Characterizing and Predicting Activity in Semantic MediaWiki Communities

Simon Walk[1] and Markus Strohmaier[2,3]

[1] Institute for Information Systems and Computer Media, Graz University of Technology, Graz, Austria
[2] GESIS - Leibniz Institute for the Social Sciences, Cologne, Germany
[3] Dept. of Computer Science, University of Koblenz-Landau, Koblenz, Germany

Abstract. Semantic MediaWikis represent shared and discretionary databases that allow a community of contributors to capture knowledge and to specify semantic features, such as properties for articles, relationships between articles, or concepts that filter articles for certain property values. Today, Semantic MediaWikis have received a lot of attention by a range of different groups that aim to organize an array of different subjects and domain knowledge. However, while some Semantic MediaWiki projects have been thriving, others have failed to reach critical mass. We have collected and analyzed a total of 79 publicly available Semantic MediaWiki instances to learn more about these projects and how they differ from each other. Further, we conducted an empirical analysis using critical mass theory on Semantic MediaWiki communities to investigate whether activity or the number of registered users (or a mixture of both) are important for achieving critical mass. In addition, we conduct experiments aiming to predict user activity and the number of registered users at certain points in time. Our work provides new insights into Semantic MediaWiki communities, how they evolve and first insights into how they can be studied using critical mass theory.

1 Introduction

Semantic MediaWikis are open repositories for structured data that can be edited by a community of users, who are interested in digitally modeling and representing domains. These Wikis have been used to capture knowledge from a wide variety of different domains, including for example beaches[4], games[5] or academic institutions[6].

Although Semantic MediaWikis have matured technologically, we still don't have a good understanding about the social processes behind them, e.g. why some Semantic MediaWiki communities are thriving and others are failing to reach critical mass. In this paper, we are using principles of critical mass theory

[4] http://beachapedia.org/
[5] http://nobbz.de/wiki/
[6] http://www.aifb.kit.edu/portal

to investigate activity and community growth in 79 publicly available Semantic MediaWikis with the goal of identifying and comparing factors that directly influence community growth and activity in said instances. In the context of online platforms, critical mass is often referred to as the amount or number of "something" (e.g., a feature or quality) that has to be reached for a system to become self-sustaining [8–10]. In terms of Semantic MediaWiki communities we want to know what this "something" is and if it is the same as it is for other systems and communities. In our empirical analysis we will look at *activity*, i.e. the accumulated number of changes contributed by the corresponding community to each Semantic MediaWiki at certain points in time. In addition, we will study the role of *community growth* via the number of accumulated unique users that have contributed to the Wikis at certain points in time. In particular, we are going to investigate whether activity or community growth (or a mixture of both) are important for achieving critical mass and predicting activity as well as community growth in Semantic MediaWikis at certain points in time. Answering these questions will fuel our understanding of how Semantic MediaWiki communities operate and evolve over time.

The remainder of this paper is structured as follows. In Section 2 we will present related work as well as work that has inspired the analysis conducted in this paper. A short characterization of the crawled Semantic MediaWiki instances and a description of the used methods for our analyses can be found in Section 3. The results and interpretations of our analyses are presented in Section 4. We conclude this paper in Section 5 and highlight future work.

2 Related Work

The work presented in this paper builds upon work in the areas of critical mass theory and collaborative ontology engineering.

2.1 Critical Mass Theory

In 1985, Oliver and colleagues [8–10] have discussed and analyzed the concept of critical mass theory by introducing so called production functions to characterize decisions made by groups or small collectives. Essentially, these production functions represent the link between individual benefits and benefits for the group.

They argue that when achieving critical mass of users, collective goods of groups are limited, thus interest can not be maintained longer than the limited (collective) resource allows for. In the case of online communities, the collective goods are not limited, theoretically allowing for an infinite increase in users. However, without users motivated in contributing, interest will decrease and critical mass will lose momentum and ultimately decelerate. In their work, three different types of production functions are identified: *Accelerating*, *decelerating* and *linear* functions (see Figure 2). The idea behind accelerating production functions is that each contribution is worth more than its preceding one. In a decelerating production function the opposite would be the case, resulting in

each succeeding contribution to be worth less than the preceding one. Until today it is still mostly unclear what these production functions look like for online communities and online production systems. Depending on the investigated or desired point of view, different aspects of these communities and online production systems can be used to calculate production functions. According to Solomon and Wash [15] it is still unclear which features of an online community characterize critical mass. One approximation they used was the activity and number of users for calculating and predicting critical mass in traditional WikiProjects. The authors argue that activity, for online production systems, after certain amounts of time is the best indicator of a self-sustaining system. In this work, we will adopt the same approach to characterizing critical mass for Semantic MediaWikis. Having an accelerating production function for the number of registered users and activity would indicate that users are interested in the collective good (e.g., the WikiProject) but also contribute to it (measured through activity). Achieving accelerating production functions for both of these factors critically promotes achieving critical mass. Once accelerating functions are reached, critical mass is likelier to follow, as interest (and pay-off) increases and user contributions rise, until the maximum potential of a system is reached.

