Circles and ties: The drivers of group
dynamics in social media

Dario Taraborelli*  Camille Roth†
d.taraborelli@surrey.ac.uk roth@ehess.fr

Abstract
Social media represent a complex ecosystem in which users can form or join communities as well as establish links with other users for the purpose of socialising and sharing content. Compared to “pure” social networking systems, media sharing platforms allow a broader range of social interactions that are mediated by content production. The dynamics of group formation and group evolution in social media hence depend on the interplay of multiple factors, spanning relationships between users and groups, users and content, users and other users. Governance measures that control the mode of participation to a group also affect how these communities evolve over time. In this study we present an analysis of the evolution of a large set of groups from Flickr, a popular media sharing Web service. We identify the influence of a number of properties (such as social network structure, demographic profile and governance structure) on the macroscopic growth in content and population of these communities, discussing which properties can be regarded as drivers or regulators of group growth. We discuss in particular the weak effects of governance measures in controlling growth compared to structural properties of a group’s social network.

1 Introduction
Social networking and social media services have been thriving since the advent of the so-called Web 2.0 and have started attracting a considerable attention in academic research in recent years. This is partly due to the explosion of interest that such services have generated in popular culture and in the media, and partly (and most importantly) to the massive availability of data on the social behaviour of Web users that these services can provide. Social scientists have extensively started to use data available via such services as a way to empirically validate hypotheses on social networks and their evolution over time (see for example Adamic and Glance, 2005; Ali-Hasan and Adamic, 2007; Golbeck, 2007; Kossinets and Watts, 2006; Kumar et al., 2006; Leskovec et al., 2008; Mislove et al., 2008, 2007).

The social Web, however, affords much more than an infrastructure for the creation of links among individuals. The Web 2.0 has become an extraordinary vehicle for the support of collaborative online communities and peer production systems (Benkler, 2006). The specificity of services supporting online communities, as opposed to “pure” social networking services, is that they provide an infrastructure that allows users not only to create new social links but also to share or contribute content, whether in the form of collaborative content production (such as in wikis or open source communities), content sharing (such as in media sharing services), content annotation (such as in

*Centre for Research in Social Simulation, University of Surrey, Guildford GU2 7XH, United Kingdom
†Centre d’analyse et de mathématique sociales (CNRS-EHESS) 54, boulevard Raspail, 75006 Paris, France
social bookmarking websites) or content-driven discussion (as in discussion forums or review-based services).

The possibility of studying social interaction mediated by content raises new challenges for social science. If, broadly speaking, the dynamics of a social network can be accounted for in terms of the creation and removal of links between individual agents, the dynamics of a content-based collaborative community can only be described by virtue of multiple kinds of entities and relations. These typically include: (a) links between agents and content (such as those resulting from the act of producing, removing, modifying, sharing and annotating content), (b) group affiliation links (such as joining or leaving a group) and (c) norms that regulate social interaction and content sharing within these communities. In spite of an increasing interest for online communities as a data source for Web-based social interaction, relatively little effort has been put in the analysis of processes at play in the evolution of content-based online communities as opposed to the evolution of pure social networks. In the present work our aim is to start bridging this gap, by focussing in particular on the relation between community dynamics and social network structure. Before delving into the question of what drives the growth of content-based online communities, we need to introduce a few caveats and conceptual distinctions.

### 1.1 Collaborative communities

When using the term “communities” we should distinguish between the notion of a community such as the one that is typically used in the field of social network analysis and the full-fledged notion of a group or collaborative community such as the one that we will be referring to in the present work.

Social network research has defined a number of formal (mostly topological) features that allow the characterisation of particular structures in a network as “communities”, “modules” or “cohesive groups” (see, *inter alia*, Alba, 1973; Freeman, 2003; Girvan and Newman, 2002; Newman, 2006; Wasserman and Faust, 1994; White et al., 1976). Such communities can be detected by considering the formal properties of a network of agents and can be used to make predictions on homophily and similarity in behaviour of specific sets of agents that fall within the scope of such communities (e.g. authors massively citing each other, users choosing similar tags to annotate content, or more generally agents that are implicitly related with each other with respect to a range of common interests).

