# Chapter 1
# Modeling User Dynamics in Collaboration Websites

Patrick Kasper, Philipp Koncar, Simon Walk, Matthias Wölbitsch, Tiago Santos, Markus Strohmaier, and Denis Helic

**Abstract** Numerous collaboration websites struggle to achieve self-sustainability—a level of user activity preventing a transition to a non-active state. We know only a little about the factors which separate sustainable and successful collaboration websites from those that are inactive or have a declining activity. We argue that modeling and understanding various aspects of the evolution of user activity in such systems is of crucial importance for our ability to predict and support success of collaboration websites. Modeling user activity is not a trivial task to accomplish due to the inherent complexity of user dynamics in such systems. In this chapter, we present several approaches that we applied to deepen our understanding of user dynamics in collaborative websites. Inevitably, our approaches are quite heterogeneous and range from simple time-series analysis, towards the application of dynamical systems, and generative probabilistic methods. Following some of our initial results, we argue that the selection of methods to study user dynamics strongly depends on the types of collaboration systems under investigation

Patrick Kasper
Graz University of Technology, e-mail: patrick.kasper@tugraz.at

Philipp Koncar
Graz University of Technology, e-mail: philipp.koncar@tugraz.at

Simon Walk
Detego GmbH, e-mail: s.walk@detego.com

Tiago Santos
Graz University of Technology, e-mail: teixeiradossantos@tugraz.at

Matthias Wölbitsch
Graz University of Technology, e-mail: m.woelbitsch@student.tugraz.at

Markus Strohmaier
RWTH Aachen University, e-mail: markus.strohmaier@humtec.rwth-aachen.de

Denis Helic
Graz University of Technology, e-mail: dhelic@tugraz.at

as well as on the research questions that we ask about those systems. More specifically, in this chapter we show our results of (i) the analysis of nonlinearity of user activity time-series, (ii) the application of classical dynamical systems to model user motivation and peer influence, (iii) a range of scenarios modeling unwanted user behavior and how that behavior influences the evolution of the dynamical systems, (iv) a model of growing activity networks with explicit models of activity potential and peer influence. Summarizing, our results indicate that intrinsic user motivation to participate in a collaborative system as well as peer influence are of primary importance and should be included in the models of the user activity dynamics.

## 1.1 Introduction

New collaboration websites continuously emerge on the Web. Users of such communities work together towards a defined goal (e.g., building a knowledge base), which sets collaboration websites apart from more common social networks. Whereas some collaboration websites reach a sufficient level of user-activity to sustain themselves, preventing a transition towards inactivity, many websites perish over time or fail to establish an active community at all. The Q&A platform StackOverflow[1] is a successful example of such a collaboration website. Users can ask questions on programming related topics or share their knowledge by answering questions from other members of the community. The explicit goal of the website states *"With your help, we're working together to build a library of detailed answers to every question about programming."*[2]. A declining community may struggle to meet this ambitious goal in an ever-growing subject field such as programming. Thus, the success of the StackOverflow website relies heavily on the active community collaborating to answer any open questions. However, we as research community still do not fully understand the factors that drive the users to participate and contribute to such websites. This understanding would allow us to support the website operators in their efforts to build a successful website around a flourishing user community.

Initial work in this field frequently concentrated on interactions between users on websites, or how information spreads through the community [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13]. Nevertheless, to predict success and potentially support websites in their efforts to reach self-sustainability, we argue that understanding as well as modeling the various aspects of user dynamics that go beyond information spreading is of crucial importance.

One of the major problems faced by both new and existing collaboration websites—such as Wikipedia or StackOverflow—revolves around efficiently

[1] https://stackoverflow.com/

[2] https://stackoverflow.com/tour

identifying and motivating the appropriate users to contribute new content. In an optimal scenario, any newly contributed content provides enough incentive on its own, triggering further actions and contributions. Once such a self-reinforced state of increased activity is reached, the system becomes self-sustaining, meaning that sufficiently high levels of activity are reached, which will keep the system active without further external impulses. StackOverflow is an example for a highly active collaboration website that has already become self-sustained (in terms of activity), evident in the steadily growing number of supporters and overall activity.

However, these self-sustaining states [14, 15, 16, 17] are neither easy to reach nor guaranteed to last. For example, Suh et al. [18] showed that the growth of Wikipedia is slowing down, indicating a loss in momentum and perhaps even first evidence of a collapse. Moreover, we generally lack the tools to properly analyze these trends in activity dynamics and thus, cannot even perform tasks such as detecting these self-sustaining system states. Therefore, we argue that new tools and techniques are needed to model, monitor and simulate the dynamics in collaboration websites.

In this chapter, we set out to shed further light on the complex user dynamics in collaboration websites. More specifically, to investigate the success and failure of collaboration websites, we are interested in the factors that govern growth and decline of the activity in such communities. Moreover, we also aim at evaluating the *robustness* and *stability* of collaborative websites.