The analysis of Oliver and colleagues [9] also highlights that different production functions can lead to very different outcomes in similar situations. For example, given an accelerating production function, users who contribute to a system are likely to find their potential contribution "profitable", as each subsequent contribution increases the value of their own contribution. Naturally, this increases the incentive to make larger contributions to begin with. Given a deceleration production function, users would not immediately see the benefit of large contributions, given that each subsequent contribution is increasing the overall value less, while more effort, in the form of larger contributions, is needed to turn a decelerating production function into an accelerating one.

Raban et al. [13] investigated factors that allow for a prediction of survival rates for IRC channels and characterized the production function of these chat channels as the best-fitting function for the curve that is generated when plotting the number of unique users versus the number of messages posted at certain (ascending) points in time.

Cheng and Bernstein [2] have analyzed concepts of activation thresholds, which resemble features that, when achieved, can help to reach and sustain self-sustainability. They created an online platform that allow groups to pitch ideas, which only will be activated if enough people commit to it.

Recently, Ribeiro [14] conducted an analysis of the daily number of active users that visit specific websites, fitting a dynamic model that allows to predict if a website has reached self-sustainability, defined through the shape of the curve of the daily number of active users over time. He uses two constants α and β, where α represents the constant rate of active members influencing inactive members to become active. β describes the rate of an active member spontaneously becoming inactive. Whenever $\frac{\beta}{\alpha} \geq 1$ a website is unsustainable and without intervention the daily number of active users will converge to zero. If $\frac{\beta}{\alpha} < 1$ and the number

of daily active users is initially higher than the asymptotic one, a website is categorized as self-sustaining.

2.2 Collaborative Ontology Engineering

The Semantic Web community has developed a number of tools aimed at supporting the collaborative development of ontologies. For example, Semantic MediaWikis [7] and some of its derivatives, such as OntoWiki and Moki [1, 4], add semantic, ontology modeling and collaborative features to traditional MediaWiki systems. In particular, OntoWiki represents a semantically enriched Wiki that supports collaborative ontology engineering, focussing on the acquisition of instance data and not the ontology or schema itself. MoKi is another collaborative tool that is implemented as an extension of a MediaWiki, which has already been deployed in a number of real world use cases.

Gil et al. [5,6] empirically analyzed different aspects of 230 different instances of Semantic MediaWikis, with a focus on the evolution of semantic features, such as properties and concepts. Among other things, they found out that in the investigated Semantic MediaWiki instances, categories were still much more popular than concepts. However, structured properties were used by all Wikis with a total of 50 instances exhibiting > 100 defined properties.

Protégé, and its extensions for collaborative development, such as WebProtégé [18] and iCAT [17], are prominent stand-alone tools that are used by a large community worldwide to develop ontologies in a variety of different projects.

To learn more about the nature of the engineering processes that occur when collaboratively developing an ontology, Pöschko, Walk and colleagues [12, 19] have created *PragmatiX*, a web-based tool to visualize and analyze a collaboratively engineered ontology.

Falconer et al. [3] investigated the change-logs of collaborative ontology-engineering projects, showing that users exhibit specific roles, which can be used to group and classify users, when contributing to the ontology. Walk et al. [20] applied Markov chains on the structured logs of changes of five collaborative ontology-engineering projects to extract sequential patterns. Pesquita and Couto [11] analyzed if the location and specific structural features can be used to determine if and where the next change is going to occur in a large biomedical ontology. Strohmaier et al. [16] investigated the hidden social dynamics when collaboratively developing an ontology providing new metrics to quantify various aspects to characterize collaborative engineering processes. Wang et al. [21] used association-rule mining to analyze user editing patterns in collaborative ontology-engineering projects.

3 Materials & Methods

We first characterize activity and community growth of our collected Semantic MediaWiki instances by applying principles of critical mass theory. We then continue our analysis and investigate if activity and community growth are good

predictors for determining the number of changes and users of Semantic MediaWikis at certain points in time. We comparing our results to what has been uncovered by Solomon and Wash [15] for WikiProjects, investigating if the number of users in the beginning stages of Semantic MediaWiki projects does play an important role for predicting activity and community growth. To study these effects in Semantic MediaWiki communities, we have crawled a total of 79 Semantic MediaWiki instances, which were all publicly available at the time of writing with the exception of *three* Wikis[7][8][9] that have already been taken offline.

