When groups blend with social networks
An analysis of community dynamics in Flickr

Dario Taraborelli
CRESS
University of Surrey
Guildford GU2 7XH
United Kindgom
d.taraborelli@surrey.ac.uk

Camille Roth
CAMS
CNRS-EHESS
54, boulevard Raspail
75006 Paris, France
roth@ehess.fr

Andrea Baldassarri
Dipartimento di Fisica
"Sapienza" Università di Roma
Piazzalele A. Moro 2
00185 Roma, Italy
andrea.baldassarri@roma1.infn.it*

Abstract

The internal dynamics of content-based Web communities depend on a range of forces that shape how these communities grow over time in content and population. In this work we present an analysis of the evolution of a large set of groups from Flickr, a popular photo sharing Web service. We show in particular the influence that a number of factors (such as social network structure, demographic profile and governance structure) may have on the growth of these communities and suggest a number of research directions on the relation between the dynamics of Web communities and their underlying social network.

Introduction

Social networking 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 behavior of Web users that these services can provide. Social scientists have extensively started using 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; Kumar, Novak, and Tomkins, 2006; Kossinets and Watts, 2006; Ali-Hasan and Adamic, 2007; Golbeck, 2007; Mislove et al., 2007, 2008; Leskovec et al., 2008).

The social Web, however, affords much more than an infrastructure for the creation of links among individuals. It has become an extraordinary vehicle for the support of online communities. The specificity of services supporting online communities, as opposed to "pure" social networking services, is that they provide an infrastructure where users are not only able to create new social links but also to *share content*, whether in the form of collaborative content production (such as in wikis or open source communities), content sharing (such as in photo, music or video sharing services), content annotation (such as in social bookmarking websites) or content-driven discussion (as in discussion forums or review-based services). Examples of popular social networking services supporting "pure" social networks include *LinkedIn*, *Facebook*, *MySpace*, *Bebo*. Services hosting content-based communities, in contrast, are often designed to meet the collaborative requirements of specific kinds of content, such as music (*Last.fm*), photography and videos (*Flickr, Ipernity, YouTube*), open source development (*Ohloh*), just to mention a few.

The possibility of studying social interaction mediated by content raises new challenges for social science. If, broadly speaking, the growth of a social network follows dynamics that depend on patterns of interaction and link creation between agents, in the case of *content-based* online communities these dynamics are also affected by patterns of interaction between agents and *content*, by group affiliation effects and by norms that regulate social interaction and content sharing within these communities. In spite of an increasing interest for online Web communities as a source for the study of social networks, relatively little effort has been put into the study of processes at play in the dynamics of simple online social networks as opposed to content-based social networks and full-fledged content-based online communities. In the present work our aim is to start bridging this gap, by focusing in particular on the relations between social network properties and community dynamics. Before delving into the question of why study content-based Web communities, we need to introduce a few caveats and conceptual distinctions.

Defining "communities"

When using the term "communities" we should distinguish between the social-network notion of a community and the full-fledged notion of a community such as the one that we will be referring to in the present work.

Social network research has defined a number of formal properties that allow the characterization of particular structures in a network as "communities" or cohesive groups (see, *inter alia*, Alba, 1973; White, Boorman, and Breiger, 1976; Wasserman and Faust, 1994; Girvan and Newman, 2002; Freeman, 2003; Newman, 2006). These formally defined communities are extracted in a bottom-up way by looking at network properties and can be used to make predictions

*DT and CR equally contributed to this paper. AB provided social network measurements, part of a larger dataset from the TAGora project, which we acknowledge for the permission to use this data.

Copyright © 2009, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

on the behavior of specific sets of agents (e.g. authors massively citing each others, users using similar tags to annotate content, agents that are implicitly related with each other with respect to a given range of interests and preferences and so on).