**Approach.** To this end, we present a diversified range of approaches, each tackling different aspects of user dynamics in collaboration websites. We use empiric data originating from various types of collaboration websites, such as StackExchange instances and Semantic MediaWikis to report our findings.

We argue that there are two factors that influence the activity of any single user in collaboration websites. First, the activity or rate of contributions of a user is influenced by their intrinsic motivation to participate in a collaborative community. This motivation may decay over time in a mechanism called *activity decay*. A previously active user may lose interest in the community and contribute less and less over time unless stimulated through other means. This behavior has been observed in many different websites [18, 19, 17]. In another scenario the intrinsic motivation of a user may remain constant or even increase with time. We summarize this phenomenon as *activity potential*. Second, *peer influence* is a mechanism in which users influence other members of the community. For example, when users post a question to StackExchange and receive helpful answers from other users, they may want to help others in the same way by answering other open questions. Note, contributions by peers are not necessarily always positive. Internet trolls may attempt to disrupt the community by adding detrimental content [20].

We discuss these influential forces and their interactions by (i) applying several tests for nonlinearity on the activity time series of various StackExchange instances to reveal *complex user behavior*. Thereafter, we (ii) apply a dynamical systems model to investigate the *long-term activity decay* (users losing interest over time) and how this decay is countered by the peer influence from the other users. Iterating upon this idea of peer influence we (iii) conduct experiments investigating the *influence of trolls* who spread negative activity through peer influence by adding detrimental content to the websites, and lastly, we (iv) present a generative probabilistic model to create synthetic activity networks and study the *emergence of clustering* in the underlying user networks.

**Contribution.** This chapter provides an overview of several methods and ideas concerning dynamics in collaboration websites. Further, we shed light on some factors contributing to their eventual success or failure. We summarize our main findings as follows. Models incorporating the user-centered concepts of *user motivation* and *peer influence* can capture crucial aspects regarding activity in collaboration websites, such as system robustness and stability. Further, depending on a particular community that we investigate the technical approaches and models need to be carefully chosen.

## 1.2 Related Work

**Analysis of Online Communities**. We know that, at some point in time, well-established collaboration websites, such as StackOverflow, have become self-sustained. There, sufficiently high levels of activity are reached, which will keep the system active without further external impulses. However, many websites never reach this state and those that do, are not guaranteed to remain there indefinitely [14, 15, 16, 17]. With the continuous growth in the number of such websites, many researchers have investigated these communities to better understand the dynamics governing growth and decline. For example, Schoberth et al. [21] and Crandal et al. [22] analyzed time-series data of websites to investigate the communication activities and social influences of their users. Analyzing the roles different types of users play, researchers characterized the users to infer properties about their communities as a whole [23, 24, 25, 26]. Using methods related to the work by Zhang et al. [27], multiple authors studied the evolution-dynamics of Web communities and their underlying networks [28, 29, 30, 31, 32, 33]. These networks often serve as a basis for dynamical systems models of the communities.

**Nonlinear Time Series Analysis.** To obtain a better understanding of the properties in high-dimensional dynamical systems, researchers have utilized nonlinear time series analysis. Bradley and Kantz [34] provided a thorough overview of applied nonlinear time series analysis. The works by Eckmann et al. [35] and Marwan et al. [36] described the use of Recurrence Plots to visually analyze complex systems. Zbilut and Webber [37, 38] further extended these visualizations with a method called Recurrence Quantification Analysis (RQA). These tools provided means to, for example, investigate the chaotic behavior in stock markets [39, 40] or predict the outcome of casino games, such as a roulette wheel [41].

Here, we present work employing various tests for nonlinearity to reveal latent nonlinear behavior in collaborative websites and their communities.

**Dynamical Systems & Activity Dynamics.** Dynamical systems in a non-network context are a well-studied scientific and engineering field. Strogatz [42] and Barrat et al. [43] provided an in-depth introduction to dynamical systems. Within the contextual scope of online communities, researchers primarily used dynamical systems to analyze and understand the diffusion of information in online social-networks for purposes such as viral marketing [9, 10, 11, 12, 13]. Recently, in the context of activity dynamics, Ribeiro [31] conducted an analysis of the daily number of active users that visit specific websites, fitting a model that allows predicting if a website has reached self-sustainability, defined by the shape of the curve of the daily number of active users over time.

In this chapter, we present a model to simulate activity as a dynamical system on online collaboration networks. Here, two forces, decay of motivation and peer influence govern the activity-potential of users. Moreover, we describe work on how these concepts facilitate the generation of synthetic networks. Online communities becoming increasingly accostable to their users does not always lead to higher overall activity. Internet trolls, for example, generate unwanted content [44, 45, 46, 47, 48, 20], creating additional strain for others who attempt to keep the community healthy.

Thus, we present an extension to the previous model incorporating the idea of trolls emitting negative peer influence and discuss how such negative activity can impact the user dynamics in collaboration websites.