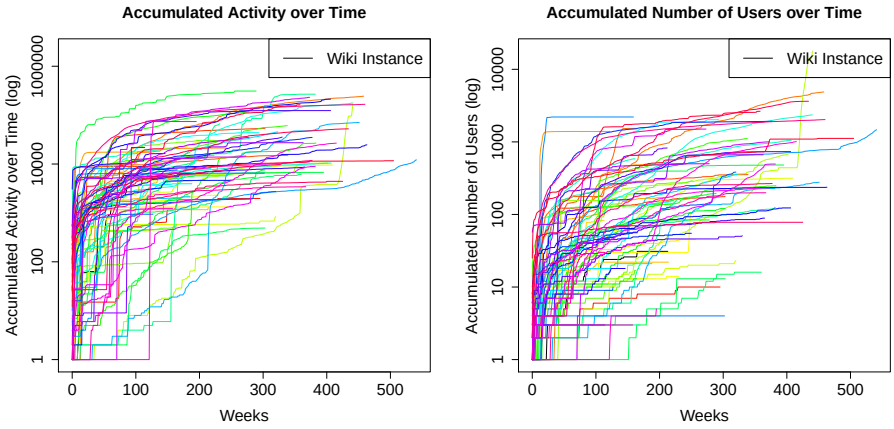
3.1 Semantic MediaWiki Datasets

The datasets used for the analyses in this paper are all randomly selected from different domains and vary in multiple aspects. Due to limitations in space we provide a summary of descriptive statistics for the entirety of our 79 Semantic MediaWikis[10] in Table 1. The number of users ranges from 1 to 85 users for our crawled Semantic MediaWiki instances with a mean of 6.7 unique users and a median of 2 users contributing to the different Wiki instances within the first month of its existence. Similar observations can be made for activity in Semantic MediaWiki communities. Initially we started our analysis with a little over 110 instances. However, due to restrictions necessary for our analyses we had to remove all Wikis with an observable lifespan of < 2 years, explaining the

[7] http://artfriendsgroup.com
[8] http://www.awaycity.com/wiki
[9] http://enlloc.net/hkp/w
[10] See http://www.simonwalk.at/wikis.html for a full list.

Table 1. Characteristics of the 79 datasets used for the prediction of activity and community growth. Community growth, represented as the number of users that have contributed at least 1 change, and activity, represented as the number of changes, are listed as average accumulated numbers over all Semantic MediaWiki instances after 1, 6, 12 and 24 months, as well as at the end of each project. Furthermore, we included the minimum (Min), median, maximum (Max) and standard deviation (SD) for each period. The differences between the Semantic MediaWikis are especially visible when looking at the standard deviation for activity and users during the first 2 years and at the end of our observation periods.

Number of	Timespan	Min	Mean	Median	Max	SD
Changes (Activity)	after 1 month	1	631.42	37	8,796	1,678.79
	after 6 months	1	2,840.91	904	63,547	7,622.06
	after 12 months	1	4,427.04	1,583	90,345	10,871.3
	after 24 months	1	10,595.18	4,694	159,502	21,264.96
	at end	459	41,338.49	11,534	310,933	68,225.41
Users (Number of Users)	after 1 month	1	6.7	2	85	13.09
	after 6 months	1	70.71	9	2,172	287.02
	after 12 months	1	102.35	23	2,203	296.25
	after 24 months	1	184.23	49	2,204	365.82
	at end	3	779.44	194	17,327	2,079.41
Duration	Weeks	113	291.44	295	541	110.08



(a) Accumulated Activity (b) Accumulated Users

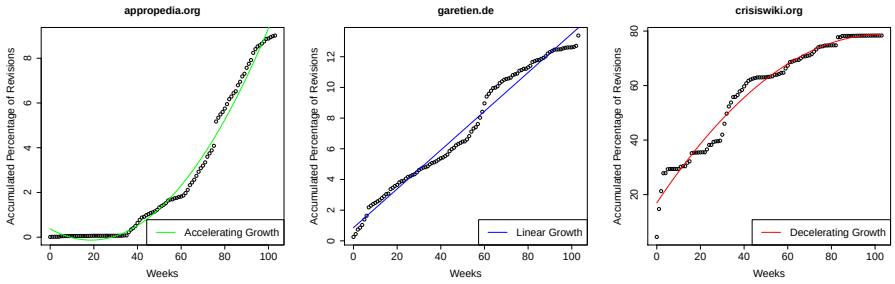
Fig. 1. Activity and Users per Semantic MediaWiki: The x-axes in both plots depict time in weeks, while the y-axes depict the accumulated amount of activity (represented as number of changes) and users during each corresponding week (log-scale). Each line represents one of the 79 Semantic MediaWiki instances. In both plots the differences in duration (x-axes), activity as well as number of users (y-axes) are visible.

minimum duration of 113 weeks. After removing all instances that did not meet the two year requirement we ended up with a total of 79 Semantic MediaWiki communities to investigate.

We have aggregated and accumulated activity and the number of users for each week from the inception of each Semantic MediaWiki until the date of the last observed change. The duration (observation period) of a Semantic MediaWiki instance starts with the first, and ends with the last change in our datasets. Figure 1(a) depicts this accumulated activity per week for every Semantic MediaWiki used in our analyses. Analogously, the accumulated number of users per week for every Wiki instance in our dataset is shown in Figure 1(b). The plots highlight the differences in observation lengths (x-axes), intensity of activity as well as number of users (y-axes, log-scale). Note that the number of users refers to all users that have contributed at least a single change. Anonymous users are represented by their ip address and are not filtered. These differences are also indicating that finding features that are suitable for fitting a general model to predict future information for Semantic MediaWiki communities is a difficult task.

