2 shows that having a relatively high number of administrators for a given content size is likely to reduce growth. There is a strong effect of the proportion of admins per page both on user and page growth. For instance, while the last quantile of admins/page ratio has near-zero growth rates over 8 months, the first quantile shows high overall rates (∼+50% for users, ∼+25% for pages). This effect may be interpreted as the impact of strong governance activity on the proliferation of content and users.

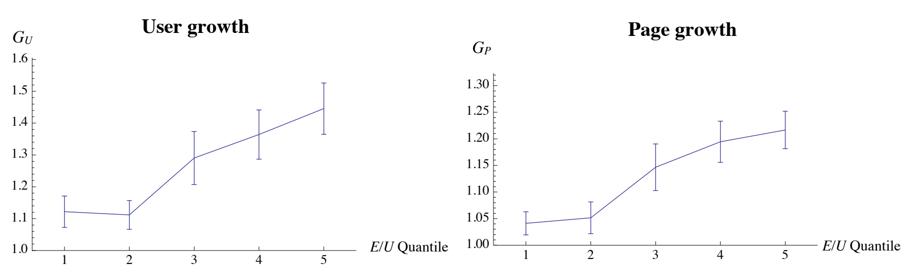

Figure 5.1: Growth landscape with respect to *user activity*, i.e. the proportion of *edits per user* (E/U).

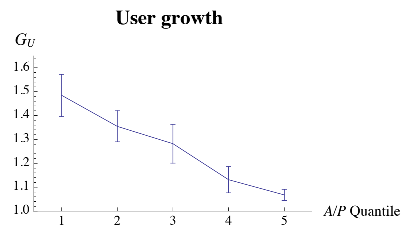

Figure 5.2: Growth landscape with respect to the proportion of *admins density* (A/P).

We identified another significant effect when we considered **editing permission**. As a binary variable, the editing permission variable generates only two groups of wikis (wikis that allow anonymous editing *versus* wikis that restrict editing to registered users only). The growth landscape is consequently limited to a one-dimensional comparison over population quantiles. The results in Figure 5.3 show that for both dimensions—population and content—having no access control is likely to favor growth. While a stronger page growth is quite unsurprising in wikis where no registration is required, the fact that this factor also fuels user registration is more puzzling. One might expect that if users can participate without the need of registration, few would be inclined to register. Our results suggest that on the contrary wikis with unrestricted registration trigger participation more easily than wikis that restrict access.

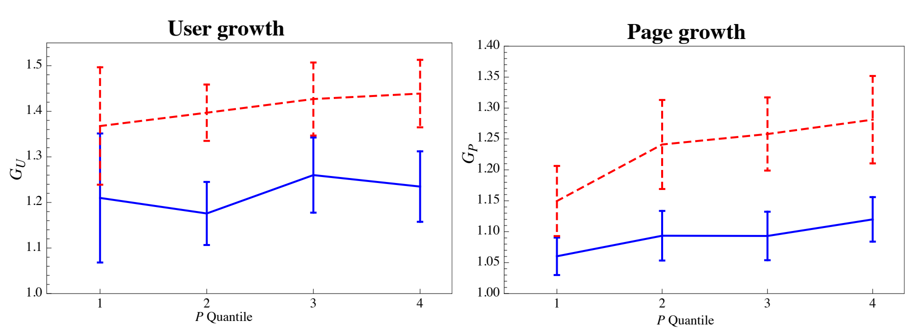

Figure 5.3: Growth landscape with respect to *editing permission* (R): *red dashed* refers to anonymously editable wikis, while *blue solid* to wikis editable by registered users only.

We also considered two indicators that showed a markedly milder correlation with wiki dynamics. On the one hand, we found that **edit density** (i.e. edits/page) correlates in a moderately negative way with user growth—with a relatively stronger effect depending on initial population size—while there is surprisingly no significant correlation with page growth. On the other hand, higher **administrator ratios** (i.e. admins/user) have no significant effect on content or population growth.

Figure 5.4 summarises the correlations found between growth rates and each of the variables we considered, by comparing the gain in the population and content sizes between the last and the first quantile for each variable (variables in Figure 5.4 are ranked from the most positively to the most negatively correlated). If we focus on *structural aspects* of wikis, we note that the higher the ratio of *edits per user* the faster the wiki grows, both in terms of content and population. Wikis with very active user communities are not only likely to grow in content, but also to attract a large number of new contributors. This result contrasts with the opposite effect produced by high user density per page.

As far as *governance factors* are concerned, we observed the singular fact that population growth is in average more than 20% faster for anonymously editable wikis. This seems to support the intuition that less barriers favor population growth. Furthermore we observed that, while too many administrators per page may hinder the growth of a wiki (in terms of content size), the proportion of administrators per user does not appear to show a significant influence on growth. In all the above cases, we observed a striking correlation between content and population growth.

This approach broadly draws attention to the remarkable intertwinement of population and content growth in a relatively large sample of (wiki-based) online communities, and constitutes a first contribution towards more comprehensive research on factors behind sustainable wiki communities, beyond

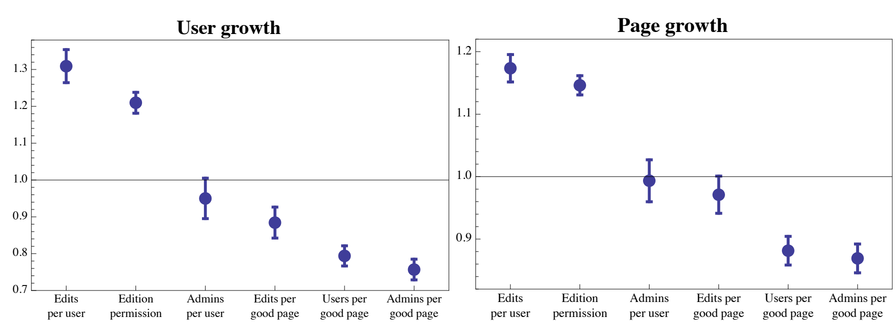

Figure 5.4: Comparison of growth rates between last and first quantiles, for each variable.

the dominant example of the *Wikipedia*. In particular, we endeavored at connecting simple quantitative features of these online groups to more qualitative characteristics mainly pertaining to simple organisational properties, including distribution of roles, modes of regulation and access control. In a more pragmatic perspective, it basically constitutes an overview of indicators that wiki communities should take into account in order to control their demographics, by paying specific attention to some variables and acting upon them when possible, while neglecting others.

From the perspective of viability theory, the dichotomy we propose between descriptive indicators and governance factors is meant to reflect the traditional distinction between *autonomous dynamics* and *control features*. In this respect, our results offer empirical grounds for the design of realistic models of the demographic evolution of these communities, the characterisation of their viable states and of factors that control their dynamics.