## 1.3 Datasets

The Web offers a multitude of ways in which people can communicate and collaborate in a group. To capture some of this diversity, we utilize empirical datasets stemming from different types of collaboration websites. Here, we provide a general overview of the empiric datasets in our experiments, and how we extract the user networks from the raw data.

**StackExchange instances.** StackExchange is a network of currently 172 Question & Answer communities. Here, users can post questions and other members of the community can provide and discuss answers. Some of the most popular instances are StackOverflow[3] and the English StackExchange[4]. We extract the network by representing each user with a node and draw an edge whenever user A replies to a post by user B. The full

[3] `https://stackoverflow.com/`

[4] `https://english.stackexchange.com/`

dataset from which we draw our networks is publicly available[5]. We denote these datasets with a *SE* suffix. For example, we call the network extracted from the English StackExchange as *englishSE*.

**Semantic MediaWikis.** The Semantic MediaWiki[6] is an extension to the MediaWiki software and allows for storing and querying structured data within the Wiki. We build the community network by representing each contributor with a node and draw an edge whenever two users work on the same page. We collected the data we use in our experiments from the live MediaWiki API, which is now unavailable. However, a comprehensive dump of the Semantic MediaWiki is publicly available[7]. We denote these datasets with a *MW* suffix. For example, we call the network extracted from the Neurolex Semantic MediaWiki as *neurolexMW*.

**SubReddits.** A SubReddit is a community within Reddit for a specific topic. While some of these communities act as recommendation platforms or Q&A sites akin to StackExchange, others aim to facilitate a platform for open discussion of various topics. We extract a network from a SubReddit by representing each user with a node and draw an edge when one user replies to a post by another user. These dumps from Reddit are publicly available[8]. We denote these datasets with a *SR* suffix. For example, we call the network extracted from the Star Wars Subreddit as *starwarsSR*.

## 1.4 Complex User Behavior in Collaboration Websites

As a first step towards the goal of identifying factors indicating successful or failing collaboration websites, we set out to identify complex (nonlinear) user behavior present in the data. To reveal and characterize any hidden nonlinear patterns, we construct the activity time series from the datasets of 16 randomly selected StackExchange instances and conduct a set of nine established tests for nonlinearity on them. This information allows for a decision on whether a standard time-series model such as the AutoRegressive Integrated Moving Average (ARIMA) is sufficient to capture and predict activity, or more complex approaches (e.g., dynamical systems) should be employed.

**Activity time series.** We construct the activity time series from a dataset by first measuring the activity—the number of questions, answers, and comments—per day. To remove outliers in the data we smooth the time series with a rolling mean over a seven-day period. Finally, we calculate the sum of the smoothed activity over all users per week, yielding a time series with one entry per week representing the activity in the corresponding community.

---

[5] `https://archive.org/details/stackexchange`

[6] `https://www.semantic-mediawiki.org`

[7] `https://archive.org/details/wiki-neurolexorg_w`

[8] `https://files.pushshift.io/reddit/`

**Experiments & Results.** To reveal hidden nonlinear patterns in our activity time series, we apply the following tests for nonlinearity on each dataset and report the results: (i) *Broock, Dechert and Scheinkman test* [49]; (ii) *Teraesvirta neural network test* [50]; (iii) *White neural network test* [51]; (iv) *Keenan one-degree test for nonlinearity* [52]; (v) *McLeod-Li test* [53]; (vi) *Tsay test for nonlinearity* [54]; (vii) *Likelihood ratio test for threshold nonlinearity* [55]; (viii) *Wald-Wolfowitz runs test* [56, 55]; (ix) *Surrogate test - time asymmetry* [57].

We apply these tests without configuration changes, except for the *Broock, Dechert, and Scheinkman* and *Wald-Wolfowitz runs* tests. As described in Zivot and Wang [58, p. 652], we compute the test statistic of *Broock, Dechert, and Scheinkman* on the residuals of an ARIMA model, to check for nonlinearity not captured by ARIMA. For the *Wald-Wolfowitz runs* test, since a run represents a series of similar responses, we define a positive run as the number of times the time series value was greater than the previous one [59].

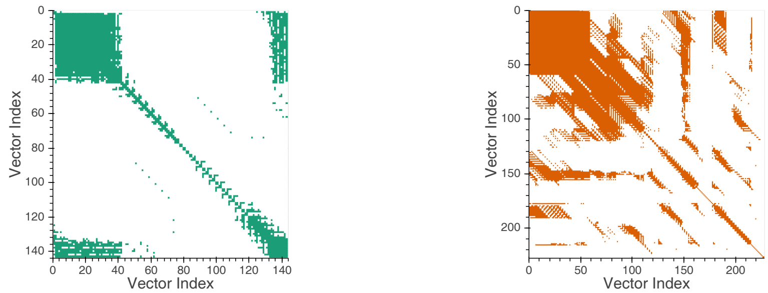
To validate the plausibility of this categorization we compare the forecast performance from three standard time series models, namely ARIMA, exponential smoothing models (ETS), and linear regression models, with nonlinear models, reconstructed from the observed activity time series.

