Activity Dynamics in Collaboration Networks

Simon Walk*1, Denis Helic2, Florian Geigl2 and Markus Strohmaier3,4

1IICM - Graz University of Technology
2KTI - Graz University of Technology
3GESIS - Leibniz Institute for the Social Sciences
4University of Koblenz-Landau

May 8, 2015

Abstract

Abstract Many online collaboration networks struggle to gain user activity and become self-sustaining due to the ramp-up problem or dwindling activity within the system. Prominent examples of such networks include online encyclopedias such as Wikipedia or (Semantic) MediaWikis, Question and Answering portals such as StackOverflow, and many others. Only a small fraction of these systems manage to reach self-sustainable activity, a level of activity that prevents the system from reverting to a non-active state. In this paper, we model and analyze activity dynamics in synthetic and empirical collaboration networks. Our approach is based on two opposing and well-studied principles: (i) without incentives, users tend to lose interest to contribute and thus, systems become inactive, and (ii) people are susceptible to actions taken by their peers (social or peer influence). With the activity dynamics model that we introduce in this paper we can represent typical situations of such collaboration networks.

*simon.walk@tugraz.at
Note that this paper is currently under review for publication.
For example, activity in a collaborative network, without external impulses or investments, will vanish over time, eventually rendering the system inactive. However, by appropriately manipulating the activity dynamics and/or the underlying collaboration networks, we can jump-start a previously inactive system and advance it towards an active state. To be able to do so, we first describe our model and its underlying mechanisms. We then provide illustrative examples of empirical datasets and characterize the barrier that has to be breached for a system before it can become self-sustaining in terms of critical mass and activity dynamics. Additionally, we expand on this empirical illustration and introduce a new metric \( p \) —the Activity Momentum—to assess the activity robustness of a collaboration network.

1 Introduction

One of the major problems faced by both new and existing online collaborative networks—such as Wikipedia—revolves around efficiently identifying and motivating the appropriate users to contribute new content. In an optimal scenario, this newly contributed content provides enough incentive for other users to contribute, triggering further actions and contributions. Once such a self-reinforced state of increasing activity is reached, we can say that a system becomes self-sustaining, meaning that sufficiently high levels of activity are reached to keep the system active without external impulses. For example, when looking at well-established collaborative websites, such as StackOverflow or Wikipedia, we already know that at some point in time, these systems have become self-sustaining (in terms of activity), evident in their steady growing number of supporters and overall activity.

However, these self-sustaining states are neither easy to reach nor guaranteed to last. For example, Suh et al. [2009] showed that the growth of Wikipedia is slowing down, indicating a loss in momentum and perhaps even first evidence of a collapse. Moreover, we typically lack the tools to properly analyze these trends in activity dynamics and thus, can not even perform such simple tasks as detecting self-sustaining system states. Therefore, we argue that new tools and techniques are needed to model, monitor and simulate activity dynamics and tendencies for collaboration networks.

The high-level contributions of this work are two-fold. First, we introduce a model that is capable of simulating activity dynamics for online collaboration networks. Second, we describe in detail how to fit the model to empiri-
Figure 1: **Intrinsic Activity and Positive Peer Influence.** Activity dynamics in collaboration networks, represented by users as nodes, collaboration as edges and activity as node size (Figure (a)), are based on two opposing principles. The *Activity Decay Rate* postulates the loss of intrinsic activity (blue color of nodes) per user over time. In contrast, the *Peer Influence Growth Rate* follows the intuition, that users in collaboration networks are (positively) influenced by their peers (yellow color of nodes) where more active peers exercise a higher influence than less active peers. We initialize the network at time $t_0$ with random intrinsic activities. Nodes with a green halo at times $t_1$ to $t_3$ represent users that exhibit a gain in their overall activity between two iterations $t_n$ and $t_{n+1}$, as the exercised positive peer influence is higher than the intrinsic loss of activity. Analogously, red halos represent decreases in overall activity. At first, very central (high degree) nodes with smaller activity values manage to increase their overall activity, while very active central nodes already start to lose activity. After $t_3$ or more iterations, due to overall decreasing activities and hence, decreasing peer influences, all nodes in the collaboration network eventually start to lose activity and inevitably converge towards zero activity.

The proposed model is based on the formalism of continuous deterministic dynamical systems—meaning that activity is modeled by a system of coupled non-linear differential equations. Each user of the system is represented by a single quantity (the current activity), and the social ties between users define the coupling of variables. For simplicity, we do not take individual differences between users into account—the dynamics and its parameters are
the same for each user in the population. This allows us to configure the model with a single parameter, which is a ratio of the following two dynamics parameters, representing two basic activity mechanisms (cf., Figure 1) in online collaboration networks:

(i) Activity Decay Rate $\lambda$, which postulates how fast a user loses her interest to contribute,

(ii) Peer Influence Growth Rate $\mu$, postulating to what extent a user is influenced by the actions taken by her peers.

A first analysis of the model shows that activity dynamics in collaboration networks have an obvious and natural fixed point – the point of complete inactivity – where all contributions of the users have seized. However, by slightly manipulating the parameters in our model we show that it is possible to destabilize the fixed point resulting in a potential increase of activity. We then outline the process of calculating the Activity Decay Rate and Peer Influence Growth Rate for existing collaboration networks, simulate their corresponding activity dynamics and expand our understanding of critical mass in collaboration networks by interpreting our findings.

The remainder of this paper is structured as follows: In Section 2 we introduce and examine our model analytically. We then continue with the model illustration by simulating activity dynamics on a synthetic dataset using our model in Section 3 and outline the process of applying our model on empirical datasets in Section 4. In Section 5 we introduce the notion of System Mass and Activity Momentum, review related work in Section 6 and summarize our findings and discuss limitations and implications for future work in Section 7.

2 Modeling Activity Dynamics

We model activity dynamics in an online collaboration network as a dynamical system on a network. Hereby, the nodes of a network represent users of the system and the links represent the fact that the users have collaborated in the past. We represent the network with an $n \times n$ adjacency matrix $A$, where $n$ is the number of nodes (users) in the network. We get $A_{ij} = 1$ if nodes $i$ and $j$ are connected by a link and $A_{ij} = 0$ otherwise. Since collaboration links are undirected, the matrix $A$ is symmetric, thus $A_{ij} = A_{ji}$, for
all $i$ and $j$. We denote the total number of links in the network with $m$, and thus we get $m = \frac{1}{2} \sum_{ij} A_{ij}$.

Now, we model activity as a continuous real-valued variable $a_i$ evolving on node $i$ of the network in continuous time $t$. The general time evolution equation can be written as follows (see also Newman [2010]):

$$ \frac{da_i}{dt} = f_i(a_i) + \sum_j A_{ij} g_i(a_i, a_j), $$

where $f(a_i)$ specifies the intrinsic activity evolution of node $i$ and $g(a_i, a_j)$ describes the influence of neighbor $j$ on node $i$. To simplify, we assume that the intrinsic activity dynamics as well as the influence of node neighbors are the same for each node $i$ and for each neighbor pair $(i, j)$. This means that we have a single intrinsic activity function $f(a_i)$ for all nodes $i$, as well as a single peer influence function $g(a_i, a_j)$ for all node pairs $(i, j)$.

In addition, we make the following assumptions:

**Intrinsic Activity Decay.** Without external incentives or without positive influence from their social connections, each user has a tendency to slowly reduce her activity, e.g. people slowly lose interest to participate in collaborative networks or exhaust their resources. An observation that specifically reflects this inherent exhaust of activity over time has been made by Danescu-Niculescu-Mizil et al. [2013] for different online communities. We model this situation by using a linear function for $f(a_i)$:

$$ f(a_i) = -\lambda a_i, \lambda > 0 $$

We call parameter $\lambda$ the *Activity Decay Rate*—the rate at which users reduce their activity per unit time, given a complete absence of other (positive) incentives. The specific form of $f(a_i)$ results in an exponential decay ($a_i(t) = a_i(t_0)e^{-\lambda t}$, with $a_i(t_0)$ being the initial activity of node $i$ at time $t_0$) of activity without any external influence. Thus, without other positive impulses the activity of every user will decay over time (see Figure 2(a)).

**Positive Peer Influence.** People tend to copy their friends Christakis and Fowler [2008], Aral and Walker [2012], Wagner et al. [2012], i.e. if neighbors of a node $i$ are active they will positively influence node $i$ to become active as well. The magnitude of the influence, or the “speed” at which the influence is
Intrinsic Activity Decay is the rate at which users reduce their activity per unit time and is represented as a linear function in the form of $f(a) = -\lambda a$, which results in an exponential decay in activity which converges to zero. Extrinsic Positive Peer Influence describes to what extent a user is influenced by the actions taken by her peers and is represented as a monotonically increasing function of a user's activity in the form of $g(a) = qa\sqrt{a^2 + a^2}$. It naturally saturates at Maximum Peer Activity Flow $q$ as activity reaches infinity and, in our simulations, can never be negative per definition (see Equation 3). When the user activity passes the point of the Critical Activity Threshold $a_c$, peer influence gains notable weight and influences neighbors to “do something” (become active). Note that (theoretically) activity for a user can never become negative.

Transferred from an active node to its neighbors will depend on two quantities (cf. Figure 2):

(i) Critical Activity Threshold $a_c$, which represents a soft threshold of activity that marks the point when a user has an activity potential, that notably exercises influence on her peers. Note that influence is exercised at all levels of $a_c$. However, once $a_c$ is reached, the influence is determined as “notable” (e.g., a level of activity that is above the average activity per user) for the corresponding peers. Hence, this critical level of activity is a system-dependent quantity. One can imagine that in a system with high user activity (e.g., a large number of changes per user) the critical activity is higher than in a system with lower levels.
of activity. For example, in the latter case the users will sooner notice a neighbor who became active recently. We model the Critical Activity Threshold as a continuous threshold. Meaning that an active user will always influence her neighbors, but will exercise more influence after she has passed the critical level of activity.