In contrast, when using the term of Web communities, we refer to online social structures that obey to stronger constraints than those described by social network science. Web communities typically require users to establish an explicit link of *affiliation* to the community and include norms on participation and social behavior that every community member should comply with.

Most online networking services support the creation of this kind of communities. Web 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 content sharing or creation). Web communities as defined in this stronger, top-down sense, allow us to characterize the notion of a community-centered social network that we can define as the network of explicit links between agents that fall within the scope of a given Web community.

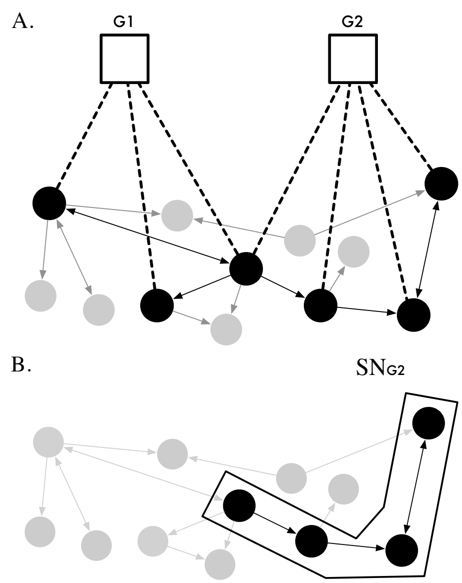

Figure 1: **A. Social network and group membership:** solid arrows represent directed agent-to-agent links, dashed arrows represent agent-to-group affiliation links; highlighted nodes represent agents with at least one group affiliation. **B. Group-centered social network:** highlighted nodes and links represent the subset of the global social network internal to group G2.

This distinction between user-centered vs. group-centered social networks as well as the bottom-up vs. top-down characterization of communities allow us to formulate a number of research questions calling for empirical clarification.

- How does the structure of group-centered social networks affect the overal growth dynamics of groups in terms of content and population?
- What is the respective contribution of user-centered and group-centered social networks in the dissemination of content in Web communities?
- Are there measurable biases on the kind of contributions selected for inclusion in a content-based Web community (e.g. a highly moderated photo community) that depend on the structure of its underlying social network?

The present work focuses on the first of these questions by looking at macroscopic community dynamics in a popular content-based Web service.

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 mutual effects of content-dependent interaction, community affiliation and social network dynamics: its user model 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 behavior mediated by *user-to-content* links; finally, by supporting with a dedicated infrastructure the creation of communities of interest or "groups", it represents an extraordinary testbed for studying the role of *user-to-group* affiliation links. Thanks to a rich and extensively documented API[1], Flickr allows the extraction of large datasets relevant for the study of all of these levels of description (content, individual users, groups).

Related work

Flickr attracted a fairly large attention (compared to its relative small age) in the research community. Most studies used Flickr as a large data source to study tagging behavior and folksonomy (Marlow et al., 2006; Nov, Naaman, and Ye, 2008; Plangprasopchok and Lerman, 2008; Sigurbjörnsson and van Zwol, 2008). A smaller number of works, more relevant to the present analysis, focused on social interactions and group-related social behavior.

Social network research on Flickr

If we leave aside studies attempting to model "pure" social network evolution in Flickr (Mislove et al., 2008; Leskovec et al., 2008), the effects of social networks on content popularity and dissemination are one of the most prominent areas of research for which Flickr has been studied. Lerman and Jones (2007) studied patterns of browsing behavior, 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 social networks of "friends" (or "contacts" as they are referred to in Flickr).

[1] http://flickr.com/services/api

A similar conclusion is reached by Cha et al. (2008), who investigated the role of social cascades in Flickr, looking at how the social network of members mediates the rapid spreading of popularity in content. Their findings support the idea that (1) 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 (2) social network structure can help predict patterns of popularity of content. This is also consistent with van Zwol (2007)'s analysis of the factors affecting the dynamics of popularity of content in Flickr (i.e. the temporal profile of the number of views a photo receives) which confirms that user-centered social networks are the most prominent vehicle of content dissemination among Flickr users.