In contrast, when using the term of online communities in the present study, we refer to social groups that obey to stronger constraints than those described by social network research. Web communities require users to establish an explicit link of affiliation to the group and they typically include norms on participation and social behaviour that each member should comply with. We follow, in this sense, the simple definition suggested by Preece et al. (2003), who refer to an online community as “a group of people who interact in a virtual environment, [...] have a purpose, are supported by technology, and are guided by norms and policies”.

Most contemporary online networking services allow the creation of this kind of communities. Communities in online networking services may form, in particular, for reasons that are extrinsic to content (and where content sharing is only indirectly supporting interactions among members of the group) whereas other communities can be said to be content-based in a strong sense (as is the case of groups of interest where social interaction is primarily driven by collaborative content creation). Web communities as defined in this stronger sense allow us to distinguish, in particular, three different classes of links that characterise their structure: user-to-user links (defining what we may call a user-centred social network), user-to-group links (defining the group affiliation network) and member-to-member links (defining a group-centred social network or the network of links between users that are affiliated with the same community). Users that have individual links with members of a group but are not themselves members of this group can be said to represent the fringe of a group-centred social
network (Backstrom et al., 2006). The distinction between these three kind of links is illustrated in Figure 1, that exemplifies the notion of a group-centred social network.

![Figure 1: (left) Social network and group membership: solid arrows represent directed user-to-user links, dashed arrows represent user-to-group affiliation links; highlighted nodes represent users with at least one group affiliation. (right) Group-centred social network: highlighted nodes and links represent the subset of the global social network internal to group G2. Links between members of the same group define what we call a group-centred social network.](image)

On the basis of the above distinctions, we can now formulate a series of research questions that call for empirical investigation.

1. How does the structure of group-centred social networks affect the overall growth of groups in content and population?

2. How do user-to-user links and group membership affect patterns of communication activity and social interaction between Web users?

3. What is the role of a group fringe in explaining a group growth potential or its ability to recruit new members?

4. What is the respective contribution of the user-centred social network, the group-centred social network and the group fringe in the dissemination of content across Web communities?

Previous research addressed issues such as group-based communication activity (Schoberth et al., 2003), group recruitment (Backstrom et al., 2006; Mislove et al., 2008) and social network-based content dissemination (Cha et al., 2008; Valafar et al., 2009). The present work aims at tackling the first of the above questions by looking at factors driving macroscopic group dynamics in a popular media sharing service.

1.2 Flickr as a case study

_Flickr.com_, one of the most popular photo and video sharing services, represents an ideal case for the study of the joint effects of content-based interaction, group affiliation and social network dynamics. Its user model allows the creation of (user-to-user) social links that can provide a direct insight into user-centred 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 “favourite”), hence offering the opportunity to study social behaviour mediated by _user-to-content_ links; finally, by supporting with
a dedicated infrastructure the creation of communities of interest or “groups”, it represents an ideal
testbed for studying user-to-group affiliation links and their effects on social interaction among users.
Thanks to a rich and extensively documented API\(^1\), Flickr allows the extraction of large datasets that
can be used to study social dynamics at each of these levels of description (content, users, groups).

2 Related work

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, 2008; Sigurbjörnsson and van Zwol, 2008). A smaller number of
works, more relevant to the present analysis, focussed on social interaction and group-driven social
behaviour.

2.1 Social network research on Flickr

Beside studies attempting to model “pure” social network dynamics in Flickr (such as Leskovec
et al., 2008; Mislove et al., 2008), the effects of social networks on content popularity and content
dissemination are one of the most prominent areas of research for which Flickr has been studied.
Lerman and Jones (2007) analysed patterns of browsing behaviour, showing how contact links form
the social backbone of content-sharing services such as Flickr and that a considerable amount of
content-related activity (such as viewing a photo, commenting a photo, marking a photo as a favourite)
is indeed mediated by user-to-user networks of “friends” (or “contacts” as they are referred to in
Flickr).