Table 1.1 lists test results on the 16 StackExchange instances. Our results reveal that on the one hand, there are StackExchange communities with mostly linear behavior, such as *englishSE* and *unixSE* as only two tests suggest nonlinearity. On the other, we see that for the communities *bicycleSE*, *bitcoinSE*, and *mathSE* the majority of tests suggest nonlinearity.

**Table 1.1: Results of statistical tests.** This table lists the activity time series length in weeks, embedding parameters $\tau$ and $m$, the number and reference of statistical tests indicating nonlinearity ($\alpha = 0.05$), and the RMSE (lower is better) of a 1 year forecast per model for each dataset. Further it lists the ranking the Friedman test for datasets with less than or five or more tests suggesting nonlinearity. The indices refer the individual tests as listed in Section 1.4.

| **Dataset** | Weeks | $\tau$ | $m$ | **Nonlin. test score** | Positive non-linearity tests | RMSE ARIMA | RMSE ETS | RMSE Linear | RMSE Nonlin. |
|---|---|---|---|---|---|---|---|---|---|
| englishSE[b] | 240 | 2 | 9 | 2/9 | (i),(v) | 0.679 | 0.445 | 0.332 | **0.308** |
| unixSE[b] | 239 | 1 | 7 | 2/9 | (i),(v) | 0.209 | 0.209 | 0.241 | **0.207** |
| chemistrySE[b] | 158 | 2 | 7 | 3/9 | (i),(v),(viii) | 0.498 | **0.253** | 0.324 | 0.461 |
| webmastersSE | 244 | 1 | 8 | 3/9 | (iv),(v),(ix) | **0.231** | 0.252 | 0.334 | 0.234 |
| chessSE | 148 | 2 | 8 | 4/9 | (i),(iv),(v),(ix) | **0.254** | –[a] | 0.562 | 0.511 |
| historySE | 177 | 1 | 9 | 4/9 | (i),(iv),(v),(viii) | 0.350 | **0.236** | 0.304 | 0.405 |
| linguisticsSE | 181 | 2 | 6 | 4/9 | (i),(iv),(v),(ix) | **0.251** | 0.270 | 0.300 | 0.328 |
| sqaSE | 200 | 3 | 9 | 4/9 | (i),(iv),(v),(ix) | 1.813 | **0.253** | 0.654 | 0.390 |
| texSE[b] | 241 | 1 | 7 | 4/9 | (v),(vi),(viii),(ix) | **0.158** | **0.158** | 0.276 | 0.275 |
| tridionSE | 107 | 1 | 7 | 4/9 | (ii),(iii),(iv),(v) | **0.271** | –[a] | 0.614 | –[a] |
| Friedman test rank on datasets with nonlin. test score < 5/9 | | | | | | 2 | **1** | 4 | 3 |
| arduinoSE | 56 | 1 | 10 | 5/9 | (i),(ii),(iii),(iv),(v) | **0.348** | –[a] | –[a] | –[a] |
| sportsSE | 159 | 1 | 7 | 5/9 | (i),(iv),(v),(viii),(ix) | **0.244** | 0.334 | 0.401 | 0.332 |
| uxSE | 239 | 2 | 8 | 5/9 | (i),(iii),(iv),(v),(vi) | 0.347 | 0.174 | 0.349 | **0.137** |
| bitcoinSE | 182 | 4 | 11 | 6/9 | (i),(ii),(iii),(iv),(v),(ix) | 0.609 | 0.554 | 0.593 | **0.578** |
| mathSE[b] | 242 | 2 | 8 | 6/9 | (i),(ii),(v),(vi),(viii),(ix) | **0.132** | 0.231 | 0.352 | 0.291 |
| bicyclesSE | 235 | 2 | 7 | 7/9 | (i),(ii),(iii),(iv),(v),(viii),(ix) | 0.297 | 0.309 | 0.325 | **0.280** |
| Friedman test rank on datasets with nonlin. test score $\geq$ 5/9 | | | | | | 2[c] | 2[c] | 4 | 1 |

[a] This activity time series is too short for a 1 year forecast with this model.
[b] This activity time series had a strong linear trend, so the results above concern the activity time series detrended with linear regression.
[c] These models achieved the same rank in the Friedman test for this group of datasets.




**(a)** bitcoinSE

**(b)** mathSE

**Fig. 1.1: Recurrence Plots (RP) for activity time series.** This figure illustrates the Recurrence Plots of the bitcoinSE and mathSE websites. Figure 1.1b shows a higher density of recurrence points in the upper left corner, gradually diminishing towards the lower right; this is a sign of a drift in the activity time series, still present after removing the linear trend. Both examples hint at non-stationary transitions in the activity time series.

A higher number of tests suggesting nonlinearity for a community indicates a better fit for models based on nonlinear time-series analysis. The prediction experiments and the Friedman test ranks [60] on datasets with mostly negative test results (less than five) indicates that for these communities ARIMA and ETS models result in the best fit. For the other datasets (more than four positive tests), nonlinear models yield the lowest error.