2.2 Social media communities

Flickr.com, one of the most popular photo and video sharing services, represents another ideal case for the study of online community viability, focusing here on the joint effects of content-based interaction, group affiliation and social network dynamics. The platform supports a dedicated infrastructure for the creation of communities of interest or “groups”, which represents an ideal testbed for studying group viability issues, as well as, more broadly, the effect of *user-to-group* affiliation links on user behaviour and social interaction among users. The user model of Flickr additionally allows the creation of (*user-to-user*) “contact” links that can provide a direct insight into user-centered social networks; it also allows interactions among users that are mediated by content (such as commenting on a picture or marking a picture as a “favorite”), hence offering the opportunity to study social behaviour mediated by *user-to-content* links; finally, thanks to a rich and extensively documented API[8], Flickr enables the extraction of large datasets that can be used to study social dynamics at each of these levels of description (content, users, groups).

Flickr attracted a fairly large attention in the research community. Most studies used Flickr as a large data source to study tagging behaviour and folksonomy (Marlow et al., 2006; Nov et al., 2008; Plangprasopchok and Lerman, 2009; Sigurbjörnsson and van Zwol, 2008). A smaller number of works, more relevant to the present analysis, focused on aspects of social interaction and group-driven behaviour (Lerman and Jones, 2007; Mislove et al., 2008; Leskovec et al., 2008; Cha et al., 2008; van Zwol, 2007; Valafar et al., 2009) . Reviewing the contribution of the literature on the understanding the functioning of communities in online social media is beyond the scope of this chapter. In this section we focus on the relative performance of groups at attracting members and securing content, an issue that has been addressed only in a tangential way by the above studies.

Flickr groups. A central social feature of Flickr, i.e. *groups*, has attracted to date a modest attention in the literature, even though it is estimated that a large part of content-mediated interactions and social interactions happens via groups. Flickr groups are of particular interest to the present analysis because, as opposed to purely user-centered social networks, they can be described as communities of interest driven by shared content. Flickr groups are specifically designed to enable *collaborative* content production and dissemination. In order to share content with the members of a group, a user is explicitly required to submit it to the group. Furthermore, groups have a governance structure consisting of at least one administrator (by default, the group creator) and an optional number of moderators. Group admins and moderators can control the rate and type of submitted content that is shared in the group, via *moderation* tools, post-submission *pruning* or *throttling* (i.e. limiting the number of posted items over a given period of time). These features make Flickr groups ideal candidates for research on content-based online social behaviour and on the coevolution of social and affiliation links. Previous research already partly addressed the role of Flickr groups from this angle (Mislove et al., 2007, Prieur et al., 2008, Schifanella et al., 2010, Negoescu and Perez, 2008, Backstrom et al., 2008, Zheleva et al., 2009).

All in all these various results suggest that group formation processes in content-based communities arise from the joint effect of a large number of factors that cannot entirely account for the evolution of a group when considered on an individual basis. The question that we wish to ask is how these different factors interact in affecting the evolution in content and population of a group.

Variety of group dynamics. There is a striking variety in growth patterns of Flickr groups when observed over time, even if we focus on macroscopic indicators such as population and content variations: some groups are characterised by a steady population growth accompanied by a null or negative content growth (which may *prima facie* suggest tight moderation or regular pruning of content); other groups rapidly grow in content but vary slowly or remain virtually constant in population (suggesting the use of groups as "dumps" of pictures with little recruitment of new members); other groups show fluctuations in both content and population (suggesting a significant portion of members who leave the group when no more active); finally, groups may display sudden bursts of growth in content and population and remain subsequently inactive for long periods (which may be the case for groups about recurring or temporally discrete events).

Groups also substantially vary in member *turnover*, i.e. the portion of a group's population that is replaced by new members joining the group over time while former members leave. Some groups have a relatively low turnover, suggesting that members tend to stick in the group and are reluctant to leave, while other groups have much faster member replacement rates.

One possibility to come to grips with this variety in global dynamics might then be to ask whether groups can be broadly categorised in a qualitative way into distinct typologies, considering for instance how content specificity or content policies affect the overall group evolution over time. The alternative approach that we take in the present study consists in assuming that similarity in temporal dynamics can be traced back to group similarity in terms of structural features.

Regardless of content, each group can be characterised as occupying at a given time a region in a multidimensional space of properties defining its demographic profile, its structure and its governance mode. These properties can pertain to a group as a whole or refer to aggregate properties of its members, such as their average degree or group affiliation spread.[9] The temporal dynamics of a group can then be studied as a trajectory across this space. Our study aims to find regularities in the observed temporal dynamics of a large set of groups by assuming that a number of initial properties of these groups can be explored as predictors of their macroscopic evolution — which, again, is taken as a preliminary description of some dimensions of their viability. The literature on group affiliation dynamics offers a number of suggestions as to how groups are generally expected to evolve over time as a function of their size, structure and properties of their membership:

P1. Larger groups tend to grow faster than smaller groups, in virtue of a preferential attachment principle.

P2. Cohesive groups tend to recruit less new members than weakly cohesive groups, because of a stronger social closure (or "cliquishness"), which also results in an increased membership inertia and less user turnover.

P3. Groups whose members are sociable tend to grow faster and attract more contributions than groups whose members have a relatively small number of friends.

P4. Highly curated groups tend to grow slower in content but faster in population because of the competitiveness produced by higher content selectivity.

P5. Groups whose members belong to many other groups grow less in content than groups with members that belong just to a few groups.

Each of these hypotheses can be empirically explored, by considering the observed growth rates over a specific time frame as a function of characteristic properties of a group.

Dataset. The data used for this study consists of a sample of 9,360 *public* Flickr groups whose variations were tracked on a daily basis for a period of 1 month between June and July 2009. The data was obtained via *Flickr Group Trackr*[10], a public Web service that we developed in order to allow Flickr group members to track the daily evolution of their community. For each group registered to the service, *Flickr Group Trackr* pulls a series of statistics from the Flickr API on a daily basis, including: size of the group pool (or number of pictures uploaded to the group), population, privacy level, moderation properties, throttling type and level. Changes along any of these variables can hence be identified with a precision of 24 hours. It should be noted that we did not consider group activity data related to discussions in group forums as this data are not available via the Flickr API. The dataset thus obtained from *Flickr Group Trackr* was complemented with a static snapshot of the same set of groups providing data on: *user-to-group* affiliation links and *user-to-user* contact links.