3.2 Critical Mass Theory

We gathered the accumulated number of revisions and unique users after 1, 6, 12 and 24 months to determine the corresponding production functions for



(a) Accelerating Production Function

(b) Linear Production Function

(c) Decelerating Production Function

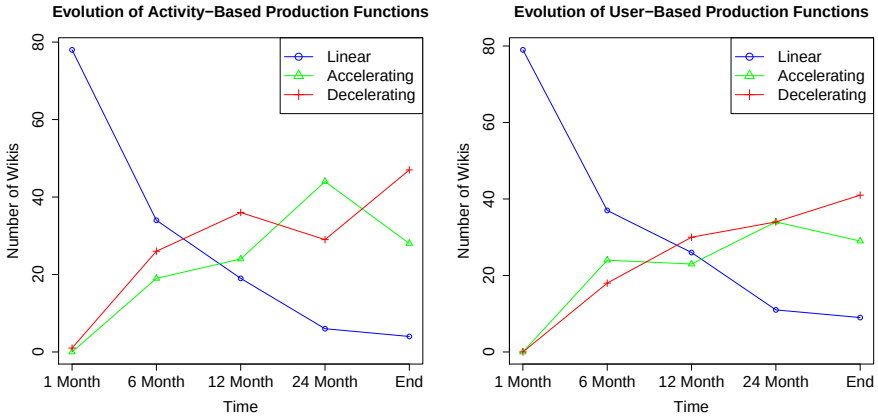
Fig. 2. Types of Production Functions: The x-axes in all three plots depict the time in weeks up to two years, while the y-axes depict the accumulated amount of activity during each corresponding week. The lines in each plot represent the best fitted linear or quadratic function for the observed data (circles).

each Semantic MediaWiki. As depicted in Figure 2, we plotted the accumulated number of users and activity versus elapsed weeks (one data point per week) and fitted a linear and squared function. As described in Solomon and Wash [15], if the squared function is not statistically significantly different from the linear function, the production function was classified as *linear*. If the difference is significant, depending on the priors of the second coefficient, representing the slope of the curve, we classified the production function as *accelerating* (positive coefficient) or *decelerating* (negative coefficient).

3.3 Activity & User Diversity Prediction

To determine if and to what extent features of Semantic MediaWiki communities are usable to determine the overall amount of activity and number of users after two years, we fit multiple regression models to the extracted activity and user data. To avoid any bias from differing overall timespans we use fixed time-intervals (1, 6 and 12 months) for extracting the input data for our regression models. Thus, we collected the accumulated amount of activity and users per week for each Semantic MediaWiki instance after 1, 6, 12 months to predict activity and the number of users after 24 months. Given that the extracted activity and number of users data from our 79 Semantic MediaWiki instances is over-dispersed, meaning that the variances are greater than the means (see Table 1), and the distribution of our extracted Semantic MediaWiki values resemble a negative binomial distribution, we can not use a standard logistic regression approach. Instead, we apply Negative Binomial Regression, which is used with count data that can not be smaller than 0 and follows a negative binomial distribution, on our datasets.

For each dependent variable, we are going to fit three negative binomial regression models, each using input data (activity and number of user) from



(a) Activity-Based (b) User-Based

Fig. 3. Evolution of Production Functions: The y-axes depict the number of Semantic MediaWikis for linear, accelerating or decelerating activity-based (Figure 3(a)) and user-based (Figure 3(b)) production functions at *Time* (x-axes). Somewhere between 6 and 12 months "something" happens that makes some Wikis increase in activity and number of user (accelerating), while others only manage to retain a steady growth (linear) or even slow down in growth (decelerating).

inception up to different points in time (1 , 6 and 12 months). The data points are collected every week and represent the *independent* regression model variables with their corresponding interaction terms. The *dependent* variables that we want to predict are (i) the accumulated number of changes after two years and (ii) the accumulated number of users after two years for each Semantic MediaWiki instance.

4 Results

In this section we present the results of the different analyses described in Section 3.

4.1 Critical Mass Theory Results

The number of Wikis classified according to the corresponding production functions can be seen in Figure 3. The further a Wiki progresses, the less likelier it will be classified with a linear growth function, both for user diversity and activity, evident in the decreasing *linear* lines in Figure 3. For the investigated Semantic MediaWikis the (significant) production functions for activity and number of users exhibit a Pearson correlation coefficient of 0.75, indicating that the two production functions are correlated for each Wiki. This observation is also evident in Figure 4, showing that the majority of Wikis, after two years, exhibit either negative (lower left quadrant) or positive (upper right quadrant) user and