Research on Flickr groups

Surprisingly, a central social feature of Flickr, i.e. *groups*, has not attracted a considerable 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 flourish on online networking services) have a particular status because, as opposed to purely user-centered social networks, they can be described as communities of interest driven by shared content. The study of content dissemination in the works cited above focuses on aspects that are social, in that they are mediated by social connections, although not strictly speaking *collaborative*. Flickr groups, on the contrary, are specifically designed to enable *collaborative* content 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 for being able to share content.[3] *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 only*, so that users can only upload content if they are explicitly invited by other group members. Groups have a governance structure made 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).

Mislove et al. (2007) conducted a general analysis of Flickr groups in the context of a comparison of high-level statistics of similar social networking services. In spite of large discrepancies in group use across these services, the same global trends were identified, showing in particular that (1) groups represent communities of users characterized by highly dense networks (as opposed to users with lower than average group participation), (2) that members of small user groups tend to be more clustered than those of larger groups and (3) that the most sociable users (those with a high 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 typologies 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.

A first extensive analysis on group participation based on a static snapshot of Flickr groups was conducted by Negoescu and Perez (2008). Their analysis illustrates in particular statistics on the *loyalty* of group contributions by the same users. The results indicate that users tend to systematically share a limited amount of photos with the same, limited number of groups. However a very high variability in user behavior suggets 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.

Unfortunately, all of these studies are based on static snapshots of Flickr groups and some of the most interesting research directions they hint at may not be empirically settled without data on the evolution over time of groups and their underlying social networks. The present study is a first step in this direction, aiming at understanding the macroscopic forces behind group dynamics and calling for similar studies on more specific aspects of group-driven online social interaction.

Method

Dataset

The data used for this study consists of a sample of 7,500+ Flickr groups whose variations were tracked on a daily basis for a period of 8 months between 2007 and 2008.

The data was obtained via Flickr Group Trackr[4], a free web service that we developed in order to allow group members to track the evolution of their group over time. For each Flickr group registered to the service, Group Trackr pulls

[2]There is disagreement on estimates of group participation with respect to the whole Flickr user base. Mislove et al. (2007) mentions 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 photo with at least one group. None of these studies seems to consider in these estimate the potentially large number of users contributing to private groups only.

[3]This may not be true any more, since Flickr introduced photo-specific invitation links that allow group administrators to request to a non-member to contribute a picture to the group pool *without* the requirement of joining the group.

[4]http://dev.nitens.org/flickr/group_trackr.php

a series of data calling the Flickr API on a daily basis, including: *size of the pool* (or *number of pictures uploaded to the group*), *population*, *privacy level*, *moderation features*, *throttling type and level*. Changes along any of these variables can hence be identified with a precision of 24hrs.

The Flickr Group Trackr dataset was complemented with a static snapshot of the same set of groups providing detailed information relative to the beginning of the tracking period on: (1) *user-to-group* affiliation links (2) *user-to-user* contact links.[5]

Sample restrictions

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 ($p < 100$), whose dynamics are too dependent on the behaviour of individual members to allow generalizations;
- to capture the natural dynamics of these groups we 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.
- 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 $3,024$ groups.

Analysis

Global regression model. To shed light on the joint contribution of various quantitative and macroscopic factors on the demographic evolution of Flickr groups, we first introduce regression models of growth rates over the whole observation period as a dependent variable. We check several model designs, based on various factors belonging to a main set of observed independent variables. These variables include:

1. demographic variables such as the number of *users* (U), the quantity of *pictures* (P);
2. structural variables related to basic topological features of the underlying social network, namely the *average degree* ($\bar{k}$), the *density* (d), and *the reciprocity index* (r).