(ii) Maximum Peer Activity Flow $q$ represents the maximum activity flow per unit time from a user to each of her neighbors. This maximum flow is reached as user activity approaches infinity. However, substantial amounts of the maximum flow are already reached whenever the user activity passes the level of the critical activity $a_c$.

Thus, to model peer influence we resort to a monotonically increasing function, where neighbors who are more active are always more influential than less active ones. Additionally, the function $g(a_j)$ saturates for sufficiently large values of activity inducing a natural limit on how much a user can be influenced by her neighbors. We model this by setting $g(a_i, a_j) = g(a_j)$ and choosing an algebraic sigmoid function with:

$$g(a_j) = \frac{qa}{\sqrt{a_c^2 + a_j^2}}, \quad q, a_c > 0. \quad (3)$$

Peer influence can also be analyzed in terms of the growth rate of $g(a)$, i.e. the derivative $\frac{dg}{da}$ of the function $g(a)$. After simplifying and rearranging, the growth rate can be calculated as:

$$\frac{dg}{da} = \frac{qa^2}{(a_c^2 + a_j^2)^{3/2}}. \quad (4)$$

In the limit of large activity $a$ the derivative of $g(a)$ tends to zero, thus peer influence saturates at $q$. On the other hand, the maximum change in influence is observed when $a = 0$ – a neighbor who becomes suddenly active will catch the most attention of her peers.

### 2.1 Dynamics Equation

With $f(a_i)$ and $g(a_j)$ defined, the activity dynamics equation becomes:

$$\frac{da_i}{dt} = -\lambda a_i + \sum_j A_{ij} \frac{qa}{\sqrt{a_c^2 + a_j^2}}. \quad (5)$$
The different parameters of the equation have dimensions, i.e. \( a_i \) and \( a_c \) have activity as unit, \( t \) has second as unit, \( \lambda \) is a rate and has inverse second as unit, and \( q \) has activity per second as unit. Further, the equation has three free parameters, which span a huge parameter space that is difficult to explore in detail. Therefore, our first step is to simplify the equation and express it in a dimensionless form, which typically also has a smaller number of parameters as only their relative ratios are of importance in that specific form. Another advantageous side-effect of a dimensionless formulation is that it eliminates the absolute values of the properties under investigation, in our case user activity, which can be difficult to interpret. Since dimensionless formulations rely on relative ratios, we also develop a notion of small, typical, or large values for activity by comparing them to 1. Thus, small activity values are \( \ll 1 \), typical values of activity are of order 1, and large activity values are \( \gg 1 \).

There are many ways to eliminate dimensions from such equations [Lin and Segel 1988]. A useful heuristic is to try to first eliminate the dimensions from the most non-linear term in the equation, which in our case is \( g(a_j) \). Thus, we begin by defining a relative activity \( x \) as the ratio between the activity \( a \) and the critical activity \( a_c \):

\[
x = \frac{a}{a_c}.
\]  

(6)

The variable \( x \) is dimensionless now, and it is easy to interpret. For example, the fact that \( x = 5 \) means that the user exercises a strong influence on her neighbors since the level of her activity is five times the critical activity \( a_c \). In fact, the influence in this case is \( g(5a_c) = \frac{5q}{\sqrt{26}} \approx 0.98q \). On the other hand if \( x \ll 1 \), e.g. \( x = 0.1 \) this then means that the influence of a user on her neighbors is much smaller, e.g. \( g(0.1a_c) = \frac{0.1q}{\sqrt{1.01}} \approx 0.1q \).

By rearranging we get:

\[
a = a_c x,
\]  

(7)

\[
da = a_c dx.
\]  

(8)
Then by substituting and simplifying ($a_c$ cancels in the second term):

\[
a_c \frac{dx_i}{dt} = -\lambda a_c x_i + \sum_j A_{ij} \frac{q a_c x_j}{\sqrt{a_c^2 + (a_c x_j)^2}}
\]

\[
= -\lambda a_c x_i + \sum_j A_{ij} \frac{q a_c x_j}{a_c \sqrt{1 + x_j^2}}
\]

\[= -\lambda a_c x_i + \sum_j A_{ij} \frac{q x_j}{\sqrt{1 + x_j^2}}. \tag{9} \]

To eliminate the dimensions from the second term we divide both sides with $q$:

\[
a_c \frac{dx_i}{q dt} = -\lambda a_c x_i + \sum_j A_{ij} x_j \sqrt{1 + x_j^2}. \tag{10} \]

The term $\frac{q}{a_c}$ is the growth rate of the function $g(a)$ evaluated at zero:

\[
\frac{dq}{da} \bigg|_{a=0} = \frac{q a_c^2}{(a_c^2 + a^2)^{3/2}} \bigg|_{a=0} = \frac{q}{a_c}. \tag{11} \]

This quantity gives the rate at which the influence on the peers grows if the user activity experiences a small displacement from the point of zero activity. Let us now define this quantity as Peer Influence Growth Rate and denote it with $\mu = \frac{q}{a_c}$ since this will simplify the algebra and will make the model interpretation more intuitive. Thus, the last equation can then be written as:

\[
\frac{1}{\mu} \frac{dx_i}{dt} = -\lambda \frac{\mu}{\mu} x_i + \sum_j A_{ij} \frac{x_j}{\sqrt{1 + x_j^2}}. \tag{12} \]

Finally, we also want to scale time $t$ and express the equation in terms of dimensionless time $\tau$. This last reformulation will further simplify the equation and allows us to interpret and compare activity dynamics over time across various systems. The latter is possible due to the usage of dimensionless time $\tau$ to scale and compare the time evolution of different systems relative to each other. Let us make the following substitution:

\[
\tau = \mu t. \tag{13} \]
We then have:

\[ d\tau = \mu dt, \quad (14) \]

\[ \frac{1}{d\tau} = \frac{1}{\mu} \frac{1}{dt}. \quad (15) \]

By substituting the last term on the left hand side in Equation 12 we arrive at the dimensionless dynamics equation:

\[ \frac{dx_i}{d\tau} = -\frac{\lambda}{\mu} x_i + \sum_j A_{ij} \frac{x_j}{\sqrt{1 + x_j^2}}. \quad (16) \]

Now, there is only one parameter in our dynamics equation, namely the ratio \( \frac{\lambda}{\mu} \). This is a dimensionless ratio of two rates: (i) The Activity Decay Rate \( \lambda \), which is the rate at which a user loses activity, and (ii) the Peer Influence Growth Rate \( \mu \), which is the rate at which a user gains activity due to the influence of a single neighbor.

The ratio between those two rates is the ratio of how much faster a user loses activity due to the decay of intrinsic activity (or interest) than she can gain due to positive peer influence of a single neighbor. For example, a ratio of \( \frac{\lambda}{\mu} = 100 \) would mean that the user intrinsically loses activity 100 times faster than she potentially can get back from one of her neighbors. If we would set \( \frac{\lambda}{\mu} = 1 \), it would mean that users would lose activity as fast as they could regain it from one of their peers. For a short description of all parameters of the activity dynamics model see Table 1.

### 2.2 Linear Stability Analysis

In general, Equation (16) is a coupled set of \( n \) (\( n \) being the number of nodes or users in the network) non-linear differential equations, for which, in a typical case, no closed form solution can be found. Therefore, we turn our attention to the properties of so-called fixed points. A fixed point \( x^* \) represents all the values for \( x_i^* \) for which the system does not change in time:

\[ \frac{dx_i}{d\tau} = -\frac{\lambda}{\mu} x_i + \sum_j A_{ij} \frac{x_j}{\sqrt{1 + x_j^2}} = 0, \forall i. \quad (17) \]

Suppose that we are able to find a fixed point \( x^* \) by solving Equation (17). One obvious fixed point in our model is \( x^* = 0 \), meaning that \( x_i^* \) has
the same value for every \( i \): \( x_i^* = x^* = 0 \), representing a simple special case: a symmetric fixed point. We can easily check that \( x^* = 0 \) is indeed a fixed point since \( f(x^*) = g(x^*) = 0 \), and this also gives \( f(x^*) + \sum_j A_{ij} g(x^*) = 0, \forall i \).

This fixed point also has a particular interpretation in our model. At this fixed point all users have zero activity, which means that they are completely inactive and the system is in an inactive or “dead” state. If the system is in such a state where all nodes are inactive and no external incentives are provided, nothing will ever change and the system will remain inactive indefinitely.

Typically, we are interested in the implications to the system if we provide a small enough impulse to leave such a steady (inactive) state. In our context, the most interesting question is if the system will move from the inactive state towards a state of lively activity or if it will just revert to an inactive state. Technically, we are interested in the stability of a fixed point. In particular, we want to know if the fixed point is attracting (meaning that the system’s activity in the proximity of the fixed point will be attracted to it) or repelling (meaning that the system’s activity close to the fixed point will be pushed away from it).

Table 1: Model and model parameters. The activity dynamics equation is in a dimensionless form and scales over relative time \( \tau \). All properties, as well as the single parameter of the model, are briefly described under Properties and Parameters.

| Equation | Name |
|----------|------|
| \( \frac{dx_i}{d\tau} = -\lambda x_i + \sum_j A_{ij} \frac{x_j}{\sqrt{1 + x_j^2}} \) | Activity Dynamics Equation |

| Properties | Name |
|-----------|------|
| \( \lambda \) | Activity Decay Rate |
| \( q \) | Maximum Peer Activity Flow |
| \( a_c \) | Critical Activity Threshold |
| \( \frac{q}{a_c} \) | Peer Influence Growth Rate |
| \( \tau \) | Relative Time Scale |

| Parameter | Name |
|-----------|------|
| \( \frac{\lambda}{\mu} \) | The ratio, describing how fast a user intrinsically loses activity compared to how fast she gets it back from her neighbors. |
To answer this question we linearize the functions in the proximity of a fixed point. We represent the value of $x_i$ close to the fixed point with $x_i = x^* + \epsilon_i$, where $\epsilon_i$ is sufficiently small. To simplify the calculations, we concentrate on the case of a symmetric fixed point, such as $x^* = 0$. Next, we perform a Taylor expansion about the fixed point and linearize by neglecting the terms of second and higher orders. After simplification we obtain (for details see e.g. Newman [2010]):

$$\frac{d\epsilon_i}{d\tau} = -\frac{\lambda}{\mu} \epsilon_i + \sum_j A_{ij} \epsilon_j,$$

(18)

where $\epsilon_i$ is the displacement of $x_i$ from the fixed point $x^*$.