The nonlinearity tests by Lee et al. [51] and Teräsvirta et al. [50] utilize neural networks and appear to be more sensitive to the presence of nonlinear dynamics than the other tests, since they test positive for nonlinearity four times more often in the dataset group with five or more tests indicating nonlinearity than in the other dataset group. We attribute the usefulness of these two tests to the well-studied ability of neural networks to model nonlinear behavior.

In a second experiment, we use with Recurrence Plots [36] to analyze the nonlinear properties for two exemplary StackExchange instances *bitcoinSE*, and *mathSE*. Both websites have a high number of positive nonlinearity tests.

Figure 1.1 illustrates the results for these two instances. Despite having the same number of positive tests for nonlinearity, these visualizations depict different patterns in their activity. In particular, Figure 1.1b shows a higher density of recurrence points in the upper left corner, gradually diminishing towards the lower right corner. This structure reveals a drift pattern which is present even after linear detrending.

**Findings.** We find that we can model activity on collaboration websites through reconstruction of their underlying, dynamical systems, with some communities showing more signs of nonlinear behavior than others. In particular, the knowledge of any drift- or periodicity patterns in the data provides information on which approach may yield the best accuracy.

For a more detailed discussion of the topic refer to Santos et al. [61].

## 1.5 Activity Decay & Peer Influence

On collaboration websites contributing users tend to lose interest over time. Wikipedia is a prominent example of such a website with a declining userbase [19]. To address this problem, we present a model based on dynamical systems where the motivation of a user decays over time (intrinsic activity decay). Danescu-Niculescu-Mizil et al. [25] were able to observe this behavior across different online communities. However, in our proposed model, users also gain activity from their neighbors through peer-influence to compensate for the intrinsic decay, which builds upon the notion that people tend to copy their friends and peers [62, 63, 64]. This activity dynamics model is capable of capturing and simulating activity in collaboration websites. We fit this model to a number of StackExchange instances and Semantic MediaWikis to simulate trends in activity dynamics. Further, we utilize the model to calculate a threshold indicating self-sustainability. Being able to monitor and measure the stability of a website with regards to user activity indicates how susceptible a system is to fluctuating members. For example, in a volatile website, a small number of highly active users (emitting a lot of peer influence) leaving, could result in activity decreasing to the point of total inactivity.

**Dynamical Systems.** The proposed model utilizes the formalism of dynamical systems—meaning that activity is modeled by a system of coupled nonlinear differential equations. Each user in the system is represented by a single quantity (the current activity), and the collaborative ties between users define the coupling of variables.

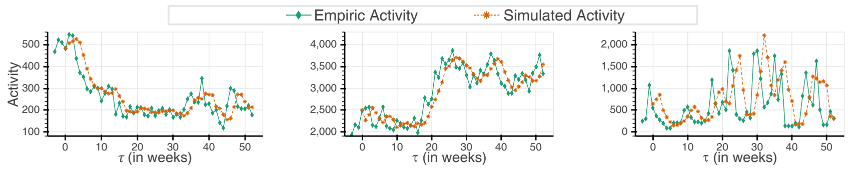
The model builds on two mechanisms which postulate that with time users lose interest to contribute and that, on the other hand, users are influenced by the actions taken by their peers.

**Modeling activity.** 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 links represent the fact that the users have collaborated in the past. We represent the network with an $n \times n$ adjacency matrix $\boldsymbol{A}$, where $n$ is the number of nodes (users) in the network. We set $A_{ij} = 1$ if nodes $i$ and $j$ are connected by a link and $A_{ij} = 0$ otherwise. Since collaboration links are undirected, the matrix $\boldsymbol{A}$ is symmetric, thus $A_{ij} = A_{ji}$, for all $i$ and $j$.

We model activity as a continuous real-valued dimensionless variable $x_i$ (representing ratio of the current activity of user $i$ over some critical activity threshold) evolving on node $i$ of the network in continuous dimensionless time $\tau$. We write the time evolution equation as follows:

$$\frac{dx_i}{d\tau} = -\frac{\lambda}{\mu} x_i + \sum_j A_{ij} \frac{x_j}{\sqrt{1 + x_j^2}}. \tag{1.1}$$

There is only one parameter in our dynamics equation, namely the ratio $\lambda/\mu$. This is a dimensionless ratio of two rates: (i) The *Activity Decay Rate*




**(a)** bitcoinSE **(b)** englishSE **(c)** neurolexMW

**Fig. 1.2: Activity simulation.** The figure depicts the results of our activity dynamics simulation for the StackExchange datasets and Semantic MediaWikis. In all our analyzed datasets, the simulated activity dynamics exhibit a notable resemblance to the empirical activity.

$\lambda$, which is the rate at which a user loses activity (or motivation), 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 users lose activity due to the decay of motivation than they can gain due to positive peer influence of a single neighbor. For example, a ratio of $\lambda/\mu = 100$ would mean that the users intrinsically lose activity 100 times faster than they potentially can get back from one of their neighbors.