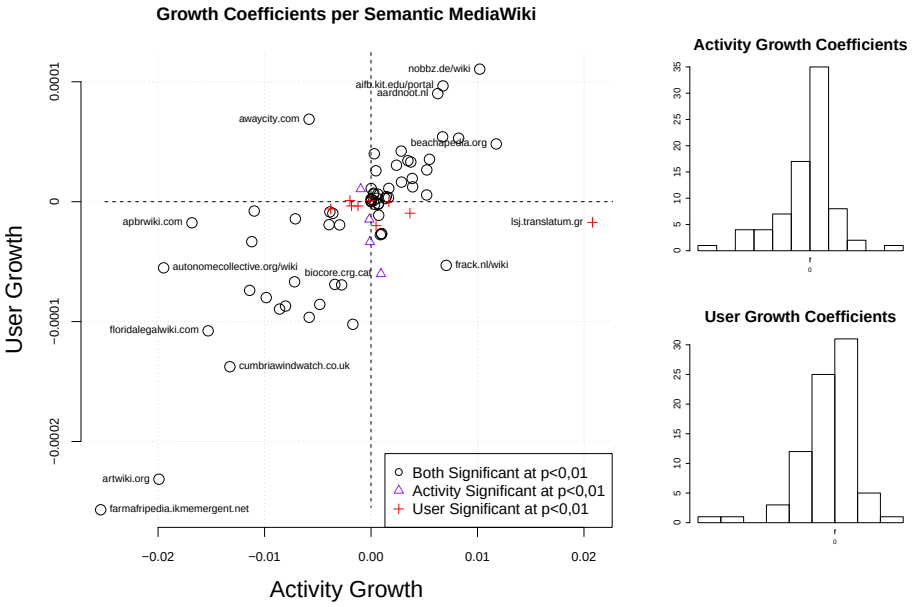
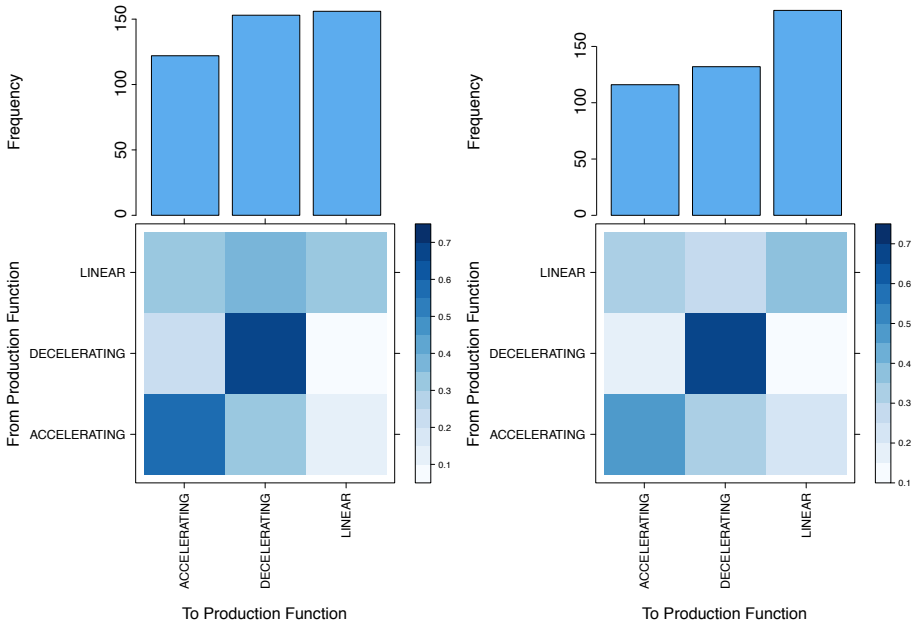


Fig. 4. The Growth Coefficient Distribution: This plot depicts the distribution of growth coefficients for our Semantic MediaWiki instances after after two years. The y-axis depicts the value of the number of users growth coefficients and the x-axis depicts the value of the activity growth coefficients. Each circle represents a Semantic MediaWiki instance with both production functions being significantly different (p-value < 0.01) from linear growth (e.g., http://aardnoot.nl/). Triangles (e.g., http://biocore.crg.cat/wiki) and crosses (e.g., http://lsj.translatum.gr) represent instances where only either the activity or user-based production function is significantly different from a linear function.

activity growth. As can be seen in the histograms, the values of the growth coefficients are equally scattered around positive and negative values. The larger the growth coefficient, the steeper the slope of the resulting production function. We calculated a Pearson correlation coefficient between (significant) user and activity-based growth coefficients of 0.75, indicating that critical mass for Semantic MediaWiki communities is constituted by an immanent correlation of the number of users and activity.

The median R^2 values for the fitted functions of activity and number of users at the different points in time range from 0.83 and 0.78 for the activity and user-based production functions in the first month to a R^2 of 0.95 for both after 2 years. These observed R^2 values represent a (rather) good fit, which also becomes more evident when looking at the sample fits in Figure 2 and the median R^2 values with input data from inception until year one.