We can also write Equation 18 in matrix form, which gives:

$$\frac{d\mathbf{\epsilon}}{d\tau} = (-\frac{\lambda}{\mu} \mathbf{I} + \mathbf{A}) \mathbf{\epsilon},$$

(19)

where $\mathbf{I}$ is the identity matrix and $\mathbf{A}$ is the adjacency matrix.

We can solve the last equation by writing $\mathbf{\epsilon}$ as a linear combination of eigenvectors $\mathbf{v}_r$ of the symmetric real matrix $(-\frac{\lambda}{\mu} \mathbf{I} + \mathbf{A})$:

$$\mathbf{\epsilon}(\tau) = \sum_r c_r(\tau) \mathbf{v}_r.$$

(20)

Equation 19 then becomes:

$$\sum_r \frac{dc_r}{d\tau} \mathbf{v}_r = (-\frac{\lambda}{\mu} \mathbf{I} + \mathbf{A}) \sum_r c_r(\tau) \mathbf{v}_r = \sum_r c_r(\tau) (-\frac{\lambda}{\mu} + \kappa_r) \mathbf{v}_r,$$

(21)

where $\kappa_r$ are the eigenvalues of the graph adjacency matrix $\mathbf{A}$. We also used the fact that the matrix $(-\frac{\lambda}{\mu} \mathbf{I} + \mathbf{A})$ has the same eigenvectors as $\mathbf{A}$, but with the eigenvalues $-\frac{\lambda}{\mu} + \kappa_r$.

The solution of the last equation for the coefficients of the linear combination is then:

$$\frac{dc_r}{d\tau} = (-\frac{\lambda}{\mu} + \kappa_r)c_r(\tau) \implies c_r(\tau) = c_r(t_0)e^{(-\frac{\lambda}{\mu} + \kappa_r)\tau}.$$

(22)

Now, the displacement from the fixed point will decay in time towards 0 if the exponents for the coefficients $c_r(\tau)$ are all negative. Thus, we arrive at
the master stability equation for the special case of a dynamical system that we defined as:

\[-\frac{\lambda}{\mu} + \kappa_r < 0, \quad \forall r, \tag{23}\]

Since the adjacency matrix has both positive and negative values, a necessary stability condition is \(\frac{\lambda}{\mu} > 0\) (which is by definition satisfied). If this condition is satisfied we can rearrange Equation 23 and obtain the following inequality:

\[\kappa_1 < \frac{\lambda}{\mu}. \tag{24}\]

where \(\kappa_1\) is the largest positive eigenvalue of the graph adjacency matrix. Note that this inequality separates the network structure (\(\kappa_1\)) from the activity dynamics (\(\frac{\lambda}{\mu}\)).

If this stability condition is satisfied, the fixed point \(x^* = 0\), in which there is no activity at all (“inactive” system), will be a stable fixed point. This also means that small changes in activity only cause the system to momentarily leave the (attracting) fixed point until it becomes inactive again. We initialized Zachary’s Karate Club Network (cf. Figure 3(a) and 3(b)) with random activities between 0 and 0.1 per node and calculated activity dynamics. If the master stability equation holds (Figure 3(c)), activity converges towards zero. However, when invalidating the master stability equation (Figure 3(d)), activity converges to a new and permanently active fixed point.

**Inactive system analysis.** We are not satisfied with an inactive system and hence, aim at making it unstable in order to permanently leave this attracting state of complete inactivity. In particular, there are two possibilities to leave the fixed point with zero activity:

(i) We provide an external impulse to the system, e.g. in the form of incentives for users to increase their activity, pushing the system far away from the fixed point of no activity (and hope that it will be attracted by another fixed point where the activity is not zero).

(ii) We compromise the stability condition by either manipulating:

   (a) the network structure (i.e., making \(\kappa_1\) larger) or
   (b) the activity dynamics (i.e., making \(\frac{\lambda}{\mu}\) smaller).
Figure 3: **Illustrative toy example.** Top Left (a): Visualization of Zachary’s Karate Club using a force-directed layout. The size and color of a node represent random (arbitrary) activity values between 0.0 and 0.1 of the corresponding nodes (bigger and darker equals higher values). **Top Right (b):** Eigenvalue spectrum of Zachary’s Karate Club network. The x-axis corresponds to the real part of eigenvalues and the y-axis represents the imaginary part of the eigenvalues. The highest eigenvalue is 6.726. **Bottom (c and d):** Evolution of activity with random initial activities (average over 10 runs; weights between 0.0 and 0.1 for each node). **Bottom Left (c):** Activity dynamics with parameters satisfying the master stability condition $\kappa_1 < \frac{\lambda}{\mu}$. Each line represents one node and the activities converge to the state of zero activity. **Bottom Right (d):** Once the activity dynamic parameters invalidate the master stability condition $\kappa_1 < \frac{\lambda}{\mu}$, activity converges towards a new and permanently active fixed point.
Structurally, we can manipulate the size of \( \kappa_1 \) by creating or removing links (and nodes) in our network. Dynamically, \( \frac{\lambda}{\mu} \) becomes smaller if either \( \lambda \) becomes smaller, meaning that the intrinsic user activity decays at a slower rate or \( \mu \) becomes larger, meaning that people copy their friends more and faster, or both.

In this paper, we focus on applying and fitting our proposed activity dynamics model to empirical datasets and discuss the obtained results and findings. At this time, we leave the investigation of the manipulation of the activity dynamics ratio \( \frac{\lambda}{\mu} \) as well as the manipulation of the network structure to invalidate the master stability equation open for future work.

3 Datasets and Experimental Setup

We are now interested in modeling and simulating activity dynamics for empirical datasets. In particular, we investigate activity dynamics for an array of different websites, consisting of instances of the StackExchange\(^1\) network as well as multiple Semantic MediaWiki\(^2\).

First, we continue by characterizing the different investigated datasets and outline our methods for the empirical estimation of the required parameters (see Table \( \text{Table 1} \)). We then fit our model to the collaboration networks and present the results of the activity dynamics simulation.

3.1 Datasets

We selected a total of four differently sized instances from the StackExchange network as well as four different Semantic MediaWiki instances to model activity dynamics (see Figure \( \text{Figure 1} \)). In particular, we concentrate our efforts on the StackOverflow StackExchange\(^3\) (SO), which also represents our largest dataset; a website for software developers to ask and discuss programming and computer related questions and topics. The English Language & Usage StackExchange\(^4\) (ESE) as well as the Mathematics StackExchange\(^5\)

\[^1\text{http://www.stackexchange.org/sites}\]
\[^2\text{http://www.semantic-mediawiki.org}\]
\[^3\text{http://www.stackoverflow.com}\]
\[^4\text{http://english.stackexchange.com}\]
\[^5\text{http://mathematics.stackexchange.com}\]
(a) StackOverflow StackExchange
(b) English Language & Use StackExchange
(c) Mathematics StackExchange
(d) History StackExchange

(e) 15Mpedia Wiki
(f) Nobbz Wiki
(g) Beachapedia Wiki
(h) CharacterDB Wiki

Figure 4: **Empirical Collaboration Networks.** Visualization of all investigated collaboration networks. Nodes represent users, edges represent collaboration between users. The node size (diameter) represents the activity weights after the first month according to Equation 33. The **top row** (a to d) depicts the different StackExchange collaboration networks, while the **bottom row** (e to h) shows the collaboration network visualizations for the different Semantic MediaWiki instances.

(MATHSE) represent two medium-sized websites and are platforms for asking and discussing questions related to the English language and mathematics respectively. Similarly, the History StackExchange (HSE), which is the smallest of the StackExchange datasets that we investigate, allows users to discuss topics and questions related to history and historical events.

We further investigate activity dynamics for the 15Mpedia Wiki (15MW)—a large Spanish Semantic MediaWiki instance that discusses a wide variety

---

6http://history.stackexchange.com
7http://wiki.15m.cc/wiki/Portada
of topics related to Spain and its different areas and regions. We have also included the medium-sized german Nobbz Wiki\(^8\) (NZ)—a structured knowledge base and discussion platform for the online game “Die Verdammten”\(^9\). The Beachapedia Wiki\(^10\) (BP) and the CharacterDB Wiki\(^11\) (CDB) are the smallest datasets in our activity dynamics analysis and strive to create structured knowledge bases for various topics on beaches in the United States and semantic data on the structure of Chinese characters respectively.

The investigated datasets are very diverse in their characteristics, for example, the number of registered users ranges from 1.8 million in SO to a total of 22 in CDB. For more detailed information see Table 2.

From each of these datasets, we extracted a collaboration network for fitting the model and simulating activity dynamics. Hence, we first parsed the change-logs for all datasets. Each user, who has contributed at least one question, answer or comment for the StackExchange datasets, or created or edited an article for the Semantic MediaWikis is represented as a node in the corresponding collaboration network. Edges between users represent collaboration and are undirected. For the StackExchange dataset, we defined collaboration as either posting an answer to a question or posting a comment on the initial question or an answer. For the Semantic MediaWiki instances, we have created an edge between users, who (chronologically and) successively changed the same article (cf. Figure 5). Edges with the same source and target user have been removed in all datasets.