   The degree measures the connectivity of a user, i.e. their total number of contacts. The density of a group is defined as the ratio between the number of existing links and the total possible number of links within the group-centered social network (i.e. [group size] $*$ ([group size] $- 1)/2$). The reciprocity index is the proportion of *reciprocated* or symmetrical contact links within the group and per group user, averaged over the whole group.
3. governance values, such as the existence of a *moderation filter* M and indices based on throttling when this notion is quantifiable, i.e. when a definite upload limit is set up such that a limited number of pictures can be uploaded per user per time period (day, week or month). This defines the *throttling index* θ.

We considered a linear regression of the logs of each variable, when possible. For instance, the regression equation underlying the first model on Tab. 3 is:

$$\log(\frac{U_{\text{end}}}{U_0}) = \lambda_0 + \lambda_1 \log(U_0) + \lambda_2 \log(\frac{U_0}{P_0}) + \lambda_3 \log(\bar{k}) + \lambda_4 \log(d) + \lambda_5 \log(1+r) \quad (1)$$

Growth landscapes for single factors. To complement the results based on this global model, we additionally examined and measured the *individual* impact of each factor on the growth of groups. To this end, we followed a methodology similar to the one adopted in Roth, Taraborelli, and Gilbert (2008) for the study of the dynamics of a large sample of wikis. Two snapshots for each group were compared at the beginning (t_0) and at the end (t_{end}) of the tracking period and group growth rates were calculated as the ratio in population and content sizes between t_{end} and t_0. We then ranked groups along a given variable in 6 quantiles, each containing therefore approximately 15% of our complete dataset. For each quantile, we calculated and plotted growth rate means for content and population. Contour plots eventually represent growth rates along a further dimension: group population, as a control for group size.

Thus, while such a framework admittedly does not render joint correlations which may occur between the various independent variables, it allows a finer, more precise observation of the potential contribution of each individual dimension.

Dynamic regression for individual groups. The previous models are meant to reveal the impact of various factors at a macro-scale, aggregated for all groups. Yet, it is impossible to tell from these results the shape of the (short-term) demographic behavior of each of these groups. To provide group-centered predictions of the possible demographic paths, we examined the peculiar growth dynamics of each Flickr group with respect to several of the above-mentioned factors.

In order to do so, we relied on information containted in group histories spanning over 8 months by applying a regression on the daily dynamics of the groups. More precisely, we appraise the impact of individual factors on the local growth between t and $t + \Delta_t$ through a group-based regression model which accounts for the daily variation in content and population as a linear combination of factors. In other words, in the case of user growth for each group we fit a model of the kind:

$$\log(\frac{U_{t+\Delta_t}}{U_t}) = \lambda_0 + \lambda_1 \log(U_t) + \lambda_2 \log(P_t) + ... \quad (2)$$

From this kind of regression we are able to assign a signature vector $\sigma_i = (\lambda_1, \lambda_2, ...)$ to group i. We distinguish

[5] This snapshot was made available from the TAGora project (*Flickr user and group membership dataset*, http://www.tagora-project.eu/data/#flickrgroups).

signatures of user growth and picture growth by using σ_i^U and σ_i^P respectively.

Each group is thus defined by a compact signature describing the effects of the various factor on its daily dynamics, and it may be represented in a parameter space by a point of coordinates σ_i^U and σ_i^P, respectively. It is next possible to determine which regions of the space are occupied by which groups, and notably make inferences on the relationships between the various estimates — for instance by exhibiting that when group dynamics are positively influenced by some factor, they are also generally negatively affected by some other factor. Note that by reason of a restriction in the scope of our dataset, social-network related data is not available on a day-to-day basis and could not be integrated in the signature computation.

Results

Regression analysis

We examined three models aiming at exhibiting the effect of (i) group-centered social network-related variables, (ii) moderation-related variables, and (iii) a combination of both. We also generally kept user and population sizes as controls in the models (see Table 3 for the detailed list of variables for each model). It should be noted that the parameter attached to the ratio of users per picture U/P is linearly dependent of those attached to U and P because of the use of logarithms; therefore, we dit not feature these three variables together.