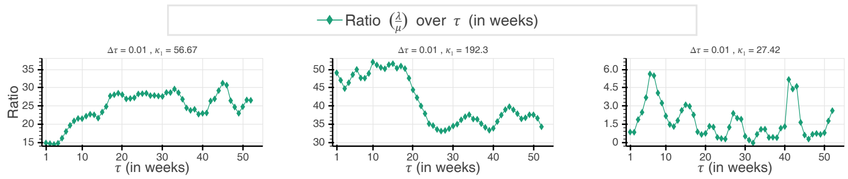
The master stability equation for our activity dynamics model is

$$\kappa_1 < \frac{\lambda}{\mu}, \tag{1.2}$$

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 ($\lambda/\mu$). If this stability condition is satisfied, the fixed point $x^* = 0$, in which there is no activity at all ("inactive" system), represents 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.

**Experiments & Results.** To estimate $\lambda/\mu$ for the empirical datasets we employ an output-error estimation method. First, we formulate the estimation of the model parameter as an optimization problem. As objective function, we use a least-squares cost function. Second, we solve the optimization problem numerically, using the method of gradient descent in combination with the Newton–Raphson method [65] to speed up the calculations. Finally, we evaluate the accuracy of the ratio estimate by calculating prediction errors on unseen data.

This prediction serves as a demonstration 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. The simplifications, such as the static network structure and average model parameters over weeks and users, entail that any results cannot be used for an accurate prediction of the activity in the system, and naturally limit the




**(a)** bitcoinSE **(b)** englishSE **(c)** neurolexMW

**Fig. 1.3: Evolution of ratios $\lambda/\mu$.** The evolution of the ratios $\lambda/\mu$ ($y$-axes) over $\tau$ (in weeks; $x$-axes) for the StackExchange datasets and for the Semantic MediaWikis. The smaller the ratio, the higher the levels of activity in Figure 1.2. Small variances in $\lambda/\mu$ over time indicate that activities of the systems are less influenced by the activity of single individuals than they are by peer influence.

accuracy of our results. These limitations are particularly visible whenever there are large and sudden increases in activity in the collaboration websites. Figure 1.2 depicts the results of the activity dynamics simulation. Overall, the results gathered from the activity dynamics simulation exhibit notable resemblance to the real activities of the corresponding datasets. Note how in some cases our simulation yields a higher activity increase than the real data (e.g., Figure 1.2c). A possible cause for this behavior is the static network structure where users might be influenced by peers who actually join the network at a later point in time.

Figure 1.3 depicts the value of the calculated ratios $\lambda/\mu$ ($y$-axis) for each week ($x$-axis) of our activity dynamics simulation. If the ratio is higher than $\kappa_1$, our master stability equation holds, and the system converges towards zero activity (over time). 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 $\kappa_1$, the master stability equation has been invalidated and the system will converge towards a new fixed point of immanent activity (cf. Equation 1.2). Robust systems are 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.

Note that one advantage of our model over other existing approaches, such as autoregression, is the interpretability of the ratio $\lambda/\mu$. For example, a ratio of 4 means that users intrinsically lose activity 4 times faster than they can get back from one of their peers, while the coefficients of the autoregression lack such interpretable characteristics. Further, using the concept of dynamical systems we can represent the underlying mechanisms in a closed form. This allows for further detailed analytical inspections, such as a linear stability analysis, which is much harder, if not impossible, to conduct for other models (i.e., agent-based models, autoregression or more complex models based on dynamical systems).

**Practical Implications.** Using our proposed model, we can characterize networks based on their susceptibility to changes in activity (referred to as Activity Momentum in [32]). Hence, community managers could use the proposed model as indicator for the robustness of their collaboration website with regards to its activity dynamics.

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 the susceptibility to changes in user activity of the collaboration network. 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. If such a state is reached combined with a low susceptibility to changes in user activity, 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.

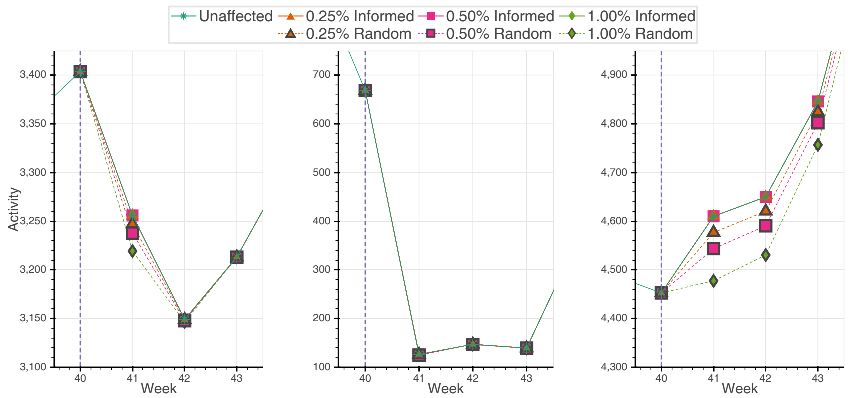
Of course, in real-world scenarios, activity will not last forever without providing additional incentives (e.g., user profile badges displaying support or expertise), as interest (and thus activity) in a system potentially decays over time. As a consequence, this would first result in an increase of $\mu$ 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.