Hence, the resulting collaboration networks consist of as many nodes as there are active users in the dataset. Following from our definition of collaboration (and thus edges), the total number of edges is limited (max) by the number of interactions. Note that for the activity dynamics analysis we assume that the structure of our collaboration networks never changes, meaning that all users are already present in the network from its inception and will not change (e.g., adding or removing users) over time.

### 3.2 Simulating Activity Dynamics

For the simulation of activity dynamics we apply Euler’s method for solving differential equations computationally. Thus, we approximate the time

---

\(^8\)http://nobbz.de/wiki
\(^9\)http://www.dieverdammten.de/
\(^10\)http://www.beachapedia.org
\(^11\)http://characterdb.cjklib.org (offline)
Figure 5: **Collaboration Network Construction.** This plot depicts the different elements of the StackExchange and Semantic MediaWiki datasets that have been classified as posts and replies (cf. Table 2) as well as the edges that have been drawn between certain entities and change-actions and represent collaboration in our collaboration networks.

The evolution of $x_i$ by iterating:

$$x_i(k+1) = x_i(k) + \left[ -\frac{\lambda}{\mu} x_i(k) + \sum_j A_{ij} \frac{x_j(k)}{\sqrt{1 + x_j(k)^2}} \right] \Delta \tau \tag{25}$$

Table 2: **Dataset statistics.** Note that all datasets differ in size, activity and observation periods. Users refers to the number of unique users that have contributed at least one post or reply to the corresponding datasets. Posts represent newly created questions in the case of the StackExchange network and newly created articles in the case of the Semantic MediaWiki datasets. Replies are either comments or answers for all StackExchange datasets and edits of existing articles for Semantic MediaWikis. $\kappa_1$ denotes the largest eigenvalue of the corresponding collaboration network.
The local approximation error for the Euler’s method is of the order $O(\Delta \tau^2)$ and the global of the order $O(\Delta \tau)$. In all our experiments involving synthetic datasets we iterate until we converge to a fixed point, which we determine by the sum of the square difference between two iteration steps—we stop the process when this sum falls beneath $10^{-16}$. For experiments involving empirical datasets we iterate for a specified number of iterations until we reach a given point in relative time $\tau$.

4 Model Fitting and Empirical Illustration

After describing the activity dynamics model, its parameters and our empirical datasets, we are interested in fitting and applying our model to empirical datasets. To that end, we first describe the necessary steps to estimate the required parameters in Section 4.1 and present the results of our simulations on empirical datasets in Section 4.2. The framework, used to estimate empirical parameters and run the activity dynamics simulation, is available on GitHub.

4.1 Fitting the Model and Initializing Simulations

As an illustration of our model and its applicability we turn our attention to the empirical estimation of the single model parameter, i.e. the ratio $\frac{\lambda}{\mu}$. After estimating this ratio we can simulate activity dynamics in our datasets.

Estimating $\frac{\lambda}{\mu}$. In a first step, for simplicity reasons, we set the absolute time scale to an arbitrary bin of one month. We then proceed by estimating Critical Activity Threshold $a_c$ and Maximum Peer Activity Flow $q$ from the data to obtain the Peer Influence Growth Rate $\mu$ (the rate at which a user gains activity due to the influence of a single neighbor) as $\mu = \frac{q}{a_c}$. This will allow us to calculate the relative time scale $\tau$ for each dataset as $\tau = \mu t$ (see also Equation 13). Additionally, we denote the number of posts per month $t$ with $p(t)$, the number of replies in month $t$ with $r(t)$, and the total time in months with $T$.

We estimate the Critical Activity Threshold $a_c$ by calculating the average activity per month per user:

\[ a_c = \frac{1}{n} \sum_{i=1}^{n} \frac{p(t_i)}{T} \]

\[ \text{where } p(t) \text{ is the number of posts in month } t \text{ and } T \text{ is the total time in months} \]

For experiments involving empirical datasets we iterate for a specified number of iterations until we reach a given point in relative time $\tau$. 

\[^{12}\text{https://github.com/cheza/ActivityDynamics}^\]
\[ a_c = \frac{\sum_{t=1}^{T} (p(t) + r(t))}{nT}, \]  
(26)

with \( n \) being the number of users as before. The intuition here is that users who are substantially more active than an average user will also exercise a significant influence on their neighbors, whereas users whose activity is below average will not influence their neighbors to a great extent.

In the next step we need to estimate Maximum Peer Activity Flow \( q \). Since \( q \) represents the maximum peer activity flow per time unit, the model has to be parametrized in such a way that the maximum activity flow that we observe in our data is possible to be achieved by the model. In other words, the sum of the activity flow alongside all the links in the model should be greater or at least equal to the maximum observed activity in a month. We can further deepen our argument by distinguishing between the initializing user activity and peer induced activity in the system, i.e. we can think about the replies (peer influenced activity) as reactions to posts (initializing user activity). Thus, what we actually require from \( q \) is to be able to support the flow of peer induced activity (replies), i.e. the sum of the activity flow alongside all the links in the model should be greater or equal to the maximum observed replies in a month, which have been caused by the posts in that month. One could argue that the posts from the previous month should be taken as the initializing activity in the current month, however, as we did not observe any significant differences in the estimated parameters we use the number of posts of the current month for simplicity reasons. Formally, with:

\[ t_{\text{max}} = \arg\max_t r(t), \]  
(27)

we require \( q \) such that:

\[ r(t_{\text{max}}) \leq \sum_{ij} A_{ij} \frac{q u(t_{\text{max}}) p(t_{\text{max}})}{\sqrt{a_c^2 + \left(\frac{p(t_{\text{max}})}{u(t_{\text{max}})}\right)^2}}, \]  
(28)

where \( u(t_{\text{max}}) \) is the number of active users in month \( t_{\text{max}} \), and hence \( \frac{p(t_{\text{max}})}{u(t_{\text{max}})} \) is the average number of posts in month \( t_{\text{max}} \). We could have calculated the average also by dividing by the total number of users \( n \), implying that the maximum number of observed replies has been caused by all users in the system. However, this would overestimate \( q \) and could lead to less precise
simulation results, since the simulation assumes that all nodes are active at all times.

Taking into account that \( \sum_{ij} A_{ij} = 2m \) we can simplify the last equation to obtain:

\[
r(t_{\text{max}}) \leq 2m \frac{q p(t_{\text{max}})}{\sqrt{a_c^2 + \left( \frac{p(t_{\text{max}})}{u(t_{\text{max}})} \right)^2}}.
\]  

(29)

At this point we can reason about the capacity utilization of the system, describing the system load during the maximum observed activity. Note that in reality, the system load depends on multiple factors, such as the user interface, the structure of the collaboration network and even the motivations and characteristics of every single user. For simplicity reason we assume that the system had a 100% capacity utilization (technically we replace less or equal relation with the equality relation). Finally, after rearranging we arrive at the expression for calculating \( q \):

\[
q = r(t_{\text{max}}) \sqrt{a_c^2 + \left( \frac{p(t_{\text{max}})}{u(t_{\text{max}})} \right)^2}.
\]

(30)

In the next step, we can calculate \( \mu = \frac{a_c}{a_c} \) and \( \tau = \mu t \) to obtain the relative time scale. It turns out, that the numerical solution to this equation is rather unstable, which is why we use a simplified and more practical approximation for estimating the ratio. First, we assume that at the end of each month the system has arrived at a fixed point, i.e. it holds that \( \frac{dx_i}{dt} = 0 \) for all \( i \). Then the Taylor expansion simplifies to the case that we introduced in the Section 2.2, and thus the final expression is as follows.

\[
\frac{\lambda}{\mu} = \kappa_1 - \frac{1}{\mu} \log \frac{x(t+1)}{x(t)}
\]

(31)

The gradient of \( g(x) \) is maximized at a fixed point and therefore the previous expression will overestimate the ratio. Hence, we have to correct the ratio for a small amount after our empirical estimation of the required parameters. In our datasets this correction proves to work well:

\[
\frac{\lambda}{\mu} = \frac{\lambda}{\mu}(t) - c \frac{x(t)}{n}.
\]

(32)
We have set $c$ at 0.03. However, the value might have to be adjusted for certain datasets, although it proved to work well at 0.03 for the StackExchange and Semantic MediaWiki datasets presented in this paper.

**Initializing simulations.** We now have to determine the number of iterations for the simulation of the calculated ratio $\frac{\lambda}{\mu}(t)$ by setting $\Delta \tau$ in relation to our relative time scale $\tau$ (which by setting $t = 1$ month is calculated as $\tau = \mu$). Thus, we want to choose an optimal $\Delta \tau$ such that we balance the efficiency of the computation and the precision of the simulation. If we set it too high—meaning that the calculations are less computationally intensive, as we have to run a smaller number of iterations—the accuracy of our simulation will decline, as the potential error per iteration, due to our approximations, becomes higher. This error can become so large that it could potentially lead to numerical instability, meaning that the overall activity in a system can become negative, which might result in activity to diverge towards $\pm \infty$. With certain combinations of the network structure, $\Delta \tau$ and the calculated ratios, activity can become negative without diverging, oscillating around the fixed point of zero activity until convergence. In contrast, if we set $\Delta \tau$ too low we end up with a very precise simulation, although the time necessary to compute the simulation will be much higher, as a much larger number of iterations will have to be executed. Thus, we set $\Delta \tau = 0.001$ for all datasets and report the other estimated empirical parameters in Table 3.