To further characterize our investigated Semantic MediaWiki instances we have plotted the user diversity and activity growth coefficients extracted from



(a) Markov Chain Transition Matrix Activity-Based

(b) Markov Chain Transition Matrix User-Based

Fig. 5. Transition Probabilities of Production Functions: Figure 5 depicts the transition probabilities (darker means higher probability) between the different production function shapes of a fitted Markov chain model of first order for all Semantic MediaWikis. Each row in the depicted transition matrix corresponds to one type of production function (linear, decelerating or accelerating). The sum of each row is 1. The plot is always read from row to column, indicated by the axes label *From Production Function* and *To Production Function*. A histogram, highlighting the total occurrences of the different production functions, is depicted on top of the transition matrix.

the previously fitted production function models, using the accumulated number of users and changes from inception until the second year, for each Wiki individually. Figure 4 allows us to plot the different growth coefficients for all 79 Semantic MediaWiki instances, including information about the "intensity" of the observed slopes. Circles represent Semantic MediaWiki instances where both production functions were significantly different from a linear function. Triangles depict Semantic MediaWiki instances with significant activity-based production functions and linear user-based production functions. The crosses follow analogously to the triangles. This means that circles in the top right quadrant are Semantic MediaWiki communities that have an accelerating activity and user diversity production function. We can also see that Semantic MediaWikis have a tendency to exhibit the same production function for activity and user diversity, evident in the number of circles in the upper right and lower left quadrant

of Figure 4. To strengthen our observation we calculated a Pearson correlation coefficient of 0.75 for the different (significant) growth coefficient distributions. Thus, critical mass might be a mixture of the number of users and activity. We have trained a (first-order) Markov chain model, using the chronologically ordered sequences of extracted production functions after 1, 6, 12 and 24 months as input, to analyze whether Semantic MediaWiki communities frequently switch between production functions. For the user-based transition matrix (Figure 5(b)) accelerating and decelerating production functions tend to stay accelerating and decelerating. Linear production functions have a higher tendency to either switch to accelerating or stay linear, than become decelerating. The activity-based production functions (Figure 5(a)) exhibit very strong tendencies to stay at the same state (accelerating and decelerating). If a linear production function was determined for a Wiki, it is similarly likely to continue to exhibit a linear activity production function or switch to an accelerating production function, and is most likely to switch to a decelerating production function. In general, Semantic MediaWikis exhibit a high tendency to stick with their decelerating and accelerating production functions.

For managers of Semantic MediaWikis, this would mean that they would have to monitor both production functions and take action if already one of them is showing first signs of deceleration.

4.2 Factors that drive activity and user diversity

Given the observations made with critical mass theory in Section 4.1 we fitted 6 negative binomial regression models to predict the number of user and activity after two years, using the gathered input data from 1, 6 and 12 months. This method allows us to analyze if activity (and the number of user) after 2 years can best be explained by activity and/or the number of users of preceding points in time. The models are described in more detail in Tables 2 and 3. The goodness of fit for both models is described by the Akaike Information Criterion (AIC) and allows for relative comparisons between the different models. The closer the data that was used for fitting the models is to the target prediction time of two years, the better the model fits the data, evident in (minimally) decreasing AIC values.

When using negative binomial regression to predict the amount of activity after two years in Semantic MediaWikis communities the models show statistically significant effects for activity in all three models (1, 6 and 12 months) on the amount of activity after two years, when holding the number of users constant. When using the model fitted with data after 12 months to predict the activity in a Semantic MediaWiki community (see Table 2) with 500 and 600 users, with an activity of $10,000$ changes, we would expect to have $12,412$ and $12,342$ changes after two years respectively. The fitted model is clearly showing that more users, in the case of our observed Semantic MediaWiki communities, do not automatically mean an increase in activity after two years, which is in contradiction to our intuition after looking at the growth coefficients from the critical mass theory results.

Analogously, when holding activity on a constant level and predicting the number of unique users (or user diversity) after two years in Semantic MediaWikis (see Table 3), the amount of users already present after 1, 6 and 12 months is showing statistically significant effects on the number of users after two years. After 12 months we can determine statistical significance for activity and the (negative) interaction term as well. Similarly, when predicting the number of users in our Semantic MediaWiki communities after two years, using the fitted model after 12 months with $10,000$ and $11,000$ performed changes and 50 users, we would expect to have 99 and 101 users after two years. In contrast to the previous prediction we can observe the positive (and statistically significant for $p < 0.05$) influence of activity on the number of users after 2 years.

This actually means that, with a general model for Semantic MediaWiki communities, activity after two years can be predicted by looking at the activity after 1, 6 and 12 months. The number of users is not significant and, at least in our fitted model, has a negative impact on activity. This would mean (according

Table 2. Predicting Activity: The table depicts the configuration and results for the negative binomial regression model used to predict activity after two years. Input data for the models was the accumulated activity, unique users and an interaction term for both variables after 1, 6 and 12 months.

		Activity After 2 Years				
		Value	Std. Err (Coeff)	Std. Err.	θ	AIC
1 month	# Revisions	0.0004399**	0.000124	0.066	0.4977	1,582.4
	# Users	not sign.	not sign.			
	Revisions:Users	not sign.	not sign.			
6 months	# Revisions	0.00008919**	0.00002084	0.0743	0.5577	1,569.4
	# Users	not sign.	not sign.			
	Revisions:Users	not sign.	not sign.			
12 months	# Revisions	0.00009944**	0.00001445	0.0827	0.6145	1,558.6
	# Users	not sign.	not sign.			
	Revisions:Users	not sign.	not sign.			