A similar conclusion is reached by Cha et al. (2008), who investigated the role of social cascades
in Flickr, looking at how user-centred social networks mediate the rapid spreading of popularity in
content. Their findings support the idea that (i) online social networks are extremely efficient at
spreading content at a very rapid rate (in particular, they are considerably more efficient at spreading
content than social networks at spreading infective diseases) and (ii) social network structure can be
used to reliably predict patterns of popularity of content. Studies analysing factors affecting content
popularity dynamics in Flickr (e.g. the temporal profile of the number of views a photo receives)
reached similar conclusions, confirming that user-centred social networks are the most prominent
vehicle of content dissemination among Flickr users (van Zwol, 2007). These results have been partly
challenged by Valafar et al. (2009) who showed that a negligible portion of actual user interaction
involving photos is mediated by the individual social network of a user: the majority of user activity
on a single user’s content is not entirely driven by the network of its friends, which raises the question
of what other channels drive user interaction and content discovery on top of user-to-user social ties.
Groups, we propose, are one such channel.

2.2 Research on 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
happen via groups.\(^2\) Flickr groups (as many other communities of interest that thrive on online net-

\(^1\)http://flickr.com/services/api

\(^2\)There is disagreement on the rate of group participation with respect to the whole Flickr user base. Mislove et al.
(2007) mention that the fraction of users that use group features based on their sample is 21%. Prieur et al. (2008) suggest
that the number of users that are members of at least one group is approximately 8% (but up to 49% if considering only
users with paying accounts). Negoescu and Perez (2008) note that 50.9% of the users in their sample shared at least one
working services) are of particular interest to the present analysis because, as opposed to purely user-centred social networks, they can be described as communities of interest driven by shared content. Research on content dissemination mentioned in the previous section focused on aspects that are social, in that they are mediated by social ties of a given kind, although not strictly speaking collaborative. Flickr groups, on the contrary, 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. In most cases (public groups), being member of a group is a necessary condition to be able to share content.\footnote{This may not be true any more, since Flickr introduced photo-specific invitation links that allow group administrators to request a non-member to contribute a picture to the group pool without the requirement of joining the group.} Private groups further restrict participation by requesting that users join the group in order to be able to see the content of a group. Groups can also be by invitation, so that users can only upload content if they are explicitly invited by other group members. 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 collaborative 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) conducted a general analysis of Flickr groups in the context of a comparison of high-level properties of several social networking services. In spite of large discrepancies in group use across these services, the same global trends were identified, showing in particular that (i) groups represent communities of users characterised by highly dense networks (as opposed to users with lower than average group participation), that (ii) members of small user groups tend to be more clustered than those of larger groups and that (iii) the most sociable users (those with the highest user-to-user link outdegree) tend to be members of a larger number of groups.

Prieur et al. (2008) focused on the relation between group topicality (as the dispersion of tags used to describe pictures in the group’s pool) and social density of groups and found a variety of group types along these two dimensions. This variety makes it possible, for instance, to use social density to tell apart geographic groups with occasional contributions by tourists (highly topical groups with low social density) by equally topical groups by residents (with higher social density). Groups with high social density may also affect the evolution of tag dispersion by inducing the use of more similar tags than groups with looser social ties, although no direct evidence is provided in support of this intriguing idea. Social and topical alignment within groups has been further investigated by Schifanella et al. (2010), who devised a method to independently assess topical alignment and the creation of social links, thereby showing that similarity in user interests (or homophily) is indeed independently driving the creation of new social links.

A further extensive analysis on group participation based on a static snapshot of Flickr groups was conducted by Negoescu and Perez (2008). This study presents in particular detailed statistics on the loyalty of group contributions by the same users, indicating that users tend to systematically share a limited amount of photos with the same, limited number of groups. However a high variability in user behaviour suggests that users sharing large sets of photos per group tend do so in only a few groups, and conversely users who are more selective about what photos to share are likely to contribute them to a higher number of groups. Determining content sharing behaviour on the basis of the spread of group affiliation has been the focus of a study by Backstrom et al. (2008), who showed the extent to which engagement is affected by the size of the group and the number of other group affiliations of a photo with at least one group. None of these studies seems to consider in these estimates the potentially large number of users contributing to private groups only.
De Choudhury (2009) found that future activity within a group (i.e. number of shared items, number of favourites added, number of comments posted, number of contact links created) as well as the potential of the group to recruit new members can be accurately predicted on the basis of a composite measure derived from observed patterns of past activity of its members. The results suggest that there is indeed a high correlation between different measures of in-group “activity”, so that the higher the overall rate of observed activity in a group is, the higher we should expect the rate of separate types of activity within that group to be. In this sense, the model can account for future activity based on observed recent activity. However it is unable to tell apart effects on group activity that can be directly ascribed to group affiliation (e.g. the ability to share photos) from effects that can be ascribed to other, group-independent factors such as social network properties or similarity in user interests: for example, a high rate of reciprocal comments between members of a group may be actually due to group-independent social connections between members rather than group co-membership. The model in other words may well provide an answer to the question of how existing members are likely to participate to a group based on their past rate of activity, but leaves open the question of understanding what drives new users to join a group, cease to actively participate in a group or decide to leave a group.