To initialize the networks with activity we extract $x(1)$, the activity over $a_c$, in the first month, normalized over the total number of users in the collaboration network and $2m$. Finally, we multiply the resulting value by $k_i$, the degree of the user in the network (see Equation 33).

$$x_i(1) = \frac{x(1)}{2mn}k_i$$

(33)

With $k_i$ being defined as

Table 3: **Parameter estimation.** A listing of all estimated parameters for the activity dynamics simulation. $\Delta \tau$ has been manually set to 0.001 for all datasets. The remainder of the listed parameters are calculated according to their corresponding formulas, as described in Section 4.1.
\[ k_i = \sum_{j=1}^{n} A_{ij} \]  

(34)

Note that we normalize the observed activity over the total number of users in the collaboration network and assign each user an initial activity value that corresponds to its degree to mitigate the bias of our static network structure and the burn-in phase from our activity dynamics calculations. In particular, nodes with higher degrees will receive (and retain) higher levels of activity over the course of our activity dynamics simulation. When initializing the networks without considering the degree, we will observe unnatural spikes in activity in the beginning of our simulation, until the activity levels have adjusted themselves to reflect the degrees of the corresponding nodes.

4.2 Empirical Illustration

After calculating \( \lambda \) and setting \( \Delta \tau \) we simulate activity in our collaboration networks. Due to our chosen approximations, the main goal of the presented illustration is neither to predict activity in collaboration networks, nor is it to accurately determine the absolute number of posts and replies for each month. Rather, we are interested in demonstrating that our assumptions regarding the Activity Decay Rate and the Peer Influence Growth Rate hold and allow us to simulate trends in activity dynamics for given and real values. Further, by modeling and simulating activity dynamics for empirical datasets we not only deepen our understanding of the model but we also—depending on the values of the parameters—potentially obtain new insights into the systems under investigation.

Figure 6 depicts the results of the activity dynamics simulation for the four StackExchange datasets and the four Semantic MediaWiki instances. The simulation (left \( y \)-axes) and the real activities (right \( y \)-axes) over time (in months; \( x \)-axes) are plotted on different scales as our simulation calculates activity over \( a_c \) and is mainly intended for illustrating the trends in activity dynamics for each dataset.

Overall, the results gathered from the activity dynamics simulation exhibit a notable resemblance to the real activities of the corresponding datasets. Due to different scales and the already discussed bias introduced by the chosen approximations and simplifications when estimating the empirical parameters for our model (i.e., static network structure and average model
parameters over all months), the simulated activity is naturally limited in its accuracy. These limitations are particularly visible whenever there are large and sudden increases of activity in the collaboration networks, as our model is only able to compensate these activity trends for a fraction of what has been observed in the real data of the collaboration networks. This is due to only considering the coefficient of the first eigenvector when estimating the empirical parameters (see Section 2.2 and Equation 31). However, such drastic increases in activity are likely to trigger the influence of multiple
(positive) eigenvectors, adding activity to the newly established fixed point, after invalidated the master stability equation, resulting in potentially larger increases in activity when necessary. Further, the assumption of a fixed network structure of our investigated collaboration networks also (negatively) influences the obtained results of our simulation. For example, it is possible for our simulation to yield higher increases in activity (see Figures 6(c) and 6(h)), especially during the first few \( \tau \) (months) of the simulations as users might be influenced by peers, who would join the collaboration network only at a much later point in time.

5 System Mass and Activity Momentum

We can further analyze the obtained ratios and parameters of our activity dynamics simulation to broaden our understanding of the collaboration networks under investigation. Figure 7 depicts the value of the calculated ratios \( \frac{\lambda}{\mu} \) (y-axis) for each month (x-axis) in relation to \( \kappa_1 \) (the largest eigenvalue; represented as dotted, horizontal line). Whenever the ratio is higher than \( \kappa_1 \) (above the dotted line), our master stability equation holds and the system inevitably loses activity. The amount of activity that is lost per iteration—and hence the speed of activity loss—is proportional to the value of the ratio and the activity already present in the network. In general, a higher ratio results in a higher and faster loss of activity.

If the ratio is smaller than \( \kappa_1 \) (below the dotted line), the master stability equation has been invalidated and the system will converge to a new fixed point of immanent activity (cf. Section 2.2). If this is the case, we can observe one of three potential behaviors, which are triggered depending on the amount of activity already present in the network and the current ratio:

(i) **An increase in activity** if the new fixed point, corresponding to the new ratio, is of higher overall activity than the activity already present in the collaboration network (see \( \tau = 14 - 16 \) in Figures 6(b) and 7(b)). This situation emerges whenever we invalidate the master stability equation from a previously stable fixed point or if the system is already stable in a situation when the new ratio is smaller than the last estimated ratio.

(ii) **A decrease in activity** if the new fixed point is of lower overall activity than the activity already present in the collaboration network
Figure 7: Evolution of ratios $\lambda/\mu$. The evolution of the ratios $\lambda/\mu$ (y-axes) over $\tau$ (in months; x-axes) for the StackExchange datasets (top row (a to d)) and for the Semantic MediaWiki instances (bottom row (e to h)). The corresponding largest eigenvalue $\kappa_1$ of each collaboration network is represented as horizontal dotted line. After the first few months, the ratios start to stabilize around $\kappa_1$, indicating that the activities of the systems are less influenced by the activity of single individuals than they are by peer influence.

(see $\tau$ 18 – 20 in Figures 6(g) and 7(g)). Again this may occur in two specific situations. Firstly, if the ratio increases, so that the master stability equation is now satisfied and the system has been previously in an unstable state. Secondly, if the system is in an unstable state but the ratio increases slightly without satisfying the stability equation.

(iii) **No change in activity** if the new fixed point corresponding to the new ratio is of the same overall activity than the activity already present in the collaboration network (see $\tau$ 27 – 35 in Figures 6(h) and 7(h)).
**System Mass.** We can now use the obtained ratios to characterize the collaboration networks and quantify their robustness in terms of their activity dynamics. Robust systems are systems with lively and high levels of activity, which are able to keep that activity even in the cases of small unfavorable changes in the dynamical parameters. Less robust systems are systems that lose their activity very quickly as a consequence of even small changes in the ratio. Thus, we calculate the standard deviation over all ratios $\sigma_\mu$ over time $t$ and normalize it over $\kappa_1$, to account for the size of the collaboration networks, and refer to it as $\rho$—the normalized standard deviation of the ratio $\lambda/\mu$ (see Equation 35).

$$\rho = \frac{\sigma_\mu}{\kappa_1} \quad (35)$$

The normalized standard deviation is a measure of system sensitivity and its inverse ($\frac{1}{\rho}$) represents a measure of system stability or *inertia* to changes in activity. Analogously to mass in classical mechanics we can call the quantity ($\frac{1}{\rho}$) the *System Mass*. We denote this quantity with $m_s$ with the subscript $s$ to distinguish it from the number of links $m$ in a collaboration network (see Table 4). The intuition for the *System Mass* is the following. In a system with a larger $m_s$ it is more difficult to induce changes in activity. In particular, this means that it is more difficult to reduce activity in an active system, as well as it is difficult to raise the activity level in an inactive system.

**Activity Momentum.** After calculating the *System Mass* $m_s$, we are now interested (again analogously to classical mechanics) in calculating the *Activity Momentum* $p$ for our collaboration networks (see Equation 36).

$$p = m_s a \quad (36)$$

For the activity we take (i) the average activity (posts and replies) per month and (ii) the activity in the last month of our observation periods (cf. Table 4) and calculate (i) the average and (ii) the current momentum.

The higher the *Activity Momentum* of a collaboration network, the more force is needed to “stop” (make it inactive) or “start” (make it active) the system. Hence, the higher the momentum, the more robust a given network. In particular, if a (sufficiently) small number of users would suddenly stop contributing to a collaboration network that exhibits a very large *Activity Momentum* $p$, activity in the overall network would not be notably
influenced. On the other hand, if the same number of users would stop contributing to a collaboration network with a (significantly) smaller Activity Momentum \( p \), chances are that their actions (or lack thereof) will have a notable influence on the overall trends in activity dynamics of the system. In particular, there are three factors that influence the Activity Momentum of collaboration networks:

(i) **The standard deviation of \( \frac{\lambda}{\mu} \).** If the ratio is very stable and does not oscillate too much around \( \kappa_1 \), the standard deviation and hence the normalized standard deviation will be very small. This also means that activity, as well as increases and decreases thereof, is equally distributed across \( \tau \) and is not (frequently) exercised in bursts.

(ii) **The largest eigenvalue \( \kappa_1 \).** Larger and denser collaboration networks exhibit a larger highest eigenvalue \( \kappa_1 \). As \( \rho \) is the normalized variance of the ratios over \( \kappa_1 \), the largest eigenvalue will directly influence \( \rho \). The notion of normalizing \( \rho \) over \( \kappa_1 \) is that large collaboration networks are less likely to exhibit sudden changes in activity than smaller ones.

(iii) **The activity.** The larger the average activity (posts and replies) per month, the higher the Activity Momentum of a collaboration network, and hence the higher the force that is needed to render the collaboration

Table 4: **System Mass and Activity Momentum.** The table depicts the results for the activity momentum analysis. \( \rho \) is the standard deviation of the calculated ratios over \( \kappa_1 \). System Mass is represented by \( \frac{1}{2} \) and Activity Momentum represents System Mass multiplied with Activity. Both columns, Activity and Momentum, depict the averages over all months as well as the values for the last observed months in brackets. SO and MATHSE exhibit the largest average and current Activity Moment, followed by ESE and 15MW. Even though 15MW exhibits a much smaller System Mass than ESE or NZ, its Activity Momentum is much larger than NZ and approximately the same as ESE.

| Dataset | Activity (last month) | \( \rho \) | System Mass | Activity Momentum (last month) |
|---------|----------------------|----------|-------------|-----------------------------|
| SO      | 723,094 (1.22 \( \times \) 10^9) | 0.00115  | 865.93      | 6.26 \( \times \) 10^6 (1.05 \( \times \) 10^6) |
| MATHSE  | 41,199 (76,013)       | 0.00198  | 505.08      | 2.08 \( \times \) 10^6 (3.83 \( \times \) 10^6) |
| ESE     | 7,008 (15,355)        | 0.00259  | 386.16      | 2.70 \( \times \) 10^8 (5.92 \( \times \) 10^8) |
| HSE     | 871 (1,193)           | 0.00733  | 136.51      | 118,913 (162,856)           |
| 15MW    | 9,567 (4,848)         | 0.00364  | 274.80      | 2.62 \( \times \) 10^6 (1.33 \( \times \) 10^6) |
| NZ      | 236 (289)             | 0.00144  | 695         | 164,141 (200,866)           |
| BP      | 514 (252)             | 0.00263  | 44.19       | 22,704 (11,135)             |
| CDB     | 17 (122)              | 0.00596  | 167.73      | 2,770 (20,463)              |

28
network inactive. Analogously, networks with a small Activity Momentum require less force to influenced (i.e., to either speed up/increase or slow down/decrease activity).