The dataset was filtered in a number of ways to obtain a more homogeneous sample. We limited our analysis to a set of medium-to-large groups with a population range of 100 to $100,000$ members; this restriction was introduced to avoid biases in the analysis due to the presence of small groups ($u_0 < 100$), whose dynamics are too dependent on the behaviour of individual members to allow any useful generalisation.[11] To capture the natural dynamics of these groups we also introduced a capping on the maximum daily growth rate in content and population, as we did for wikis, excluding those groups displaying an instantaneous growth of more than 5% of their pool size or population (possibly resulting, again, from extrinsic events such as contests or administrator bulk decisions). Groups that switched to *private* access control mode during the tracking period were also excluded from the sample. As a result of these restrictions, the final dataset used here consists of $9,167$ groups.

The dataset also contains a complete snapshots of the population of each of the tracked groups at t_0 as well as the complete list of contacts and affiliations for each member of these groups. The union of members of the groups in the dataset spans a total population of 1,267,874 unique users. Group pools size and group populations in our dataset follow a log-normal distribution.[12]

Variables. The metrics that we used as independent variables to study the drivers of group dynamics throughout the present study are described in Tab. 5.3. Among demographic metrics, *ms* (membership spread) indicates the number of other groups a group member is affiliated with, averaged over the whole group population. Among structural metrics or metrics related to topo-

Table 5.3: Flickr group metrics used as independent variables

demographic metrics	u_0	number of group members at time t_0
	p_0	number of photos in the group pool at time t_0
	ms	average membership spread of group members
structural metrics	k	average directed degree of group members
	c_3	average clustering coefficient
	r	reciprocity index
governance metrics	adm	number of group administrators
	mod	number of group moderators
	μ	moderation filter
	θ	throttling index

logical properties of the group social network: k refers to the direct degree for group members calculated on contact links that are internal to the group social network; r measures the proportion of *reciprocated* or symmetrical contact links within the group and per group member, averaged over the group population. Among governance metrics: mod indicates the number of superusers other than administrators who can accept photos submitted to the group's moderation queue; μ indicates the presence of a moderation queue, by which photos submitted to a group are reviewed by moderators before being published in the group pool; θ is a quantitative indicator of the maximum number of photos that can be contributed to the group per time period (day, week or month), also denoted as "throttling index".

Group growth indicators can be defined in multiple ways. Growth can be assessed in absolute terms as the difference in the total number of members and photos between t_0 and t_1, i.e.: $u_1 - u_0$ and $p_1 - p_0$ respectively. Alternatively, one may focus on relative growth or "growth rate" over the observation period, or the variation in members and content normalised by the initial size of the group: $\frac{u_1 - u_0}{u_0}$ and $\frac{p_1 - p_0}{p_0}$. Finally, one may consider the actual turnover or the number of unique users who joined (u_+) and leaved a group (u_-) over the observation period. The turnover itself can be considered in absolute $(u_+ - u_-)$ or relative terms $\frac{u_+ - u_-}{u_0}$.

Table 5.4: Flickr group growth indicators

Group size variation	Δu	absolute population variation $(u_1 - u_0)$
	Δp	absolute content variation $(p_1 - p_0)$
Population turnover	u_+	number of users who joined the group
	u_-	number of users who left the group

For the sake of the present study and contrarily to the wiki case, we decided to focus on absolute rather than relative growth indicators (see Tab. 5.4) for a number of reasons. First of all, we wanted to take all groups at face value as equally prone to recruit new members and measure size-dependent effects as only one among several possible assumptions on growth driving factors. Although other studies showed that the size of a group population plays a central role in the recruitment of new members (Backstrom et al., 2008), this assumption can be challenged on the basis of the significant number of group members who do not appear to have any social connection with other members ($k = 0$). Evidence of the existence of such members, as pointed out by Zheleva et al., 2009, suggests that "social recruitment" is only one among possible mechanisms that attract new members to a group. Second, we wanted to study the specificity of member turnover as indicators of a stable or volatile community, and for this reason we also decided to opt for absolute figures as opposed to relative growth rates. A final reason not to focus on relative growth rates was that results using these rates as dependent variables were not statistically significant in several cases, suggesting that for the timeframe that we considered absolute variations were the most appropriate to focus on.

Aggregate analysis of growth-driving factors. To investigate the joint contribution of demographic, structural and governance-related factors on the temporal dynamics of groups, we performed a regression analysis of absolute group growth over the whole observation period as a dependent variable. We used four different models aiming at measuring the respective effect of a series of independent variables on absolute user variation (Δu), content variation (Δp) as well as member turnover (u_+ and u_- respectively). We used the initial population and content size as control variables in each of the models (see Tab. 5.5 for the detailed list of variables included in each model). The general regression equation underlying each model (barring specific variable exclusions) is:

$$\begin{aligned}\log(y) = \lambda_0 + \lambda_{u_0}\log(u_0) + \lambda_{p_0}\log(p_0) \\ + \lambda_r(r) + \lambda_{c_3}(c_3) + \lambda_k\log(1+k) + \lambda_{ms}\log(ms) \\ + \lambda_\mu\mu + \lambda_{mod}\log(1+mod) + \lambda_\theta\log(\theta) + \lambda_{adm/u0}\log(1+(adm/u_0))\end{aligned} \tag{5.1}$$

We thus considered a linear regression of the logs of each variable, when applicable and relevant: logs were essentially used for quantitative variables spanning over one or several orders of magnitude (such as u_0) in order to make them comparable in the regression with variables evolving in e.g. $[0,1]$ (such as c_3). For each dependent variable $y \in \{\Delta u, \Delta p, u_+, u-\}$, we started with an equation specified by the full model of Eq. 5.1. Variables corresponding to non-significant p-values were then iteratively excluded, generally resulting in a change in R^2 of less than 1%.

Table 5.5: Results of regression analysis

	POPULATION VARIATION (Δu)		CONTENT VARIATION (Δp)		POPULATION TURNOVER joining (u_+)		leaving (u_-)	
parameter	*value*	*p*	*value*	*p*	*value*	*p*	*value*	*p*
λ_{u_0}	0.87	***			0.77	***	0.78	***
λ_{p_0}			0.94	***	0.11	***	-0.03	***
λ_r	0.99	***	2.09	***			-0.19	**
λ_{c_3}	-1.87	***	-1.73	***	-1.49	***	-1.27	***
λ_k	0.10	*	0.18	***			0.23	***
λ_{ms}	-0.57	***	-0.33	***	-0.43	***	0.35	***
λ_μ					0.08	**	0.07	***
λ_{mod}	0.05	**	0.09	***	0.08	***	0.02	***
λ_θ			0.10	***				
λ_{adm/u_0}					-0.06	***		
R^2		*0.65*		*0.75*		*0.68*		*0.82*

Log-linear regression model on absolute population variation (Δu), absolute content variation (Δp) *(left)* and population turnover measured as absolute number of joining members (u_+) and leaving members (u_-), respectively *(right)*. Significance: *, ** and *** mean a p-value smaller than 0.05, 0.01 and 0.001 respectively.