$^{*}p < 0.05$; $^{**}p < 0.001$

Table 3. Predicting Users: The table depicts the configuration and results for the negative binomial regression model used to predict the number of users after two years. Analogously to the negative binomial regression model used to predict activity after 2 years, we have accumulated the number of users and activity for each Semantic MediaWiki after 1, 6 and 12 months and used this data (including an interaction term), as input for the listed regression models.

		Users After 2 Years				
		Value	Std. Err (Coeff)	Std. Err.	θ	AIC
1 month	# Revisions	not sign.	not sign.	0.0688	0.5159	947.56
	# Users	0.05386**	0.01839			
	Revisions:Users	not sign.	not sign.			
6 months	# Revisions	not sign.	not sign.	0.0892	0.6501	922.18
	# Users	0.009738**	0.001258			
	Revisions:Users	−0.0000004807**	0.0000001128			
12 months	# Revisions	0.00003025*	0.00001309	0.105	0.751	907.01
	# Users	0.006745**	0.001002			
	Revisions:Users	−0.0000001848*	0.00000008617			

$^{*}p < 0.05$; $^{**}p < 0.001$

to our model) that administrators and managers of Semantic MediaWikis should try to get as much content as possible, as soon as possible into their Wikis to ensure later activity. Critical mass for activity at later stages in a Semantic MediaWiki solely depends on activity in the beginning of a Wiki.

To predict the number of user after 2 years, the number of users after 1, 6 and 12 months are a significant factor. From month 1 to month 12 we can also observe a significance for the interaction term, which further increases in significance until activity becomes significant for the prediction at month 12. For increasing the number of users in a Semantic MediaWiki community, both, the number of users and activity (after a year) have to exhibit a positive (and significant) influence.

5 Conclusions & Future Work

The *main contribution* of this work is the characterization of activity and number of users using approaches of critical mass theory to gauge the viability of Semantic MediaWiki communities. We have studied 79 Semantic MediaWiki projects and their respective production functions over time. In addition, we have fitted negative binomial regression models to predict activity and the number of users after two years. Our approach is not specific to the projects under investigation but can be applied to other (Semantic) MediaWiki projects or collaborative online production systems at scale. In summary, we have found the following:

Semantic MediaWikis exhibit a wide range of evolving production functions: We have shown that the majority of observed Semantic MediaWikis start off with linearly growing activity and numbers of users. This changes within the first 6 to 12 months, which also apparently marks the timeframe where "something" determines if a Wiki will exhibit accelerating, decelerating or linear production functions after two years. At this point we leave it up to future work to further investigate, analyze and determine these influential factors.

Semantic MediaWikis suffer decaying information system lifecycles: The results obtained from the critical mass analysis, as well as the prediction experiment suggest that Semantic MediaWikis are prone to suffer from the vicious circles of decaying information systems. Meaning that Semantic MediaWiki instances that exhibit a decelerating production function (user and/or activity-based) are very likely to keep this decelerating production function, resulting in either less active users or lesser activity, which in turn triggers again less activity or less active users.

Successful Semantic MediaWiki communities start small: Our analysis suggests that the more content is produced by as few users as early as possible, the likelier it is for (our observed) Semantic MediaWikis to reach critical mass and exhibit the highest amount of activity after two years. This also means that the higher the number of users that contribute to a Wiki early on, the lower the amount of activity after two years is going to be. Surprisingly, after 12 months, the amount of activity becomes (positively) significant for the total number of users after 2 years. This indicates that after a certain amount of time (12 months), to attract more users, high activity in a Semantic MediaWiki has a positive effect.

One hypothesis to explain our observations could be that small groups around structured data projects are usually much more focused and devoted, as they need more background knowledge to contribute. However, this could imply that they do not necessarily need to reach critical mass for the number of users, but rather only in terms of activity, as their interest in creating a structured knowledge base already outweighs the efforts of contributing.

Summarizing, we believe that the work presented in this paper represents an important first step towards a better understanding of the factors that drive Semantic MediaWiki communities and their evolution. While our analysis has been initially performed on 79 Semantic MediaWikis and has been limited to user growth and activity, our method can be applied on a wider scale. Future work might focus on investigating additional instances, semantic properties, the evolution of the underlying knowledge base, different kinds of communities and types of Semantic MediaWikis with different motivations and interests, structural properties or additional dimensions of activity, such as passive usage logs (where *visits* are studied in addition to *edits*) or different kinds of activities and specific non-trivial phenomena, such as "edit wars", as well as other log data to expand our understanding of social and community dynamics in such systems.