Hence, we can use the calculated Activity Momentum $p$ as an indicator of the activity level as well as the tendency of a system to stay at that activity level in the future. For example, SO exhibits the most robust collaboration network of our datasets regarding changes in activity, with an Activity Momentum of orders $10^8$ (average) and $10^9$ (last month). ESE and 15MW both exhibit similar average Activity Momentum of orders $10^6$. However, when looking at the Activity Momentum of the last months, ESE is roughly four times as hard to stop as 15MW.

In contrast, HSE exhibits a larger activity per month than NZ and a lower Activity Momentum, indicating that less force is needed to render HSE inactive than it would be needed to render NZ inactive. MATHSE, BP and CDB follow analogously.

On the other hand, the higher the System Mass of a collaboration network and the lower the corresponding Activity Momentum (e.g., HSE and CDB), the harder it is to accelerate or jump-start the system.

6 Related Work

The work presented in this paper was inspired by and builds upon work presented in the areas of critical mass theory and dynamical systems on networks.

6.1 Critical Mass Theory

In 1985 and 1988, Oliver et al. [1985], Oliver and Marwell [1988], Marwell et al. [1988] 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. Fundamentally, these production functions represent the link between individual benefits and benefits for the group.

They argue that one very important aspect of critical mass is the natural limitation of collective goods for groups such as housing, food, fuel or oil. Hence, the capacity of users (and thus critical mass) for such a group or system is naturally limited by the corresponding resource. However, collective
(digital) goods are not (or only artificially) limited for online communities; theoretically allowing for an infinite increase in users and interest. Without users motivated to contribute, interest will decrease and critical mass will lose momentum and ultimately decelerate until all interest vanishes. In their work they identified multiple different types of production functions, with the most important ones being: Accelerating, decelerating and linear functions. 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, while contributions to linearly growing functions are always worth the same. Until today it is still mostly unclear what these production functions look like for online communities (e.g., StackOverflow) and online production systems (e.g., Semantic MediaWikis).

Depending on the investigated or desired point of view, different characteristics of these communities and online production systems can be used as basis for calculating production functions. The analysis of Oliver et al. [1985] 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.

One approximation for critical mass by Solomon and Wash [2014] involved the investigation of the number of changes – as activity – and number of users – as growth of a community – for calculating production functions for WikiProjects. The authors argue that activity in online production systems, after certain amounts of time, is the best indicator of a self-sustaining system. In this work, we will extend the analysis presented by Solomon and Wash and specifically define the point of when an online system has reached critical mass and has become self-sustaining in terms of its activity dynamics. Walk and Strohmaier [2014] recently conducted a similar analysis to characterize critical mass for Semantic MediaWikis.

Raban et al. [2010] investigated factors that allow for a prediction of
survival rates for IRC channels and identified the production function of these chat channels regarding the number of unique users versus the number of messages posted at certain times, as the best predictor.

Cheng and Bernstein [2014] 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.

We use the notion of critical mass to define the barrier, that has to be overcome, for collaboration networks to become self-sustaining in terms of activity.

6.2 Dynamical Systems on Networks

Dynamical systems in a non-network context are a well-studied scientific and engineering field. Generally, a dynamical system is any system that changes in time, whose behavior is determined by some specific rules or (differential) equations over a set of quantifiable variables. We distinguish between continuous and discrete as well as deterministic and stochastic systems. Strogatz [1994] and Barrat et al. [2008] provide excellent introductions and analyses of dynamical systems.

Different social and economic processes, which take place both offline and online, have been modeled with the use of dynamical systems. In the context of the Web, the primary focus of dynamical systems was set on analyzing and understanding the diffusion of information in online social networks Leskovec et al. [2007, 2009], Myers et al. [2012], Vespignani [2012], including the analysis of online memes and viral marketing.

On the other hand, the Bass Model Bass [1976] describes how novel products are accepted and adopted in a network and has seen a wide variety of applications in different fields of research and also for practical use. The model consists of two parameters, the propensity for innovation and the propensity for imitation. A product will be successfully accepted and adopted by the community, depending in the ratio between these two parameters.

Acerbi et al. [2012] investigated factors that determine how social traits propagate within a specific popularity. Iribarren and Moro [2009] conducted a viral email experiment, allowing them to track the diffusion of information in a social network. They showed that due to heterogeneity in human activity, the most common and simple growth equation from epidemic models is not suitable to model information diffusion in social networks.
Recently, in the context of activity dynamics, Ribeiro [2014] conducted an analysis of the daily number of active users that visit specific websites, fitting a model that allows to predict if a website has reached self-sustainability, defined by the shape of the curve of the daily number of active users over time. He uses two constants $\alpha$ and $\beta$, where $\alpha$ represents the constant rate of active members influencing inactive members to become active. $\beta$ 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.

The model presented in this paper to simulate activity dynamics heavily relies on the concept of dynamical systems on networks. We strongly believe that by modeling and understanding activity dynamics, we will gain a better understanding of the processes involved in and around the concept of peer influence in collaboration networks. Other areas of application for dynamical systems on networks are the modeling and simulation of diseases in the form of epidemic models, and opinions or traits of a person, also known as opinion dynamics.

6.2.1 Epidemic Models

Modeling the outbreak of diseases can be seen as a special case of dynamical systems. At first, epidemic models dealt with the spreading of diseases in social (real life) networks May and Anderson [1984], Hethcote [1978], Anderson and May [1991], Bolker and Grenfell [1993, 1995], Lloyd and May [1996], Keeling and Rohani [2002], Ferguson et al. [2003], ignoring the underlying network aspect, simulating contractions and outbreaks via random encounters of the whole population under investigation. For an exhaustive survey of epidemic models refer to Pastor-Satorras et al. [2014].

Henceforth, these models have been extended to include the structure and other aspects of the underlying networks Rvachev and Longini [1985], Ferguson et al. [2003], Hufnagel et al. [2004], Longini et al. [2005], Ferguson et al. [2005], Colizza et al. [2006], limiting the spread and outbreaks according to different factors. Further, epidemic models were also utilized to simulate the spread for a plethora of properties in different kinds of networks, such as viruses spreading in computer networks Kephart et al. [1993, 1997], Pastor-Satorras and Vespignani [2001a], Aron et al. [2002], Pastor-Satorras [2014].
and Vespignani [2007] and information propagation (e.g., memes) Leskovec et al. [2007] among others.

In general, epidemic models are based on the intuition that a disease propagates through a social network with a given infection rate, defining the probability that a neighbor of an already infected node contracts the disease. Different models have been developed and analyzed to simulate epidemic outbreaks in a population or network Bailey et al. [1975], Anderson and May [1991], Hethcote 2000, Newman [2010], which can only transfer on contact. Typically, such an outbreak is modeled using a small number of possible states for each node and a fixed probability of contraction (e.g., $\beta, \gamma$), which defines the probability or “threshold” that has to be reached for a node to change to a different state. For example, the SI model consists of only two states – susceptible and infected – and one probability parameter $\beta$, that determines when the transition from susceptible to infected is initiated. Note that transitions in the SI model can only occur from susceptible to infected while already infected nodes remain infected indefinitely. As the infection rate is relative to the population under investigation, epidemic simulations with a small number of originally infected hosts usually start-off by slowly contracting the disease until exponential growth is reached. Once the majority of the population carries the disease, the infection process slows down again until the whole population is infected.

A more sophisticated extension to the SI model is the SIR model Anderson and May [1991], Murray [2002], which additionally introduces the recovered (or removed) state as well as an additional parameter $\gamma$ to model the transition from infected to recovered. Again, transitions only occur from susceptible to infected to recovered. As the name suggests, this newly introduced state allows nodes to become immune to the disease and will not be infected in the future, nor be able to infect other nodes. Other models for simulating epidemic outbreaks are the SIS and SIRS models, where the population can recover but does not become immune (SIS) or stays immune but still has a chance to become susceptible for infection again (SIRS) Britton 2010, Dietz 1967.

Since their introduction, epidemic models have seen a wide array of application. For example, to analyze how computer viruses spread Kephart and White [1991, 1993], Newman et al. 2002 or the study of epidemics in complex (scale-free, power-law) networks Pastor-Satorras and Vespignani 2001a,b, Moreno et al. 2002.

Among others Wang et al. 2003 as well as Ganesh et al. 2005 demon-
strated the importance of the networks spectra (eigenvalues and eigenvectors of the network adjacency matrix) for epidemic and dynamical network models [Chung et al. 2003a,b]. We show a similar dependency of activity dynamics on eigenvalues in this paper in Section 2.

6.2.2 Collective Behavior & Opinion Dynamics

Another important field of application of dynamical systems on networks are opinion dynamics. They are used to model collective behavior and influence, usually in the form of a consensus-reaching task, at every point in time. The main idea behind the concept of social influence is that interacting agents strive to become more alike [Festinger 1950].