**Findings.** Using our proposed model to simulate activity dynamics, we show that the overall activity in collaboration websites appears to be a composite of the *Activity Decay Rate* and the *Peer Influence Growth Rate*. A first analysis of the model suggests 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 in activity.

For a more detailed presentation and discussion of factors such as *System Mass* and *Activity Momentum*, see Walk et al. [32].

## 1.6 Negative Activity in Collaboration Websites

While most users in collaboration networks contribute by adding helpful content—in the case of StackExchange by asking questions or providing helpful answers—Internet Trolls post unwanted content for their own amusement [20]. We investigate such unwanted users and how they affect collaboration websites by adapting the Activity Dynamics model presented in Section 1.5. Thus far, we considered peer influence as a purely positive force. In this proposed modification, introduced trolls emit negative activity to their




**(a)** englishSE **(b)** neurolexMW **(c)** starwarsSR

**Fig. 1.4: Effects of trolls on total system activity.** This figure depicts the impact trolls have for the first weeks after their introduction to the network. When trolls connect to highly active users (informed strategy) the effect on the systems activity is minimal, whereas with the random connection strategy we observe a noticeable impact on the activity in the network.

neighbors. As an example, a troll may post a nonsensical question on StackExchange, or deliberately post wrong answers. Other users now have to spend time to either report or remove the unwanted post. We argue that this consumes the time these users could have potentially used to answer an open question. Understanding how trolls can disrupt the activity in collaboration websites can be used to derive strategies to prevent or minimize their impact.
**Modeling Troll-Users.** We model the impact of disruptive content in the form of negative activity. A troll-user emits negative activity to their connected users and simultaneously receive productive positive activity as their neighbors try to compensate for it. Further, we argue that trolls commit to their cause and therefore do not lose motivation on their own. Thus, we disable the *motivation decay* for these users. Within a network, we define the total number of normal users as $N$ and the number of trolls as $T$. Thus, $N$ remains constant regardless how many trolls enter the network.

Whenever a troll enters a network at the beginning of our experiments, they connect to a number of existing users ($\alpha$). We define two methods for this process; First, with the *random strategy* the troll connects to other users uniformly at random ($P = \frac{\alpha}{N}$). To achieve this, the troll may extract a list of all users within the network and then perform a random selection. Second, with the *informed strategy* the troll specifically targets and connects to highest degree users. Here the troll observes the collaboration website for some time before selecting their targets according to this strategy.

The negative activity of a troll absorbs the positive activity spread via peer influence. Note that, when a normal user receives enough negative activity, their own activity can become negative for some time. Whenever the incoming peer influence received by a troll exceeds their outgoing activity, the troll is defeated, and we remove their corresponding node from the network.

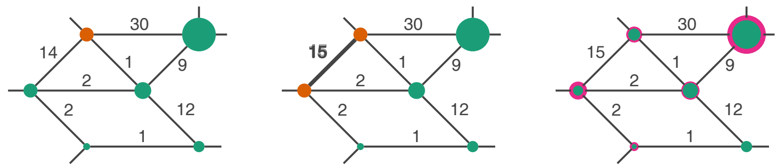
**Experiments & Results.** In this experiment, we aim to determine how trolls affect the overall activity in networks. For the initial 39 weeks we calculate the activity within a network unaltered (akin to the model in Section 1.5). After this point, we introduce the troll-users and observe their impact. For each troll, we set their starting activity at week 40 to $-5$ and conduct each experiment twice. Once, with trolls following the random strategy, and once with informed connection strategy. Further, we fix the parameter $\alpha$ (number of connections per troll) to be equal to the mean degree of all existing users. In total, we add trolls equal to the of 0.25%, 0.50%, and 1.00% of existing users ($N$) and investigate their initial impact.

Figure 1.4 illustrates the simulated system activity for the first three weeks after we add the trolls to the networks. Our results suggest, that trolls connecting to highly active users do not affect the overall activity in the network. We attribute this to the peer influence emitted by the troll being comparably insignificant. However, when we connect the trolls at random, users are more heavily influenced. A sporadic contributor may lose interest upon exposure to trolls. Small, but well-connected networks may lack a sizable body of casual users. Figure 1.4b illustrates such a network (neurolexMW). Due to this strong structure, newly introduced trolls fail to disrupt the system regardless of their connection strategy.

**Findings.** Up to a threshold, highly active users can compensate for the negative activity trolls emit, whereas random users can be more susceptible and may even temporarily spread negative content on their own further reducing the activity in a network. Below this threshold, the negative activity is nullified over time. However, once enough trolls connect to the highly active users within a network and overwhelm their positive activity, the networks collapse rapidly, ending all productive contribution.