Zheleva et al. (2009) present a model that aims to account for the coevolution of the social and affiliation network in Flickr groups. The analysis used to empirically support the model shows that a large portion of group members are actually singletons, thus suggesting that group formation processes can only be partly ascribed to “social recruitment” processes driven by the social network of group members. Their findings are at odds with results by Backstrom et al. (2006) and Hui and Buchegger (2009) who identified in social network properties the main drivers of group formation, showing in particular the role of friends within a group to recruit new members from the group’s “fringe” (i.e. users who are closely connected with group members but not yet affiliated with the group) and the influence of the topological structure of the group-centred social network on these recruitment processes.

All in all these various results suggest that group formation processes in collaborative communities result 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.

### 3 Variety of group dynamics

The initial rationale for our study was the striking variety in growth patterns of Flickr groups when observed over time. One may expect that the vast majority of groups in media sharing services display an uninteresting monotonic growth tendency. However, upon closer inspection, groups display a surprising variety of temporal dynamics, that become evident 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. In many cases turnover is not sufficient to secure a steady growth in the group population, suggesting that leaving members may actually outnumber new recruits.

This variety in group dynamics is hardly surprising, if we think that Flickr groups—as full-fledged social communities—can serve a large range of social activities, spanning technical discussions or political campaigns (which typically induce high user recruitment, a high rate of discussion but a relatively little growth in uploaded content) as well as generic thematically or geographically-driven content sharing (where recruitment is proportional to growth in content and results in little discussion activity in the group forum).

One possibility to come to grips with this variety in global behaviour might then be to ask whether groups can be broadly categorised in distinct typologies, determined by similarity in individual usage patterns of their members and topicality. Such “group types” could then be identified in a qualitative way by considering 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.

Figure 2: Examples of group dynamics: Relative daily variations (in percentage) in content (thin purple line) and population (thick blue line) in 9 public Flickr groups over one year. Groups (5), (7) and (8) display an overall negative content variation, while groups (8) and (9) display a negative variation in population.
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.\textsuperscript{4} 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. 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. In the following sections we propose a framework to identify types of groups that display similar dynamics from the analysis of a large dataset of Flickr groups.

### 4 Method

#### 4.1 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\textsuperscript{5}, a public Web service allowing 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: (i) user-to-group affiliation links (ii) user-to-user contact links.

\textsuperscript{4}For a comparison with the idea of an agent-level sociodemographic space, see McPherson et al. (1992).

\textsuperscript{5}http://dev.nitens.org/flickr/group_trackr.php
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. To capture the natural dynamics of these groups we also introduced a capping on the maximum daily growth rate in content and population, excluding those groups displaying an instantaneous growth of more than 5% of their pool size or population (which we assumed could only result from extrinsic events such as contests or other forms of group promotion resulting in short-term mass-recruitment). 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 for the present study consists of 9,167 groups.

4.2 Variables

4.2.1 Group metrics

The metrics that we used as independent variables to study the drivers of group dynamics throughout the present study are described in table 1.