We first notice that social-network features have a sensible and significant effect (models #1 & #3), to the contrary of moderation variables which display almost no correlation (models #2 & #3) — a point we will explore further below with growth landscapes. In more details, both the average degree and the reciprocity index induce a negative correlation with growth. This suggests that group cohesiveness, as measured distinctly by these two variables, tends to limit growth. The role of density, which is positively correlated, is more difficult to interpret as this value also partly depends on group size. Second, parameters linked to moderation variables exhibit a surprisingly small significance (both in terms of magnitude and statistical evaluation), a fact further evidenced by the negligeable R^2 values for the model where moderation variables alone are tested (model # 2).

Global growth landscapes

Growth landscapes based on the consideration of individual factors display a number of significant effects. The first striking feature is the strong correlation of growth landscapes for content and population with respect to the same indicators. With the only exception of groups with high values of reciprocity (quantiles 4 and 5 in fig. 6) that appear to have a fairly different impact on content and population growth, all variables seem to affect growth rates with the same strength. This may not come as a surprising result as growth rates in content and population for content-based communities appear to be very tightly interdependent (see for analogous results in the case of wikis: Roth, Taraborelli, and Gilbert, 2008).

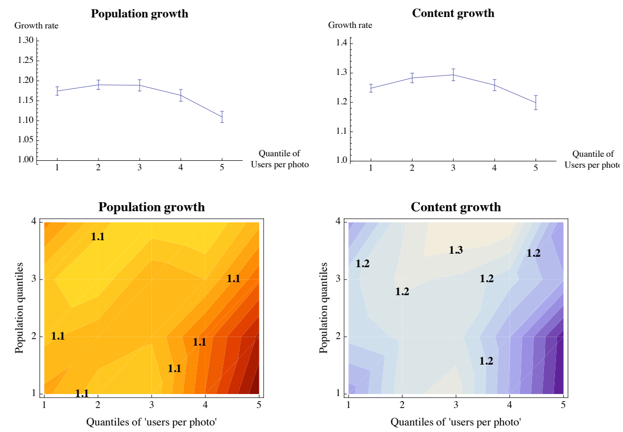

Figure 2: Growth landscapes as a function of the ratio of *users per photo*.

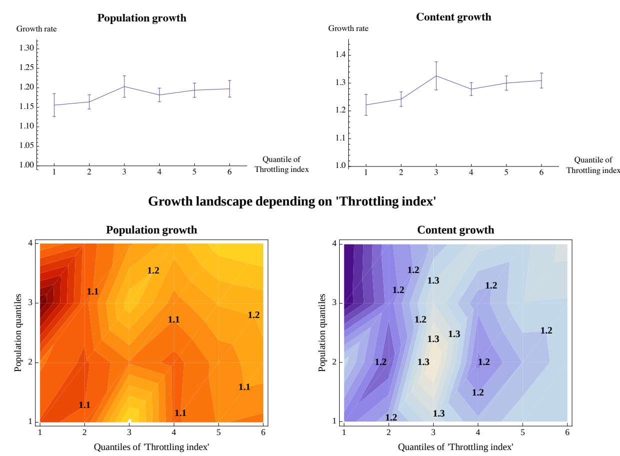

Figure 3: Growth landscapes as a function of throttling index.

Still, the breakdown by population of the impact of these variables indicates that effects are very unevenly distributed across population sizes: particularly high growth rates appear to be confined to large groups in some cases (as for density, see fig. 5, top right quantile) or conversely slow growth rates be tied to groups with a small population (e.g. users per photo, see fig. 2, bottom right quantile). Additionally, the particular "flat" profile of some indicators (especially for groups of medium size) provides evidence in support of the hypothesis that individual social network properties cannot be taken as reliable predictors of particularly strong group growth.