For example, agents in the Ising model for ferromagnets [Binney et al. 1992, Barthélémy 2011] are influenced by the state/opinions of the majority of their peers. This influence naturally drives the system towards an ordered state where all agents are either positive or negative (ferromagnets). Hence, the model can be interpreted as a very simple model for simulating (binary) opinion dynamics. However, the transition probabilities of the Ising model are influenced by temperature, representing the modeling of external or influential factors. In particular, if the temperature is above a certain threshold, consensus-finding, in terms of magnetization, becomes an unstable process that never converges. The Potts model [Wu 1982, Dorogovtsev et al. 2008] further extends the Ising model by increasing the number of potential states an agent can assume from two (positive or negative) to an arbitrary number greater than two. Other factors that might influence the process of reaching consensus is the size of the system under investigation [Tessone and Toral 2009]. In particular, this means that differently sized (or connected) systems potentially need different strategies to reach consensus.

Opinions are usually represented as a set of words or numbers for each agent individually. [Weidlich 1971] introduced such a model, based on sociodynamics, in 1971. [Galam et al. 1982, Galam and Moscovici 1991] analyzed the potential applications of the Ising model for simulating opinion dynamics starting in 1982.

The most wide-spread and adapted models to simulate (among others) opinion dynamics are the voter model [Clifford and Sudbury 1973, Holley and Liggett 1975], the Axelrod model [Axelrod 1997] as well as The Naming Game [Baronchelli et al. 2006].

The voter model constitutes that each agent is equipped with a binary
variable. At each step in time, the binary variable of one (randomly chosen) agent is synchronized with one of its neighbors variable. Introducing the concept of social influence for opinion dynamics. The voter model has since been adapted and extended by many researchers to fit an array of different purposes (e.g., Mobilia [2003], Mobilia and Georgiev [2005], Mobilia et al. [2007], Vazquez et al. [2003], Vazquez and Redner [2004], Castelló et al. [2006]).

The Axelrod model [Axelrod 1997] combines the notion of social influence – individuals becoming more similar upon frequent interactions – and the tendency that similar individuals will have a higher tendency (and frequency) to interact with each other. Each agent is endowed with a set of characterizing variables. The more variables are shared among two agents, the more similar they are. Given this description, one would assume that the described notions are self-reinforcing dynamics and hence, will inevitably produce stable networks with only identical agents. However, Castellano et al. [2000] have shown that the resulting number of different states is dependent on the number of characterizing variables. Large numbers are likely to result in very few similar individuals (high agent diversity). Analogously to the voter model, the Axelrod model has been extensively adapted, analyzed and expanded by researchers to broaden our understanding of the spread of (cultural) traits across agents (e.g., Klemm et al. [2003b], Flache and Macy [2007]).

The Naming Game originates from the idea to analyze and explore the evolution of language [Steels 1995]. Baronchelli et al. [2006] introduced the most basic version of The Naming Game in 2006, where a group of agents that communicate via a complete network, try to reach consensus when naming an entity. Each agent holds a list of synonyms or words associated with the entity, also referred to as vocabulary, under investigation. Every iteration (or step in time), two agents are chosen. One agent is assigned the role of the speaker, who randomly choses a word of her vocabulary. If the other agent – the listener – knows (i.e., also has the word in her vocabulary) the chosen word, both agents discard all other words in their vocabulary and “agree” on the common word. However, if the listener does not know the word of the speaker, the word is appended to her vocabulary and no words are discarded. In the next step another pair of nodes is chosen and process is repeated until either consensus is found or a predetermined number of steps (time) have passed. The Naming Game has spurred a complete line of dynamical models with a variety of different parameters, that each address different problems and tasks (e.g., Abrams and Strogatz [2003], Minett and Wang [2008], Wang
and Minett [2005], Castelló et al. [2006]). For an excellent and comprehensive introduction to opinion dynamics (among others) we refer the interested reader to Castellano et al. [2009].

7 Discussion, Limitations & Future Work

We have developed a model of activity behavior and applied it to the collaboration network extracted from the Zachary’s Karate Club (see Figure 3) dataset to illustrate its core mechanics. Subsequently, we continued with a linear stability analysis (cf. Section 2.2) and depicted the behavior that can occur when the master stability equation is invalidated (see Figure 3). Using our proposed model to simulate activity dynamics, we have shown that the overall activity in collaboration networks appears to be a composite of the Activity Decay Rate and the Peer Influence Growth Rate, as described in Section 2. In Section 3 we have fitted our model on empirical datasets to simulate activity dynamics trends.

The presented results are destined to be interpreted only and solely as an indicator for trends in activity dynamics, rather than absolute values that can be used for accurately predicting the activity for a given system. This is a direct result of the different approximations and simplifications (cf. Section 3) that we have made when estimating the parameters for our activity dynamics simulation. For future work we plan on extending the ability of our model to not only reflect on changes in activity dynamics but also properly cope with structural changes in the underlying collaboration networks. One additional limitation of the presented approach is the fact that nodes with a very small degree, which are not connected to the largest connected component, inevitably will lose activity until they reach the point of total inactivity. Including the structural evolution of a collaboration network in our analyses will allow us to mitigate this effect, as users will only be added to the collaboration network and considered in our calculations, once they have actually become active. One potential approach involves the investigation of snapshots of the collaboration networks at every $\tau$, providing additional insights on the evolution of the parameters of our model and the investigated systems. Additionally, we assume that peer influence is a symmetric property. This means that posts and replies exercise the same amount of influence on peers as we do not differentiate between different types of activity and influence will always traverse along both directions of the edges in our
Furthermore, all of our estimated parameters are calculated for the collaboration networks as a whole. Future work will also include extending the activity dynamics model to calculate the ratio $\frac{\lambda}{\mu}$ on a user level, rather than on a network level. This modification not only potentially increases the accuracy of our model but would also allow us to gather additional information for each user of the corresponding networks. Further, with an increased accuracy in our simulations it will be possible to conduct activity prediction experiments and emulate network attacks as well as optimize (arbitrary) cost-strategies for increasing activity in these systems.

The ratio $\frac{\lambda}{\mu}$—describing how fast a user loses activity (Activity Decay Rate $\lambda$) over how fast she regains activity over her neighbors (Peer Influence Growth Rate $\mu$)—fluctuates around the corresponding highest eigenvalue $\kappa_1$ for all investigated empirical datasets. Negative peaks in this ratio represent periods of time ($\tau$; in our case months) where activity grew faster than could be compensated by the Peer Influence Growth Rate. It naturally follows that a decrease of $\lambda$—resulting in less activity-loss per contribution for each user—is necessary to accomplish such drastic increases of activity. If the network itself is of a smaller scale and/or these negative peaks occur on a frequent basis, the activity dynamics of the corresponding networks are depending on the contributions (and thus influence) of single (individual) users. To compare the stability of the activity dynamics across multiple networks we calculated the System Mass and Activity Momentum $p$—indicating the required force to accelerate or render the corresponding collaboration networks inactive.

When comparing $p$ and the results of our empirical illustration (cf. Figures 6 and 7) between the different datasets, we can see that the Activity Momentum is very small for datasets that either (i) exhibit only a very small number of changes and are close to inactivity or (ii) exhibit a small $\kappa_1$ (see Figure 6 and 7). This suggests that we can use Activity Momentum as an indicator for the robustness of a collaboration network with regards to its activity.

Further, we can characterize the potential of a collaboration network to become self-sustaining by comparing the calculated ratios of $\frac{\lambda}{\mu}$ with the corresponding $\kappa_1$ and Activity Momentum. If the ratio is below $\kappa_1$, our master stability equation is invalidated, pushing the system towards a new fixed point where the forces of the Activity Decay Rate and the Peer Influence
Growth Rate reach an equilibrium so that the network converges towards a state of immanent and lasting activity (see Figure 3). If such a state is reached and combined with a high Activity Momentum, the corresponding collaboration network has reached critical mass of activity and has become self-sustaining; no external impulses are required to keep the network active. Of course, in real world scenarios, activity will not last forever without providing additional incentives as interest (and thus activity) in a system potentially decays over time. As a consequence, this would first result in an increase of $\mu$ and inevitably, with a sufficiently large $\mu$, the collaboration network would return to its stable fixed point, once our master stability equation holds again, and activity would once more converge towards zero. Once we extend our model to allow for user-based calculations, we will be able to not only calculate Activity Momentum for collaboration networks, but also for single and individual users.

Acknowledgements

This research was in part funded by the FWF Austrian Science Fund research projects P24866.

References

D. M. Abrams and S. H. Strogatz. Linguistics: Modelling the dynamics of language death. Nature, 424(6951):900–900, 2003.

A. Acerbi, S. Ghirlanda, and M. Enquist. The logic of fashion cycles. PloS one, 7(3):e32541, 2012.

R. M. Anderson and R. M. May. Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, USA, 1991.

S. Aral and D. Walker. Identifying influential and susceptible members of social networks. Science, 337(6092):337–341, 2012.

J. L. Aron, M. O’leary, R. A. Gove, S. Azadegan, and M. Schneider. The benefits of a notification process in addressing the worsening computer virus problem: Results of a survey and a simulation model. Comput. Secur.,
C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Reviews of modern physics*, 81(2):591, 2009.

X. Castelló, V. M. Eguíluz, and M. San Miguel. Ordering dynamics with two non-excluding options: bilingualism in language competition. *New Journal of Physics*, 8(12):308, 2006.

J. Cheng and M. S. Bernstein. Catalyst: Triggering collective action with thresholds. 2014.

N. A. Christakis and J. H. Fowler. The collective dynamics of smoking in a large social network. *New England journal of medicine*, 358(21):2249–2258, 2008.

F. Chung, L. Lu, and V. Vu. Eigenvalues of random power law graphs. *Annals of Combinatorics*, 7(1):21–33, 2003a.

F. Chung, L. Lu, and V. Vu. Spectra of random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, 100(11):6313–6318, 2003b.

P. Clifford and A. Sudbury. A model for spatial conflict. *Biometrika*, 60(3):581–588, 1973.

V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, 103(7):2015–2020, 2006.