Table 1: Group metrics used as independent variables

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographic</strong></td>
<td></td>
</tr>
<tr>
<td>$u_0$</td>
<td>number of group members</td>
</tr>
<tr>
<td>$p_0$</td>
<td>number of photos in the group pool</td>
</tr>
<tr>
<td>$ms$</td>
<td>average membership spread of group members</td>
</tr>
<tr>
<td><strong>Structural</strong></td>
<td></td>
</tr>
<tr>
<td>$k$</td>
<td>average directed degree of group members</td>
</tr>
<tr>
<td>$c_3$</td>
<td>average clustering coefficient</td>
</tr>
<tr>
<td>$r$</td>
<td>reciprocity index</td>
</tr>
<tr>
<td><strong>Governance</strong></td>
<td></td>
</tr>
<tr>
<td>$adm$</td>
<td>number of group administrators</td>
</tr>
<tr>
<td>$mod$</td>
<td>number of group moderators</td>
</tr>
<tr>
<td>$\mu$</td>
<td>moderation filter</td>
</tr>
<tr>
<td>$\theta$</td>
<td>throttling index</td>
</tr>
</tbody>
</table>

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 topological 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; $\mu$ 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; $\theta$ is a quantitative indicator of the maximum number of photos that can be contributed to the group per time period (day, week or month).

A similar rationale for focussing on mid-size population groups is proposed by Backstrom et al. (2006).
4.2.2 Group growth indicators

The growth of a group over the observation period can be characterised 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 left a group (\( u_- \)) over the observation period. The turnover itself can be considered in absolute (\( u_+ - u_- \)) or relative terms \( \frac{u_+ - u_-}{u_0} \). One should also consider that growth and turnover (whether defined in absolute or relative terms) can be positive or negative, the latter being the case when the number of members leaving a group is higher than the number of newly recruited members.

Table 2: Group growth indicators

<table>
<thead>
<tr>
<th>Group size variation</th>
<th>( \Delta u ) absolute population variation ((u_1 - u_0))</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Delta p ) absolute content variation ((p_1 - p_0))</td>
<td></td>
</tr>
<tr>
<td>Population turnover</td>
<td>( u_+ ) number of users who joined the group</td>
</tr>
<tr>
<td></td>
<td>( u_- ) number of users who left the group</td>
</tr>
</tbody>
</table>

For the sake of the present study we decided to focus on absolute rather than relative growth indicators (see table 2) for a number of reasons. First of all, we took all groups at face value as equally likely to recruit new members and measure size-dependent effects as only one among several possible assumptions on growth driving factors. Although other studies have showed that the size of a group population plays a central role in the recruitment of new members (see 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.

5 Results and discussion

5.1 Global sample properties

The dataset 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 pool sizes and group populations in our dataset follow a log-normal distribution (best fit: \( \mu_P = 8.11, \sigma_P = 1.83; \mu_U = 6.30, \sigma_U = 1.42 \). See figure 3, left). 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. 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).

If we consider the absolute growth over the observation period, we notice that, although the vast majority of groups in our sample display a positive growth, a remarkable number of groups show an overall negative growth in population \(N = 1,471\) and photos \(N = 667\), suggesting that several groups suffer member drop-off and content removal, respectively.

Figure 3: Global sample properties. (left) Group populations \(u_0\), blue) and group pools \(p_0\), purple) follow a log-normal distribution. Note that the histogram of populations is truncated, as only groups with a minimum of 100 members were considered (see section 4.1). (right) Distribution of the absolute growth in population \(\Delta u\) and content \(\Delta p\) over the observation period \(t_1 - t_0\). Intervals highlighted in grey indicate groups with null absolute growth.

Table 3: Number of groups with positive, null and negative absolute growth

<table>
<thead>
<tr>
<th></th>
<th>(\Delta &gt; 0)</th>
<th>(\Delta = 0)</th>
<th>(\Delta &lt; 0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population ((\Delta u))</td>
<td>7,122 (77.7%)</td>
<td>574 (6.3%)</td>
<td>1,471 (16.0%)</td>
</tr>
<tr>
<td>Content ((\Delta p))</td>
<td>8,198 (89.4%)</td>
<td>302 (3.2%)</td>
<td>667 (7.2%)</td>
</tr>
</tbody>
</table>