Growth signatures

Signatures of both user growth and content growth were computed for a series of models, including raw demographic variables (U and P), (delayed) growths on these variables ($U_t/U_{t-\Delta_t}$ and $P_t/P_{t-\Delta_t}$), and governance-related vari-

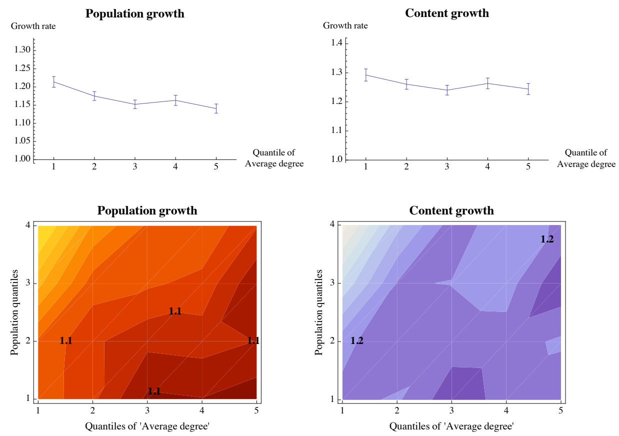

Figure 4: Growth landscapes as a function of average degree.

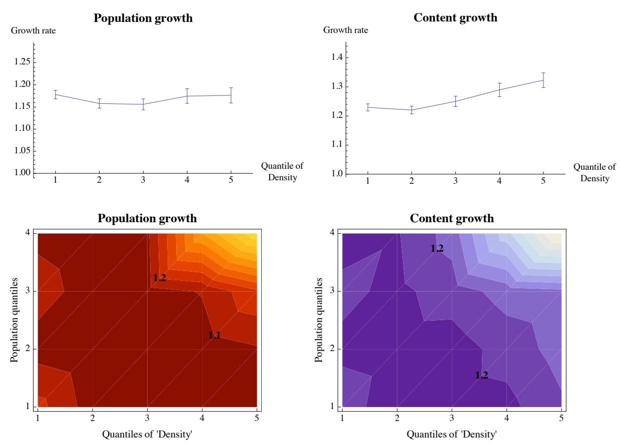

Figure 5: Growth landscapes as a function of density.

ables (moderation, throttling) — as said before, dynamic social network data was not available.

Signatures were generally based on significant estimates in a manner consistant with the global analysis: that is, demographic variables exhibited significant effects, whereas governance factors seemed to yield largely less significant estimates.

In particular, we focused on the following models:

$$\log(\frac{U_{t+\Delta_t}}{U_t}) = \lambda_0 + \lambda_U \log(U_t) + \lambda_P \log(P_t) \qquad (3)$$

which provided particularly significant correlations between the estimates. A scatterplot of the corresponding user growth signatures $\sigma^U(\lambda_U, \lambda_P)$ is drawn on Fig. 7. A similar scatterplot can be observed for picture growth signatures σ^P (correspondingly based on the same set of dependent variables), not plotted here. As can be seen, the negative correlation between the two variables is sensible, with Pearson correlation coefficients of $-.69$ and $-.53$, respectively.

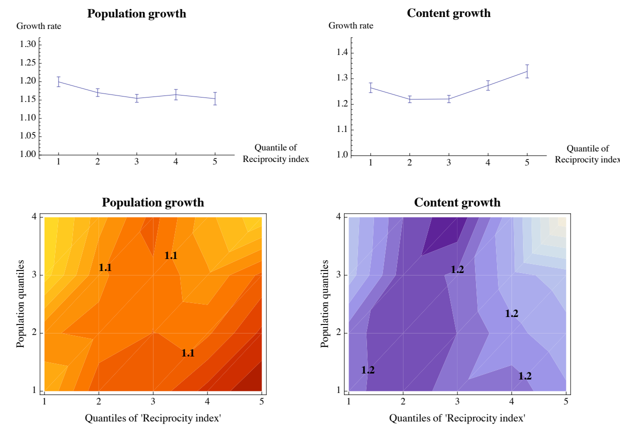

Figure 6: Growth landscapes as a function of reciprocity.