References

1. S. Auer, S. Dietzold, and T. Riechert. OntoWiki–A Tool for Social, Semantic Collaboration. In *Proceedings of the 5th International Semantic Web Conference (ISWC 2006)*, volume LNCS 4273, Athens, GA, 2006. Springer.
2. Justin Cheng and Michael S Bernstein. Catalyst: Triggering collective action with thresholds. 2014.
3. Sean M. Falconer, Tania Tudorache, and Natalya Fridman Noy. An analysis of collaborative patterns in large-scale ontology development projects. In Mark A. Musen and scar Corcho, editors, *K-CAP*, pages 25–32. ACM, 2011.
4. Chiara Ghidini, Barbara Kump, Stefanie Lindstaedt, Nahid Mahbub, Viktoria Pammer, Marco Rospocher, and Luciano Serafini. MoKi: The Enterprise Modelling Wiki. In Lora Aroyo, Paolo Traverso, Fabio Ciravegna, Philipp Cimiano, Tom Heath, Eero Hyvönen, Riichiro Mizoguchi, Eyal Oren, Marta Sabou, and Elena Paslaru Bontas Simperl, editors, *Proceedings of the 6th European Semantic Web Conference on The Semantic Web: Research and Applications 2009*, pages 831–835, Berlin, Heidelberg, 2009. Springer.
5. Yolanda Gil, Angela Knight, Kevin Zhang, Larry Zhang, and Ricky J. Sethi. An initial analysis of semantic wikis. In Jihie Kim, Jeffrey Nichols, and Pedro A. Szekely, editors, *IUI Companion*, pages 109–110. ACM, 2013.
6. Yolanda Gil and Varun Ratnakar. Knowledge capture in the wild: a perspective from semantic wiki communities. In V. Richard Benjamins, Mathieu d'Aquin, and Andrew Gordon, editors, *K-CAP*, pages 49–56. ACM, 2013.
7. Markus Krötzsch, Denny Vrandecic, and Max Völkel. Semantic MediaWiki. In *Proceedings of the 5th International Semantic Web Conference 2006 (ISWC 2006)*, pages 935–942. Springer, 2006.
8. Gerald Marwell, Pamela E Oliver, and Ralph Prahl. Social networks and collective action: A theory of the critical mass, ill. *American Journal of Sociology*, 94(3):502–534, 1988.

9. Pamela Oliver, Gerald Marwell, and Ruy Teixeira. A theory of the critical mass. i. interdependence, group heterogeneity, and the production of collective action. *American journal of Sociology*, pages 522–556, 1985.
10. Pamela E Oliver and Gerald Marwell. The paradox of group size in collective action: A theory of the critical mass. ii. *American Sociological Review*, pages 1–8, 1988.
11. Catia Pesquita and Francisco M. Couto. Predicting the extension of biomedical ontologies. *PLoS Comput Biol*, 8(9):e1002630, 09 2012.
12. Jan Pöschko, Markus Strohmaier, Tania Tudorache, and Mark A. Musen. Pragmatic analysis of crowd-based knowledge production systems with iCAT Analytics: Visualizing changes to the ICD-11 ontology. In *Proceedings of the AAAI Spring Symposium 2012: Wisdom of the Crowd*, 2012. Accepted for publication.
13. Daphne R. Raban, Mihai Moldovan, and Quentin Jones. An empirical study of critical mass and online community survival. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, CSCW '10, pages 71–80, New York, NY, USA, 2010. ACM.
14. Bruno Ribeiro. Modeling and predicting the growth and death of membership-based websites. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14, pages 653–664, Republic and Canton of Geneva, Switzerland, 2014. International World Wide Web Conferences Steering Committee.
15. Jacob Solomon and Rick Wash. Critical mass of what? exploring community growth in wikiprojects. 2014.
16. Markus Strohmaier, Simon Walk, Jan Pöschko, Daniel Lamprecht, Tania Tudorache, Csongor Nyulas, Mark A. Musen, and Natalya F. Noy. How ontologies are made: Studying the hidden social dynamics behind collaborative ontology engineering projects. *Web Semantics: Science, Services and Agents on the World Wide Web*, 20(0), 2013.
17. T. Tudorache, S. M. Falconer, C. I. Nyulas, N. F. Noy, and M. A. Musen. Will Semantic Web technologies work for the development of ICD-11? In *Proceedings of the 9th International Semantic Web Conference (ISWC 2010)*, ISWC (In-Use), Shanghai, China, 2010. Springer.
18. Tania Tudorache, Csongor Nyulas, Natalya F. Noy, and Mark A. Musen. WebProtégé: A Distributed Ontology Editor and Knowledge Acquisition Tool for the Web. *Semantic Web Journal*, 4(1/2013):89–99, 2013.
19. Simon Walk, Jan Pöschko, Markus Strohmaier, Keith Andrews, Tania Tudorache, Csongor Nyulas, Mark A. Musen, and Natalya F. Noy. PragmatiX: An Interactive Tool for Visualizing the Creation Process Behind Collaboratively Engineered Ontologies. *International Journal on Semantic Web and Information Systems*, 2013.
20. Simon Walk, Philipp Singer, Markus Strohmaier, Tania Tudorache, Mark A Musen, and Natalya F Noy. Discovering Beaten Paths in Collaborative Ontology-Engineering Projects using Markov Chains. *Journal of Biomedical Informatics*, January 2014.
21. Hao Wang, Tania Tudorache, Dejing Dou, Natalya F Noy, and Mark A Musen. Analysis of user editing patterns in ontology development projects. In *On the Move to Meaningful Internet Systems: OTM 2013 Conferences*, pages 470–487. Springer, 2013.