The results of the regression analysis summarised in Tab. 5.5 indicate some salient effects of various initial properties of groups on their dynamics. For a given dependent variable, an empty cell indicates that the corresponding independent variable had eventually been excluded from the regression. If we focus on **population and content growth**, we first notice a (somewhat unsurprising) correlation in the effect of different variables on population growth on the one hand and content growth on the other hand, which is consistent with the above findings in wiki-based communities. As to structural/demographic factors, we observe indeed that population (u_0) and pool size (p_0) are important drivers of absolute growth: the larger the population of a group, the stronger its absolute growth over the observation period (consistently with **P1.**). The average spread of group affiliation for group members (*ms*) displays a negative correlation, suggesting that groups whose members also belong to many other groups tend to grow slower and the effect is actually stronger on population growth than it is on content (**P5.**): this is consistent with the idea that groups whose members are selective (i.e. choose to join a smaller number of groups) are likely to attract more members than groups that mostly function like content dumps for occasional members. In terms of topological properties of the group-centered social network, we observe that cohesiveness as measured by the average clustering coefficient of the group-based network (c_3) has a remarkable negative correlation with growth (**P2.**) and is by far the variable displaying the strongest effect across all analyses. Conversely, a high rate of reciprocity (r) and a larger presence of (popular) high-degree nodes in a group (k) have the effect of boosting growth (**P3.**). Possibly the most striking finding is the overall negligible effect of moderation properties on the observed growth. In many cases the effect of moderation factors ($\mu, mod, \theta, adm/u_0$) is not statistically significant; in those cases in which it is, the observed effects are considerably weaker than those related to other group properties, which is partly at odds with our expectations (**P4.**).

The analysis of factors affecting **member turnover** provides further insights. The strongest effect on turnover is that of cohesiveness, that not only appears to hinder new recruits but also to work as a barrier against user drop-off, as indicated by its negative effect on both components of the turnover: as such, cohesiveness (or the "cliquishness" of a group) works as a factor measuring the social inertia of a group membership, suggesting a higher level of commitment by its members that are more reluctant to leave than in less cohesive groups (**P2.**). The level of engagement is also measured by the symmetric effect that affiliation spread has with respect to member recruitment and drop-off: a higher spread increases the probability that more members will be leaving the group and less new members joining.

Individual drivers of group growth. Whereas the regression results can be used for a global assessment of the contribution of different factors to the dynamics of a group, we can address each of the hypotheses P1–5, following the methodology adopted before for wiki communities. We tackled the implications of this regression model on each hypothesis through an analysis of the individual impact of each metric on the observed growth and turnover of a group. Two snapshots for each group were compared at the beginning (t_0) and at the end (t_1) of the tracking period and group growth rates were calculated as the absolute variation in population and pool size between these two snapshots (Δu and Δp respectively). We then ranked groups along each independent variable in 9 quantiles, each containing therefore $1/9$ of the groups in our dataset. The first quantile represents groups with the lowest values for the considered variable, whereas the last quantile refers to groups with the highest values. The analysis of individual effects should be taken as evidence of how effective each factor would be under the assumption that all other factors had an equal effect on growth.

- **P1: Size matters.** The breakdown of the effects of size on the observed growth (Figure 5.5) shows indeed that the expected growth of a group in content and population follows monotonically from its size. This allows us to discard the null assumption that we made that all groups should in

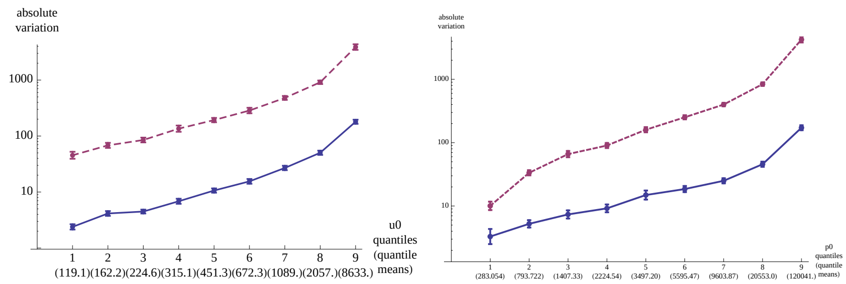

Figure 5.5: **P1**: Size matters. Effects of u_0 and p_0 on absolute variation of users (*solid line*) and photos (*dashed line*).

principle be considered at face value as having an equal probability of attracting new members and new content: size does matter, which can be explained by e.g. herding behaviour.

- **P2: Effects of cohesiveness on group growth.** Figure 5.6 (*left*) shows the breakdown of the effects of cohesiveness on group growth. Consistently with the regression analyses, cohesiveness as measured by the average clustering coefficient for the group-centered network works as a growth-regulating factor. Groups where cohesiveness is high display a higher inertia.

- **P3: Are sociable users growth attractors?** An effect conflicting with cohesiveness is related to individual sociability as measured by the average *within-group* degree of members in the group-centered contact social network (Figure 5.6, *right*). Note that degree has only been measured within groups: a related hypothesis, assessing sociability through a de-

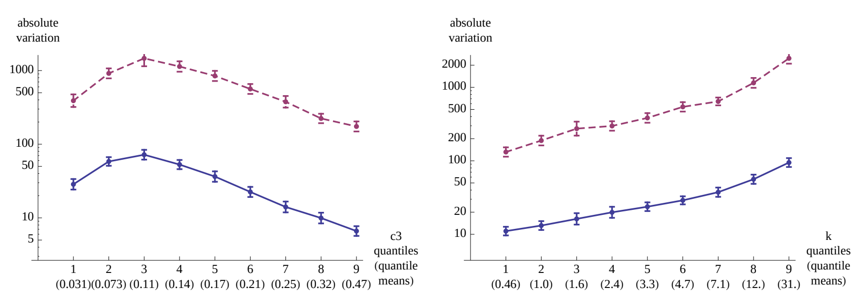

Figure 5.6: (*left*) **P2**: Effects of cohesiveness (c_3), (*right*) **P3**: Effects of sociability on absolute variation of users (*solid line*) and photos (*dashed line*) (k).