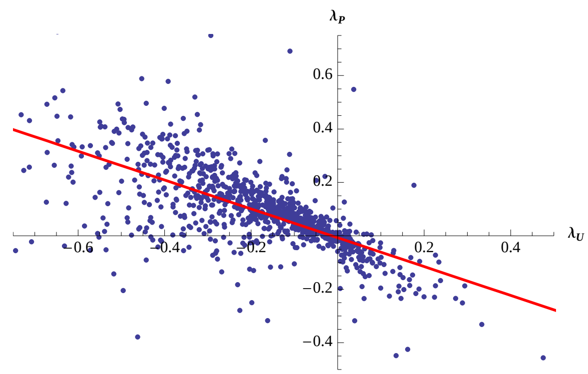

Figure 7: Scatterplot of user growth signatures (σ^U) in a (λ_U, λ_P) parameter space.

As a result of the negative correlation, it is possible to roughly dichotomize groups into two classes, depending on the sign of λ_U, generally opposite to that of λ_P. These findings tend to suggest that, in terms of instantaneous growth, the roles of both underlying demographic variables U and P are mutually exclusive: the more the population size is positively correlated with growth, the more the content size affects growth negatively — and conversely. In other words, if a group tends to grow with higher *population values*, it tends to correspondingly grow less with higher *content sizes*. Additionally, averages of (λ_U, λ_P) are $(-0.19, 0.09)$ for user growth and $(-0.85, 0.84)$ for picture growth, which is therefore more significantly affected by this effect.

Conclusions

Content-based online communities and their relation to social networks represent a promising new area of investigation for social scientists that can benefit from a range of data sources available from online networking services. The specific nature of interactions afforded by these communities (i.e. content-dependent agent-to-agent interactions and group affiliation links) calls for the development of new models to account for the evolution of these communities, beyond those that traditionally apply to the explanation of social network dynamics. This is, in particular, a challenge for research attempting to understand the processes behind collaborative knowledge production as well as community-based participatory decision making.

We presented an empirically assessment of the intertwinement of demographic factors, social network properties and governance structure among the forces jointly driving the macro-level growth of Web communities. Further research will need to investigate the specific role of social network features in the micro-evolution of such communities.

Acknowledgments

We are grateful to Nigel Gilbert, the members of CRESS and the participants in the 2008 Dagstuhl seminar "Social Web Communities" for valuable feedback and insights on earlier versions of this work. The present research was partly supported by the PATRES (NEST-043268) and TAGora (FP6-IST5-34721) projects funded by the European Commission.

References

Adamic, L. A., and Glance, N. 2005. The political blogosphere and the 2004 U.S. election: divided they blog. In *LinkKDD '05: Proceedings of the 3rd international workshop on Link discovery*, 36–43. New York, NY, USA: ACM Press.

Alba, R. D. 1973. A graph-theoretic definition of a sociometric clique. *Journal of Mathematical Sociology* 3:113–126.

Ali-Hasan, N., and Adamic, L. A. 2007. Expressing social relationships on the blog through links and comments. In *International Conference on Weblogs and Social Media*.

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.

Freeman, L. C. 2003. Finding social groups: A meta-analysis of the Southern women data. In Breiger, R.; Carley, K.; and Pattison, P., eds., *Dynamic Social Network Modeling and Analysis*. Washington, D.C.: The National Academies Press. 39–97.

Girvan, M., and Newman, M. E. 2002. Community structure in social and biological networks. *PNAS* 99(12):7821–7826.

Golbeck, J. 2007. The dynamics of web-based social networks: Membership, relationships and change. *First Monday* 12(11).