C. Danescu-Niculescu-Mizil, R. West, D. Jurafsky, J. Leskovec, and C. Potts. No country for old members: User lifecycle and linguistic change in online communities. In *Proceedings of the 22Nd International Conference on World Wide Web*, WWW ’13, pages 307–318, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-2035-1. URL [http://dl.acm.org/citation.cfm?id=2488388.2488416](http://dl.acm.org/citation.cfm?id=2488388.2488416).

K. Dietz. Epidemics and rumours: A survey. *Journal of the Royal Statistical Society. Series A (General)*, pages 505–528, 1967.

40
S. N. Dorogovtsev, A. V. Goltsev, and J. F. Mendes. Critical phenomena in complex networks. *Reviews of Modern Physics*, 80(4):1275, 2008.

N. M. Ferguson, M. J. Keeling, W. J. Edmunds, R. Gani, B. T. Grenfell, R. M. Anderson, and S. Leach. Planning for smallpox outbreaks. *Nature*, 425(6959):681–685, 2003.

N. M. Ferguson, D. A. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, S. Iamsirithaworn, and D. S. Burke. Strategies for containing an emerging influenza pandemic in southeast asia. *Nature*, 437(7056):209–214, 2005.

L. Festinger. *Social pressures in informal groups: A study of human factors in housing*. Stanford University Press, 1950.

A. Flache and M. W. Macy. Local convergence and global diversity: The robustness of cultural homophily. *arXiv preprint physics/0701333*, 2007.

S. Galam and S. Moscovici. Towards a theory of collective phenomena: consensus and attitude changes in groups. *European Journal of Social Psychology*, 21(1):49–74, 1991.

S. Galam, Y. Gefen, and Y. Shapir. Sociophysics: A new approach of sociological collective behaviour. i. mean-behaviour description of a strike. *Journal of Mathematical Sociology*, 9(1):1–13, 1982.

A. Ganesh, L. Massoulié, and D. Towsley. The effect of network topology on the spread of epidemics. In *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, volume 2, pages 1455–1466. IEEE, 2005.

H. W. Hethcote. An immunization model for a heterogeneous population. *Theoretical population biology*, 14(3):338–349, 1978.

H. W. Hethcote. The mathematics of infectious diseases. *SIAM review*, 42(4):599–653, 2000.

R. A. Holley and T. M. Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *The annals of probability*, pages 643–663, 1975.
L. Hufnagel, D. Brockmann, and T. Geisel. Forecast and control of epidemics in a globalized world. Proceedings of the National Academy of Sciences of the United States of America, 101(42):15124–15129, 2004.

J. L. Iribarren and E. Moro. Impact of human activity patterns on the dynamics of information diffusion. Physical review letters, 103(3):038702, 2009.

M. J. Keeling and P. Rohani. Estimating spatial coupling in epidemiological systems: a mechanistic approach. Ecology Letters, 5(1):20–29, 2002.

J. O. Kephart and S. R. White. Directed-graph epidemiological models of computer viruses. In Research in Security and Privacy, 1991. Proceedings., 1991 IEEE Computer Society Symposium on, pages 343–359. IEEE, 1991.

J. O. Kephart and S. R. White. Measuring and modeling computer virus prevalence. In Research in Security and Privacy, 1993. Proceedings., 1993 IEEE Computer Society Symposium on, pages 2–15. IEEE, 1993.

J. O. Kephart, S. R. White, and D. M. Chess. Computers and epidemiology. Spectrum, IEEE, 30(5):20–26, 1993.

J. O. Kephart, G. B. Sorkin, D. M. Chess, and S. R. White. Fighting computer viruses: Biological metaphors offer insights into many aspects of computer viruses and can inspire defenses against them. Scientific American, 1997.

K. Klemm, V. M. Eguíluz, R. Toral, and M. San Miguel. Global culture: A noise-induced transition in finite systems. Physical Review E, 67(4):045101, 2003a.

K. Klemm, V. M. Eguíluz, R. Toral, and M. San Miguel. Nonequilibrium transitions in complex networks: A model of social interaction. Physical Review E, 67(2):026120, 2003b.

J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. ACM Transactions on the Web (TWEB), 1(1):5, 2007.

J. Leskovec, L. Backstrom, and J. Kleinberg. Meme-tracking and the dynamics of the news cycle. In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 497–506. ACM, 2009.
C. Lin and L. Segel. *Mathematics Applied to Deterministic Problems in the Natural Sciences*, volume 1. SIAM, 1988.

A. L. Lloyd and R. M. May. Spatial heterogeneity in epidemic models. *Journal of theoretical biology*, 179(1):1–11, 1996.

I. M. Longini, A. Nizam, S. Xu, K. Ungchusak, W. Hanshaoworakul, D. A. Cummings, and M. E. Halloran. Containing pandemic influenza at the source. *Science*, 309(5737):1083–1087, 2005.

G. Marwell, P. E. Oliver, and R. Prahl. Social networks and collective action: A theory of the critical mass, ill. *American Journal of Sociology*, 94(3):502–534, 1988.

R. M. May and R. M. Anderson. Spatial heterogeneity and the design of immunization programs. *Mathematical Biosciences*, 72(1):83–111, 1984.

J. W. Minett and W. S. Wang. Modelling endangered languages: The effects of bilingualism and social structure. *Lingua*, 118(1):19–45, 2008.

M. Mobilia. Does a single zealot affect an infinite group of voters? *Physical Review Letters*, 91(2):028701, 2003.

M. Mobilia and I. T. Georgiev. Voting and catalytic processes with inhomogeneities. *Physical Review E*, 71(4):046102, 2005.

M. Mobilia, A. Petersen, and S. Redner. On the role of zealotry in the voter model. *Journal of Statistical Mechanics: Theory and Experiment*, 2007 (08):P08029, 2007.

Y. Moreno, R. Pastor-Satorras, and A. Vespignani. Epidemic outbreaks in complex heterogeneous networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 26(4):521–529, 2002.

J. D. Murray. *Mathematical biology*. Springer, New York, 2002. ISBN 0387952233 9780387952239. URL [http://www.worldcat.org/search?qt=worldcat_org_all&q=9780387952239](http://www.worldcat.org/search?qt=worldcat_org_all&q=9780387952239)

S. A. Myers, C. Zhu, and J. Leskovec. Information diffusion and external influence in networks. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 33–41. ACM, 2012.
M. Newman. *Networks: an introduction*. Oxford University Press, 2010.

M. E. Newman, S. Forrest, and J. Balthrop. Email networks and the spread of computer viruses. *Physical Review E*, 66(3):035101, 2002.

P. Oliver, G. Marwell, and R. 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.

P. E. Oliver and G. Marwell. The paradox of group size in collective action: A theory of the critical mass. ii. *American Sociological Review*, pages 1–8, 1988.

R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical review letters*, 86(14):3200, 2001a.

R. Pastor-Satorras and A. Vespignani. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63(6):066117, 2001b.

R. Pastor-Satorras and A. Vespignani. Epidemic dynamics in finite size scale-free networks. *Physical Review E*, 65(3):035108, 2002.

R. Pastor-Satorras and A. Vespignani. *Evolution and structure of the Internet: A statistical physics approach*. Cambridge University Press, 2007.

R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani. Epidemic processes in complex networks. *arXiv preprint arXiv:1408.2701*, 2014.

D. R. Raban, M. Moldovan, and Q. 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. ISBN 978-1-60558-795-0. doi: 10.1145/1718918.1718932.

B. 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. ISBN 978-1-4503-2744-2. doi: 10.1145/2566486.2567984. URL [http://dx.doi.org/10.1145/2566486.2567984](http://dx.doi.org/10.1145/2566486.2567984).
L. A. Rvachev and I. M. Longini. A mathematical model for the global spread of influenza. *Mathematical biosciences*, 75(1):3–22, 1985.

J. Solomon and R. Wash. Critical mass of what? exploring community growth in wikiprojects. 2014.

L. Steels. A self-organizing spatial vocabulary. *Artificial life*, 2(3):319–332, 1995.

S. H. Strogatz. *Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry, And Engineering (Studies in Nonlinearity).* Studies in nonlinearity. Perseus Books Group, 1994.

B. Suh, G. Convertino, E. H. Chi, and P. Pirolli. The singularity is not near: slowing growth of wikipedia. In *Proceedings of the 5th International Symposium on Wikis and Open Collaboration*, page 8. ACM, 2009.

C. J. Tessone and R. Toral. Diversity-induced resonance in a model for opinion formation. *The European Physical Journal B-Condensed Matter and Complex Systems*, 71(4):549–555, 2009.

F. Vazquez and S. Redner. Ultimate fate of constrained voters. *Journal of Physics A: Mathematical and General*, 37(35):8479, 2004.

F. Vazquez, P. L. Krapivsky, and S. Redner. Constrained opinion dynamics: Freezing and slow evolution. *Journal of Physics A: Mathematical and General*, 36(3):L61, 2003.

A. Vespignani. Modelling dynamical processes in complex socio-technical systems. *Nature Physics*, 8(1):32–39, 2012.

C. Wagner, S. Mitter, C. Körner, and M. Strohmaier. When social bots attack: Modeling susceptibility of users in online social networks. *Making Sense of Microposts (# MSM2012)*, page 2, 2012.

S. Walk and M. Strohmaier. Characterizing and predicting activity in semantic mediawiki communities. In *SWCS14 Third International Workshop on Semantic Web Collaborative Spaces*, 2014, page 21, 2014.

W. S. Wang and J. W. Minett. The invasion of language: emergence, change and death. *Trends in ecology & evolution*, 20(5):263–269, 2005.
Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos. Epidemic spreading in real networks: An eigenvalue viewpoint. In Reliable Distributed Systems, 2003. Proceedings. 22nd International Symposium on, pages 25–34. IEEE, 2003.

W. Weidlich. The statistical description of polarization phenomena in society. British Journal of Mathematical and Statistical Psychology, 24(2):251–266, 1971.

F.-Y. Wu. The potts model. Reviews of modern physics, 54(1):235, 1982.