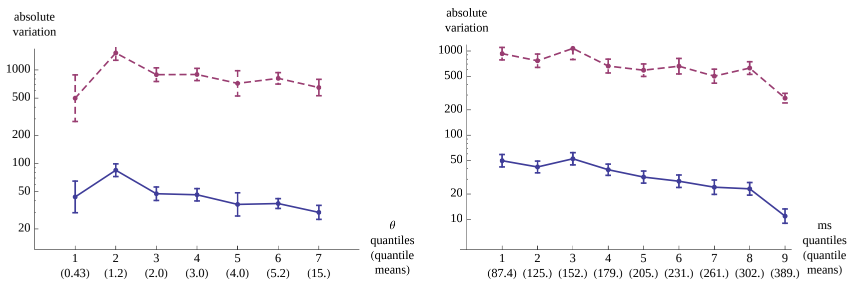

Figure 5.7: (*left*) **P4**: Weak effect of governance, exemplified by throttling (θ). (*right*) **P5**: Level of user engagement measured by affiliation spread (*ms*). Plots represent absolute variation of users (*solid line*) and photos (*dashed line*)

gree computed over the whole network of the Flickr population would actually allow one to answer the question whether groups in which (global) high-degree nodes (or very social/popular users) are concentrated are more likely to attract members than groups where the degree is more uniformly distributed.

- **P4: The poor effects of governance.** Possibly the most striking findings of the present study are the negligible effects of the moderation and governance structure on group growth. Figure 5.7 (*left*) exemplifies the virtually flat growth landscape that emerges as a function of θ. This is not to deny the effectiveness of curators' strategies in actually enforcing norms about content and participation on group members. However, from a purely quantitative perspective, these results suggest that in social media sharing systems social-network factors are likely to drive to a much larger extent recruitment and participation than what group administrator and moderators can control with the help of governance tools. This result contrasts with the above findings on wikis, which raises the question of what differences in terms of user interaction modes and collaborative behaviour may explain this discrepancy.

- **P5: User engagement and attention.** The marginal role of governance-related factors suggests that the main drivers of group dynamics in social media sharing systems need to be found elsewhere. In addition to social ties, individual and collective attentional spans may influence group growth. We saw that affiliation spread (*ms*) has a globally significant effect on group growth; analysing growth as a function of different values of affiliation spread (Figure 5.7, *right*) indicates that this effect is robust also at an individual basis: groups whose members tend to spread their contribution over many other groups are consistently slower in growth than groups whose members are more selective in their affiliations.

From a viability theory perspective, these results diverge from the relatively clear-cut distinction between descriptive factors and governance factors which we discussed in the case of wikis. Indeed, while it would not be hard to consider plain demographic metrics such as population and content size as characterising of the autonomous dynamics of a model of social media groups, typical governance measures (such as throttling or moderation), appear to have little influence on the evolution of these groups. In this regard the topological structure of group-centred social networks exerts a more important effect, however indirect. As noted above, although these factors can be seen as defining the autonomous dynamics of the system (by assuming that social ties are the result of spontaneous social interaction among users) they can also be regarded as genuine control factors (insofar as the design of the system or the implementation of specific policies can favour the creation of some kind of links, e.g. transitive links, or increase the cohesiveness of the members' social network).

3. A simple model of viable Web communities

In this section we propose to formulate a simple model in which the viability of collaborative communities could be assessed against constraints on group population and group content size. In other words, we sketch the initial steps needed to formally account for the intertwinement of these constraints, the autonomous demographic dynamics of such systems and the possible control actions which may be adopted to ensure and/or restore their viability. We leave issues specifically related to model solving and simulation-based approximation outside of this chapter. To this end, the interested reader may nonetheless apply the tools and techniques presented in Part. 2 of the present book.

3.1 Model variables

From the two examples discussed in the previous section we may abstract a list of variables that can be applied to describe the structure and governance of collaborative systems in general. Apart from the fundamental demographic metrics consisting of population and content sizes, such communities are generally characterised by the fact that some users (so-called "administrators") have special privileges; besides, they often feature similar governance mechanisms that limit the amount of contributions (filtering processes, registration requirements before contribution, etc.).

Model variables are to be based upon the corresponding quantities, at least for those features that can be described in quantitative terms. Among these variables, we have to distinguish *state variables* from *control variables* (upon which, for instance, administrators may act in order to influence the dynamics of the community they manage). Additionally, we must be able to define a viability domain, wherein we expect state variables should remain in order to consider the system viable.

State variables. The state space *per se* is essentially made of:

- U**:** denoting the population of members or participants in a community.
- P**:** representing the size of content contributed by members (pages, photos, etc.).

Control variables. Those include:

- A**:** the number of administrators in a community. For convenience, we also call *administrator ratio* the proportion of administrators with respect to the whole population, denoted as a; similarly, the *administrator density* is defined as the proportion of administrators with respect to content size, denoted as b. We then have: $aU = A$ and $bP = A$.
- m**:** moderation constraint. The moderation constraint represents a mechanism that filters contributions, such as: e.g. edition permission on wikis ($m = R$), or moderation queues in Flickr ($m = \mu$).

Eventually, the status of:

- c**:** cohesiveness (for instance measured through the clustering coefficient: $c = c_3$ in Flickr groups)

as a control variable remains relatively unclear: on the one hand, it may indeed be possible for administrators to favour some kinds of interactions between users, in such a way that the group becomes more cohesive. Yet, it is likely that the evolution of this variable could also be dictated by the autonomous dynamics of the group — for instance and all other things being equal, an increasing population is likely to induce a weaker cohesiveness.

Viability domain As discussed in the first section, the definition of a "desirable state" as a precondition to studying group viability is bound to have a large number of potential interpretations in social systems or, at least, be more debatable than in the case of e.g. physical systems. In the following preliminary model, we choose to adopt a rather simple approach to viability by stylising and extrapolating a plausible trend observed both in Flickr and wiki groups: that groups tend to roughly grow along a diagonal of constant ratio U/P. In particular, we notice that the largest groups (both in terms of content and population) are concentrated along this diagonal. We suggest that these groups should have had a successful development, at least at some point, in order to reach this area of the (U,P) state space.