5.2 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 \(\Delta u\), content variation \(\Delta p\) 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 Table 4 for the detailed list of variables included in each model). The general regression equation underlying each model (barring specific variable exclusions) is as follows:

\[
\log(y) = \lambda_0 + \lambda_{u_0} \log(u_0) + \lambda_{p_0} \log(p_0) + \lambda_r(r) + \lambda_{c_3}(c_3) + \lambda_k \log(1 + k) + \lambda_{ms} \log(ms) + \lambda_{\mu} \mu + \lambda_{mod} \log(1 + mod) + \lambda_{\theta} \log(\theta) + \lambda_{adm/u_0} \log(1 + (adm/u_0))
\]

\(^7\)Prieur et al. (2008) estimated this to represent the 8% of the total Flickr population in 2006.
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. 1. Variables corresponding to non-significant \(p\)-values were then iteratively excluded, generally resulting in a change in \(R^2\) smaller than 1%.

Table 4: Results of regression analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Population Variation ((\Delta u))</th>
<th>Content Variation ((\Delta p))</th>
<th>Population Turnover joining ((u_+))</th>
<th>Population Turnover leaving ((u_-))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\lambda_{u_0})</td>
<td>0.87 ***</td>
<td>0.77 ***</td>
<td>0.78 ***</td>
<td></td>
</tr>
<tr>
<td>(\lambda_{p_0})</td>
<td>0.94 ***</td>
<td>0.11 ***</td>
<td>-0.03 ***</td>
<td></td>
</tr>
<tr>
<td>(\lambda_r)</td>
<td>0.99 ***</td>
<td>2.09 ***</td>
<td>-0.19 **</td>
<td></td>
</tr>
<tr>
<td>(\lambda_{c_3})</td>
<td>-1.87 ***</td>
<td>-1.73 ***</td>
<td>-1.49 ***</td>
<td>-1.27 ***</td>
</tr>
<tr>
<td>(\lambda_k)</td>
<td>0.10 *</td>
<td>0.18 ***</td>
<td>0.23 ***</td>
<td></td>
</tr>
<tr>
<td>(\lambda_{ms})</td>
<td>-0.57 ***</td>
<td>-0.33 ***</td>
<td>-0.43 ***</td>
<td>0.35 ***</td>
</tr>
<tr>
<td>(\lambda_{\mu})</td>
<td>0.05 **</td>
<td>0.09 ***</td>
<td>0.08 **</td>
<td>0.07 ***</td>
</tr>
<tr>
<td>(\lambda_{mod})</td>
<td>0.09 ***</td>
<td>0.08 ***</td>
<td>0.02 ***</td>
<td></td>
</tr>
<tr>
<td>(\lambda_{\theta})</td>
<td>0.10 ***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\lambda_{adm/u_0})</td>
<td></td>
<td></td>
<td>-0.06 ***</td>
<td></td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.65</td>
<td>0.75</td>
<td>0.68</td>
<td>0.82</td>
</tr>
</tbody>
</table>

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

The results of the regression analysis summarised in table 4 indicate some salient effects of various initial properties of groups on their dynamics. 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: this is not only consistent with the overall distribution of group pool sizes and group populations but also with previous findings on growth drivers in wiki-based communities (Roth et al., 2008a,b) that showed that the effect of each variable on population growth is often linearly correlated with its effect on content growth. 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-centred 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 ($\mu, \text{mod}, \theta, \text{adm}/u_0$) 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.

5.3 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 presented in section 3. 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. We followed to this purpose a methodology proposed for a study of growth landscapes in a large sample of wikis in Roth et al. (2008b). 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 ($\Delta u$ and $\Delta p$ 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 4) 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 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 as a result of preferential attachment (as further exemplified by our analyses of social network effects) as well as herding behaviour.

P2: Effects of cohesiveness on group growth

Figure 5 (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-centred network works as a growth-regulating factor. Groups where cohesiveness is high display a higher inertia The non-monotonic trend in the effects of clustering may actually suggest that there is a critical threshold of cohesiveness (corresponding to the 3rd quantile) below which a high cohesiveness actually boosts growth. Once this threshold is passed, a saturation of social links produces an overall slower growth (quantiles 4-9).
P3: Are sociable users growth attractors?