Kossinets, G., and Watts, D. J. 2006. Empirical analysis of an evolving social network. *Science* 311:88–90.

Kumar, R.; Novak, J.; and Tomkins, A. 2006. Structure and evolution of online social networks. In *Proc. 12th ACM SIGKDD*, 611–617. 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 *KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 462–470. New York, NY, USA: ACM.

Marlow, C.; Naaman, M.; Boyd, D.; and Davis, M. 2006. Ht06, tagging paper, taxonomy, Flickr, academic article, to read. In *HYPERTEXT '06: Proceedings of the seventeenth conference on Hypertext and hypermedia*, 31–40. New York, NY, USA: ACM Press.

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*, 29–42. New York, NY, USA: ACM.

Mislove, A.; Koppula, H. S.; Gummadi, K. P.; Druschel, P.; and Bhattacharjee, B. 2008. Growth of the Flickr social network. In *WOSP '08: Proceedings of the first workshop on Online social networks*, 25–30. 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*, 417–426. New York, NY, USA: ACM.

Newman, M. E. J. 2006. Modularity and community structure in networks. *PNAS* 103(23):8577–8582.

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*, 1097–1100. New York, NY, USA: ACM.

Plangprasopchok, A., and Lerman, K. 2008. Constructing folksonomies from user-specified relations on Flickr.

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

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: Proceedings of the 4th International Symposium on Wikis*. New York, NY, USA: ACM.

Sigurbjörnsson, B., and van Zwol, R. 2008. Flickr tag recommendation based on collective knowledge. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, 327–336. New York, NY, USA: ACM.

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

Wasserman, S., and Faust, K. 1994. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

White, H. C.; Boorman, S. A.; and Breiger, R. L. 1976. Social-structure from multiple networks. I: Blockmodels of roles and positions. *American Journal of Sociology* 81:730–780.

Quantile	1	2	3	4	5	
Users per picture	0.052	0.095	0.15	0.26	3.43	
Average degree	1.11	2.66	4.58	8.25	23.6	
Reciprocity	0.29	0.44	0.53	0.63	0.78	
Density	0.0031	0.008	0.015	0.027	0.089	
	1	**2**	**3**	**4**	**5**	**6**
Throttling index	0.57	1	2	3	5	10

Table 1: Mean values of quantiles.

Effect	Model 1		Model 2		Model 3	
	param.	s.e.	param.	s.e.	param.	s.e.
Intercept	0.20	(0.04)	0.32	(0.04)	0.20	(0.06)
Users	0.22	(0.01)	-0.006	(0.01)	0.25	(0.01)
Users per photo	-0.04	(0.00)	—		-0.03	(0.01)
Average degree	-0.22	(0.01)	—		-0.28	(0.01)
Reciprocity	-0.18	(0.07)	—		-0.11	(0.10)
Density	0.21	(0.01)	—		0.25	(0.01)
Moderation			-0.02	(0.02)	0.00	(0.02)
Throttling index			0.03	(0.01)	0.01	(0.01)
R^2		*0.27*		*0.01*		*0.33*

Table 2: Regressions for population growth.

Effect	Model 1		Model 2		Model 3	
	param.	s.e.	param.	s.e.	param.	s.e.
Intercept	0.17	(0.05)	0.46	(0.04)	0.12	(0.06)
Pictures	0.28	(0.01)	-0.01	(0.00)	0.32	(0.01)
Users per photo	-0.25	(0.01)	—		-0.30	(0.01)
Average degree	-0.32	(0.01)	—		-0.37	(0.01)
Reciprocity	-0.38	(0.08)	—		0.56	(0.11)
Density	0.28	(0.01)	—		0.32	(0.01)
Moderation			-0.06	(0.02)	-0.07	(0.02)
Throttling index			0.04	(0.01)	0.02	(0.01)
R^2		*0.30*		*0.02*		*0.40*

Table 3: Regressions for picture growth.