Table 5.6: Model variables and parameters.

	name	*description*	*range*
Variables	U	group population	$\mathbb{R}^{+*}$
	P	group content size	$\mathbb{R}^{+*}$
	A	administrator population	$\mathbb{R}^{+*}$
	m	moderation constraint	$\{0,1\}$
	c	cohesiveness of contacts within the group	$[0,1]$
	r	proportion of reciprocated contacts	$[0,1]$
Parameters	λ_a	slope of growth vs. population size	$\mathbb{R}^{+*}$
	λ_b	slope of growth vs. content size	$\mathbb{R}^{+*}$
	ν_a	relative spread of administrator ratios (per user)	$[0,1]$
	ν_b	relative spread of administrator densities (per page)	$[0,1]$
	g_0, g_1	effect of resp. the absence or presence of moderation	$\mathbb{R}^{+*}$
	ρ^-, ρ^+	minimum and maximum population/content ratio	$\mathbb{R}^{+*}$

We thus define viable a group such that its U/P ratio remains within a given boundary:

$$\rho = [\rho^-, \rho^+]$$

3.2 Viable dynamics

The model aims at stylising several effects initially observed in the empirical data. In its present version however, it values mathematical tractability over realism. From a generic point of view, the system of differential equations governing the evolution of state variables could be written as:[13]

$$\begin{cases} \dfrac{dU_t}{dt} & = & \alpha(U_t, P_t, A_t, c_t, m_t, r_t) \\ \\ \dfrac{dP_t}{dt} & = & \beta(U_t, P_t, A_t, c_t, m_t, r_t) \end{cases} \tag{5.2}$$

From this point, various equations and diverse models can be proposed depending on different collaborative models one wants to describe, peer production systems (such as wikis) vs. social media groups (such as Flickr communities). We choose to focus on wikis and, in particular, their user dynamics, dU_t/dt. In the absence of empirical results based on social network properties in wikis we should ignore the role of cohesiveness in the present model. Eventually, we can write:

$$dU_t/dt = \alpha_{\mathrm{w}}(U_t, P_t, A_t, m_t)$$

In our previous analyses, we made the hypothesis that wiki growth over the *whole* observation period was directly influenced by a constant and permanent impact of the values of control and state variables measured at the *beginning* of the given period. The most straightforward way to account for this type of growth is to assume that groups experience an exponential growth depending on initial conditions: in other words, given U_{t_0}, P_{t_0}, A_{t_0} and m_{t_0}, growth is assumed to be an exponential function of t.

The following equation can therefore be proposed, for a proper function ϕ:

$$\frac{dU_t}{U_t} = \phi(U_{t_0}, P_{t_0}, A_{t_0}, m_{t_0})dt \quad \Leftrightarrow \quad U_t = U_{t_0} e^{(t-t_0)\phi(U_{t_0}, P_{t_0}, A_{t_0}, m_{t_0})} \tag{5.3a}$$

where ϕ could be estimated for instance from Roth et al., 2008 as a function depending (i) linearly upon U_{t_0}, (ii) affinely and in a monotonously decreasing manner upon A_{t_0}/P_{t_0} (or, rather, quantiles thereof) and (iii) upon a given step function of m_{t_0}. Put differently, ϕ can be schematically written as:

$$\begin{aligned} \phi(U_{t_0}, A_{t_0}, P_{t_0}, m_{t_0}) &= \lambda_a U_{t_0} \left(\frac{b_{t_0} - b_{\max}}{b_{\min} - b_{\max}} \right) g_{m_{t_0}} \\ &= \frac{\lambda_a}{\nu_b} U_{t_0} (1 - \frac{A_{t_0}}{b_{\max} P_{t_0}}) g_{m_{t_0}} \end{aligned} \tag{5.4}$$

where $\lambda_a \in \mathbb{R}^+$ is a given constant, $b_{\min}$ and $b_{\max}$ are respectively the minimum and maximum administrator densities over all groups, $\nu_b = \frac{b_{\max} - b_{\min}}{b_{\max}}$ is the constant relative spread of administrator ratios, and g_0 and g_1 are two given constants such that $g_0 < g_1$ (in practice, from Fig. 5.3 it roughly seems that $g_1/g_0 \in [2,3]$).

Viable group dynamics: a coevolutionary sketch. By extrapolating on this empirically-based relationship, we propose a dynamic model featuring the following system of equations:

$$\begin{cases} \dfrac{dU_t}{dt} = \dfrac{\lambda_a}{\nu_b} g_{m_t} (1 - \dfrac{A_t}{b_{\max} P_t}) {U_t}^2 \\ \\ \dfrac{dP_t}{dt} = \dfrac{\lambda_b}{\nu_b} g_{m_t} (P_t^2 - \dfrac{A_t P_t}{b_{\max}}) \end{cases} \tag{5.5}$$

while $\forall t \geq 0$, the following double inequality should hold:

$$\rho^- < U_t/P_t < \rho^+$$

Possible refinements, given specific cases and under specific subsets of assumptions, where for instance A_t itself could be connected to U_t and P_t, are left to the reader. The same goes for the viability domain: to account for the need for a group to keep receiving contributions, constraints should be put onto dP_t/dt — for instance, by requiring that it remains above a certain level of activity and below a certain threshold of cognitive effort for group members (so that they can reasonably keep track of ongoing contributions). Similarly, exact solutions of this class of models are left to the interest of the reader. Approximate and simulation-based exploitation of these kinds of dynamics can be exploited through software sucg as Kaviar, to allow performing computations of the corresponding viability kernels (see Chap. 10).

4. Conclusions

This chapter presented a methodology aimed at empirically appraising one possible dimension and understanding of "viability" (growth-related viability), in two paradigmatic cases of collaborative communities. In this respect, it constitutes a preliminary framework and a necessary step towards defining and modelling pattern resilience and viability in the context of social systems.

In particular, we assessed the interplay of demographic factors and governance structure and their role among forces driving the macro-level growth of content-based Web communities; in the case of social media communities, we emphasised the role of social network properties as a class of indirect control factors. Despite a largely shared ontology (governance features, demographic factors) between the two case studies and the application of a similar analysis taking growth as a proxy for viability, we noticed several macroscopic discrepancies: most importantly, we observed that demographic and governance properties are good predictors of growth in the case of wikis, but are surprisingly poor at predicting growth rates in social media communities. Some possible reasons for this discrepancy may depend on the specific methodology that we adopted here (e.g. unnoticeable effects on relative growth rates as opposed to absolute growth). The observed inconsistency in the effectiveness of control factors may be due to distinct underlying modes of peer production—i.e. genuinely *collaborative*, in the case of wikis (as users jointly modify shared content) vs. *atomistic* in the case of media sharing (as users contribute individual contents to a shared pool)—or distinct types of social interaction (suggesting the effects of the underlying social network as a stronger recruitment factor in social media groups).

All in all, we characterised the relationships between structural, demographic and governance-related variables of these online groups. We interpreted these relationships from the perspective of the evolution, stability and sustainability of these Web communities. We essentially showed how data can be analysed and how, in principle, the results of such analysis could be used as input for highly aggregated mathematical models: within the framework of viability theory, we argue that this approach makes it possible to define hypotheses pertaining to the *autonomous dynamics* of such systems and the factors that control them, and therefore serve as a first step towards the design and computation of the corresponding "capture basins" or regions of the variable space where viability can be ensured. More broadly, we suggest that it constitutes a preliminary framework and a necessary step towards defining and modelling pattern resilience and viability in the context of social systems.