An effect conflicting with cohesiveness is related to individual sociability as measured by the average degree of members in the group-centred contact social network (Figure 5, right). It should be noted that as degree has only been measured within groups, this analysis only informs us on its effect relative to the group-centred network. In this study we did not explore a related hypothesis, assessing the effects of sociability as measured by the degree calculated over the complete network of the Flickr population. This 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 other members than groups where the degree is more uniformly distributed.

P4: The poor effects of governance

The most striking findings of the present study are the negligible effects of the moderation and governance structure on group growth. Figure 6 (left) exemplifies the virtually flat growth landscape that emerges as a function of $\theta$, a throttling index measuring how many contributions users are allowed to make to a group within a given time period. With the only exception of low values which have a significant differential effect (quantiles 1 and 2), higher values of throttling display no significant difference in how they affect growth (quantiles 3-7). A possible explanation of this result is that looser throttling constraints may have no substantial consequence as the limit that they impose on the maximum number of contributions per user within a group is actually higher than the natural average.
contribution of group members. As self-regulating communities, one may speculate, social media groups are more likely to organise themselves as the result of the spontaneous behaviour of users than in response to governance measures. All in all, at a global level governance factors tend to show a consistently flat growth landscape and a very weak overall effect on population and content dynamics (as measured in the regression analyses). 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 earlier findings that show that governance and moderation constraints have a strong effect in the control of social and content dynamics of peer production systems such as wiki-based communities (Roth et al., 2008a,b), which raises the question of what differences in terms of user interaction modes and collaborative behaviour may explain this discrepancy.

Figure 6: (left) P4: Poor effects of governance, exemplified by throttling ($\theta$). (right) P5: Level of user engagement exemplified by affiliation spread ($ms$).

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. We suggested that social ties are at the same time key drivers of recruitment and likely causes of inertia in membership. However another important factor that affects the evolution of these systems is the way in which they determine the (individual and collective) attentional span. The question of how social computing systems shape and are shaped by collective attention has attracted an increasing interest over the last years (see Huberman et al., 2007; Wu and Huberman, 2007). Social media sharing systems tap into individual attentional limits because of the sheer volume of content, annotations and opportunities for social link creation that prompt decisions by individual users. Properties that are likely to strain individual attentional limits (for instance, an overwhelming volume of content in a group) are likely to have an effect on the social dynamics of the groups. Backstrom et al. (2006), for one, observed that members of smaller groups typically display a higher level of engagement than members of larger groups. We saw that affiliation spread ($ms$) has a globally significant effect on group growth; the analysis of growth as a function of different values of affiliation spread (Figure 6, 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.
6 Conclusions

The interplay of affiliation networks, social networks and content sharing patterns represents a promising new area of investigation for social scientists that may benefit from a range of data sources available from online networking services. The specific nature of interactions afforded by collaborative communities (i.e. content-dependent agent-to-agent interactions as well as multiple affiliation links) calls for the development of new models to account for the evolution of these communities, beyond those traditionally developed for the explanation of social network dynamics. This is a particularly important challenge for research attempting to explain processes behind collaborative knowledge production, collective decision making as well as the impact of governance measures on self-organising peer production systems.

We presented in this study an empirically assessment of demographic factors, social network properties and governance structures that jointly drive the macro-level growth of collaborative communities. Our analysis did not encompass a number of properties that are likely to play a central role in controlling growth, such as membership duration (Backstrom et al., 2008; McPherson et al., 1992), group-driven communication activity (as opposed to direct user-to-user communication) (Schoberth et al., 2003) or the distribution of activity and passive “lurking” within groups (De Choudhury, 2009; Nonnecke et al., 2006). Future research should clarify how each of these variable affects the temporal evolution of collaborative communities. Another complementary issue that we expect future research to focus on is the problem of individual drivers of affiliation behaviour, i.e. what properties of a user’s affiliation and social networks are likely to increase or decrease her probability to join or leave a group. Understanding the factors underpinning multiple affiliation in social media and peer production systems may help shed light on how to effectively govern these communities and predict their long-term evolution.

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, Nic Geard, Przemek Grabowicz, the members of the Centre for Research in Social Simulation (University of Surrey) and the participants in the 2008 Dagstuhl seminar “Social Web Communities” for valuable feedback on earlier versions of this work.

References