Acknowledgments

This work was partly supported by the PATRES project (`NEST-043268`) funded by the FP6 programme of the European Commission and by the Future and Emerging Technologies programme FP7-COSI-ICT of the European Commission through project QLectives (grant no.: `231200`). We are grateful to Nigel Gilbert, Volker Grimm, Nic Geard, Przemek Grabowicz, as well as members of the Centre for Research in Social Simulation (University of Surrey) and participants in the 2008 Dagstuhl seminar "Social Web Communities" for valuable feedback and insights on earlier versions of this work.

Notes

1. We will use in what follows "online groups" and "Web communities" as synonyms to refer to social systems that require an explicit act of affiliation for users to become members, as opposed to communities that can be "detected" on the basis of topological properties of a network (see for example Edling, 2002; Newman, 2006), see Taraborelli and Roth, 2010 for a discussion.

2. The question of the mapping between real, offline identities and online identities in Web communities is beyond the scope of the present discussion. In our work we refer to a member of an online community as a user that can be identified by a unique online identifier, no matter who actually owns and controls that identifier.

3. For an example of qualitative, community driven attempts at understanding patterns that affect in a positive or negative way the evolution of a collaborative community, see Mader, 2007

4. For an extensive discussion of results presented in this section, see Roth et al., 2008

5. As per our considerations in the first section of this chapter, it should be noted that sheer content growth per se may not be a good indicator of a sustainable wiki, as studies on wiki proliferation also seem to suggest Happel and Treitz, 2008.

6. This initial dataset includes among others a large set of *Wikipedias*.

7. `http://s23.org/wikistats/largest_html.php`. The database is maintained by a user called "Mutante" who graciously granted us the permission to harvest this data.

8. `http://flickr.com/services/api`

9. For a comparison with the idea of a sociodemographic space, see McPherson et al., 1992.

10. `http://dev.nitens.org/flickr/group_trackr.php`

11. A similar rationale for focussing on mid-size groups is in Backstrom et al., 2006.

12. It should be noted that, compared to other studies that considered a random sample of the global Flickr user and group population, our study focused on *public groups* (i.e. groups with public content, flagged as "safe" and hence open to recruit any Flickr user as a potential member) and users that engage in actual social activity such as being member of at least one public group (Prieur et al., 2008 estimated this to represent the 8% of the total Flickr population in 2006). This explains the mismatch between the global statistics reported by other studies that include private groups and non-social users (i.e. users who may only use social media services as a way to dump private content not meant for public consumption).

13. For the sake of differentiability, we obviously have to make a continuous approximation on U, P and A by considering that they evolve in $\mathbb{R}$; one may still consider that the population at t equals $\lfloor U_t \rfloor$.

References

Almeida, R., Mozafari, B., and Cho, J. (2007). On the evolution of Wikipedia. In *Proc. Intl. Conf. Weblogs and Social Media (ICWSM '07)*.

Anthony, D., Smith, S., and Williamson, T. (2005). Explaining quality in internet collective goods: Zealots and good samaritans in the case of Wikipedia. URL http://web.mit.edu/iandeseminar/Papers/Fall2005/anthony.pdf

Backstrom, L., Huttenlocher, D., Kleinberg, J., and Lan, X. (2006). Group formation in large social networks: membership, growth, and evolution. In *Proc. 12th ACM SIGKDD Intl. Conf. Knowledge discovery and data mining*, pages 44–54, New York, NY, USA. ACM Press.

Backstrom, L., Kumar, R., Marlow, C., Novak, J., and Tomkins, A. (2008). Preferential behavior in online groups. In *Proc. Intl. Conf. Web search and web data mining (WSDM '08)*, pages 117–128, New York, NY, USA. ACM.

Benkler, Y. (2002). Coase's penguin, or Linux and the nature of the firm. *Yale Law Journal*, 112(3).

Benkler, Y. (2006). *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press.

Brandes, U. and Lerner, J. (2008). Visual analysis of controversy in user-generated encyclopedias star. *Information Visualization*, 7:34–48.

Bryant, S. L., Forte, A., and Bruckman, A. (2005). Becoming Wikipedian: Transformation of participation in a collaborative online encyclopedia. In *Group'05, Sanibel Island, FL, USA*.

Capocci, A., Servedio, V. D. P., Colaiori, F., Buriol, L. S., Donato, D., Leonardi, S., and Caldarelli, G. (2006). Preferential attachment in the growth of social networks: The internet encyclopedia Wikipedia. *Phys. Rev. E*, 74(3):036116.

Cha, M., Mislove, A., Adams, B., and Gummadi, K. P. (2008). Characterizing social cascades in Flickr. In *Proceedings of the 1st ACM SIGCOMM Workshop on Social Networks (WOSN'08)*, New York, NY, USA. ACM.

Ciffolilli, A. (2003). Phantom authority, self-selective recruitment and retention of members in virtual communities: The case of Wikipedia. *First Monday*, 8(12).

Edling, C. R. (2002). Mathematics in sociology. *Annual Review of Sociology*, 28:197–220.

Forte, A., Larco, V., and Bruckman, A. (2009). Decentralization in Wikipedia governance. *Journal of Management Information Systems*, 26(1):49–72.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, 438(7070): 900–901.

Godfrey, M. and Tu, Q. (2001). Growth, evolution, and structural change in open source software. In *Proc. 4th Intl. Workshop on Principles of Software Evolution (IWPSE '01)*, pages 103–106, New York, NY, USA. ACM.

Godfrey, M. W. and Tu, Q. (2000). Evolution in open source software: A case study. *ICSM*, 00:131.

Halim, F., Yongzheng, W., and Yap, R. (2009). Wiki credibility enhancement. In *WikiSym '09: Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, pages 1–4, New York, NY, USA. ACM.

Happel, H.-J. and Treitz, M. (2008). Proliferation in enterprise wikis. In *Proceedings of the 8th International Conference on the Design of Cooperative Systems (COOP 08)*, Carry-le-Rouet, France.

Hars, A. and Ou, S. (2002). Working for free? motivations for participating in open-source projects. *Int. J. Electron. Commerce*, 6(3):25–39.

Kittur, A., Chi, E., Pendleton, B. A., Suh, B., and Mytkowicz, T. (2007). Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. In *ALT CHI 2007*, San Jose, CA.

Kittur, A., Pendleton, B., and Kraut, R. E. (2009). Herding the cats: The influence of groups in coordinating peer production. In *WikiSym '09: Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, pages 1–9, New York, NY, USA. ACM.

Lam, S. K. and Riedl, J. (2009). Is Wikipedia growing a longer tail? In *GROUP '09: Proceedings of the ACM 2009 international conference on Supporting group work*, pages 105–114, New York, NY, USA. ACM.

Lerman, K. and Jones, L. (2007). Social browsing on Flickr. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM '07)*.

Leskovec, J., Backstrom, L., Kumar, R., and Tomkins, A. (2008). Microscopic evolution of social networks. In *Proc. 14th ACM SIGKDD Intl. Conf. on Knowledge discovery and data mining (KDD '08)*, pages 462–470, New York, NY, USA. ACM.

Levrel, J. (2006). Wikipédia, un dispositif médiatique de publics participants. *Réseaux*, 24(138):185–218.

Lih, A. (2004). Wikipedia as participatory journalism: Reliable sources? Metrics for evaluating collaborative media as a news resource. In *5th Intl Symp on Online Journalism*, Austin, TX, USA.

Mader, S. (2007). *WikiPatterns. A Practical Guide To Improving Productivity and Collaboration In Your Organization*. Wiley.

Marlow, C., Naaman, M., Boyd, D., and Davis, M. (2006). HT06, tagging paper, taxonomy, Flickr, academic article, to read. In *HYPERTEXT '06: Pro-*

ceedings of the seventeenth conference on Hypertext and hypermedia, pages 31–40, New York, NY, USA. ACM Press.

McPherson, J. M., Popielarz, P. A., and Drobnic, S. (1992). Social networks and organizational dynamics. *American Sociological Review*, 57(2):153–170.

McPherson, M. (1983). An ecology of affiliation. *American Sociological Review*, 48(4):519–532.

Mislove, A., Koppula, H. S., Gummadi, K. P., Druschel, P., and Bhattacharjee, B. (2008). Growth of the Flickr social network. In *WOSP '08: Proc. 1st workshop on Online social networks*, pages 25–30, New York, NY, USA. ACM.

Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., and Bhattacharjee, B. (2007). Measurement and analysis of online social networks. In *IMC '07: Proc. 7th ACM SIGCOMM conf. on Internet measurement*, pages 29–42, New York, NY, USA. ACM.

Negoescu, R. A. and Perez, D. G. (2008). Analyzing Flickr groups. In *CIVR '08: Proc. 2008 Intl Conf. Content-based image and video retrieval*, pages 417–426, New York, NY, USA. ACM.

Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582.

Nonnecke, B., Andrews, D., and Preece, J. (2006). Non-public and public online community participation: Needs, attitudes and behavior. *Electronic Commerce Research*, 6(1):7–20.

Nov, O., Naaman, M., and Ye, C. (2008). What drives content tagging: the case of photos on Flickr. In *CHI '08: Proceeding of the Twenty-Sixth annual SIGCHI Conference on Human Factors in Computing Systems*, pages 1097–1100, New York, NY, USA. ACM.

Plangprasopchok, A. and Lerman, K. (2009). Constructing folksonomies from user-specified relations on flickr. In *WWW '09: Proc. 18th Intl. Conf. on World Wide Web*, pages 781–790, New York, NY, USA. ACM.

Prieur, C., Cardon, D., Beuscart, J.-S., Pissard, N., and Pons, P. (2008). The strength of weak cooperation: A case study on Flickr. arXiv:0802.2317v1

Reagle, J. M. (2007). Do as I do: leadership in the Wikipedia. In *Proceedings of WikiSym '07, 3rd Intl Symposium on Wikis*, New York, NY, USA. ACM Press.

Roth, C. (2007). Viable wikis: struggle for life in the wikisphere. In *WikiSym '07: Proceedings of the 2007 international symposium on Wikis*, pages 119–124, New York, NY, USA. ACM.

Roth, C., Taraborelli, D., and Gilbert, N. (2008). Measuring wiki viability. An empirical assessment of the social dynamics of a large sample of wikis. In *WikiSym '08: Proc. 4th Intl. Symposium on Wikis*, New York, NY, USA. ACM.

Schifanella, R., Barrat, A., Cattuto, C., Markines, B., and Menczer, F. (2010). Folks in folksonomies: Social link prediction from shared metadata. In *Proc. 3rd ACM Intl. Conf. on Web Search and Data Mining (WSDM'10)*, New York City, New York, USA. ACM.

Sigurbjörnsson, B. and van Zwol, R. (2008). Flickr tag recommendation based on collective knowledge. In *Proc. 17th Intl. Conf. on World Wide Web (WWW '08)*, pages 327–336, New York, NY, USA. ACM.

Stvilia, B., Twidale, M. B., Smith, L. C., and Gasser, L. (2008). Information quality work organization in Wikipedia. *Journal of the American Society for Information Science and Technology*, 59(6):983–1001.

Suh, B., Chi, E. H., Kittur, A., and Pendleton, B. A. (2008). Lifting the veil: improving accountability and social transparency in wikipedia with wikidashboard. In *Proc. 26th SIGCHI conference on Human factors in computing systems (CHI '08)*, pages 1037–1040, New York, NY, USA. ACM.

Taraborelli, D. and Roth, C. (2010). Circles and ties. Drivers of group dynamics in social media. *Submitted*.

Valafar, M., Rejaie, R., and Willinger, W. (2009). Beyond friendship graphs: a study of user interactions in Flickr. In *Proc. 12th ACM SIGKDD on Online social networks (WOSN '09)*, pages 25–30, New York, NY, USA. ACM.

van Zwol, R. (2007). Flickr: Who is looking? In *IEEE/WIC/ACM International Conference on Web Intelligence*, pages 184–190.

Viegas, F., Wattenberg, M., Kriss, J., and van Ham, F. (2007). Talk before you type: Coordination in Wikipedia. In *Proc. 40th Hawaii Intl Conf on System Sciences*.

Voss, J. (2005). Measuring Wikipedia. In *Proc. Intl. Conf. International Society for Scientometrics and Informetrics (ISSI '05)*, Stockholm.

Wilkinson, D. and Huberman, B. (2007). Assessing the value of cooperation in Wikipedia. *First Monday*, 12(4).

Wöhner, T. and Peters, R. (2009). Assessing the quality of Wikipedia articles with lifecycle based metrics. In *WikiSym '09: Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, pages 1–10, New York, NY, USA. ACM.

Zheleva, E., Sharara, H., and Getoor, L. (2009). Co-evolution of social and affiliation networks. In *KDD '09: Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1007–1016, New York, NY, USA. ACM.

Zlatic, V., Bozicevic, M., Stefancic, H., and Domazet, M. (2006). Wikipedias: Collaborative web-based encyclopedias as complex networks. *Phys. Rev. E*, 74(1):016115.