Based on these findings, website administrators may entrust highly active users to moderate their communities. These moderators would be instructed in how to deal properly with occurring trolls, making them less susceptible for distractions (unproductive activity). Additionally, moderators could support peripheral users targeted by trolls and handle other detrimental factors, such as spam bots or illegal content. Offering incentives, such as additional functionality on their website or even money, could help motivating users to become moderators. A different approach would be to use machine learning techniques to automatically detect occurring trolls, for example, by identifying fake profiles [66] or by inspecting textual contents of comments [67, 68].

For further discussion of this subject and experiments on how trolls affect and infect users, see Koncar et al. [69].




**(a)** $t = x$ **(b)** $t = x$ **(c)** $t = x + 1$

**Fig. 1.5: Illustrative model example.** The highlighted node in Fig. 1.5a becomes active and interacts with the second highlighted node in Fig. 1.5b, reinforcing the tie between them. The outlines in Fig. 1.5c depict the additional peer-influenced activation probability in the next iteration.

## 1.7 Peer Influence in Temporal Networks

Thus far, we have represented the collaboration websites and their user networks in a static form. However, in the real world, new users frequently join the communities, while other members leave after a while. We approach this dynamic user-base by presenting a generative model to create synthetic networks. Existing network generators incorporating the concept of activity often solely consider the intrinsic activity potentials as sources of activity [70, 71]. But, we have shown in the previous sections, that interaction between users is an important factor to consider. Thus, we present a generative model that incorporates peer influence (similar to Section 1.5) and tie strength (how frequent two users interact) as explicit mechanics. With this model for generating synthetic networks we are able to explore new ideas and conduct experiments before verifying them on empiric networks.

**Generating activity networks.** We model the influence that a node receives from their neighbors in each time step (iteration $t$) as the increase in the activity potential according to the number of active neighbors in the previous iteration ($t-1$) and the tie strengths.

The equation for the peer influence for a node $v_i$ is:

$$p_i(t) = \frac{\alpha_i(t)\, q}{\sqrt{\alpha_i^2(t) + \theta^2}}, \tag{1.3}$$

where $\alpha_i(t)$ is the weighted fraction of active neighbors and $q$ is the parameter for the maximum peer influence. Further, ($\theta > 0$) denotes a critical threshold, determining the required fraction of active neighbors to set the peer influence probability close to this maximum.

Any node can become active based on either their own intrinsic activity or on the peer influence. When they do they select a new node as the partner for the interaction and either create or reinforce the tie between them.

The resulting network exhibits structures seen in real-world networks, due to the partner selection, which follows a set of rules. First, a memory effect as described by Karsai et al. [72] (depending on the number of currently

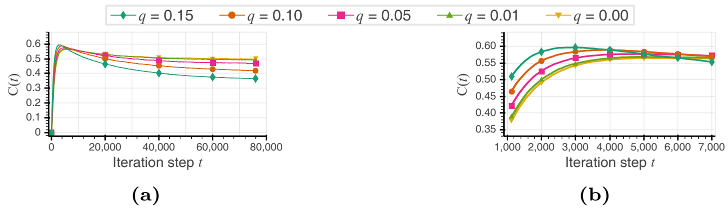



**Fig. 1.6: Average clustering coefficient ($C(t)$) evolution.** This figure depicts the average clustering coefficient (y-axis) at each iteration step (x-axis) over various values for the maximum peer influence ($q$). Higher values for $q$ result in stronger peer influence effects. Note that the value of $q$ also affects the time until convergence. Figure 1.6b illustrates the timespan where the $C(t)$ is maximal.

existing ties for the node and the memory strength parameter $c$) defines the probability to reinforce an existing tie. More precisely, this probability is equal to $\frac{c}{k_i+c}$, where $k_i$ is the number of current neighbors. Second, if a node wants to form a new tie, it tries to perform a cyclic closure [73]—by interacting with a randomly selected neighbor of a neighbor—with the probability $p_\Delta$, or a focal closure [74], which emulates homophily (i.e., similar users connect to each other). The latter is performed with a probability of $1 - p_\Delta$, or if there are no suitable candidates for a cyclical closure. This is, for example, the case if a node becomes active for the first time.

Figure 1.5 illustrates these mechanics. Figure 1.5a describes the network at iteration step $t = x$, where the numbers along the edges represent the tie strengths and the site of the nodes indicates their intrinsic activity potential. In this example, the highlighted node (top left) becomes active. It selects the newly highlighted node (left) in Figure 1.5b as the partner, which becomes active as due to this interaction. As a result, they reinforce the tie between them. At the start of the next iteration, the nodes receive peer influence from their neighbors active in the last iteration (outlines in Figure 1.5b). Note how the node in the top right corner receives a high amount of peer influence due to its strong ties.

To prevent the network from becoming fully connected after a sufficient amount of iterations, every node has a probability to be removed. In this case, we delete the node from the network and introduce a new node (without any existing ties). As a result, the total number of nodes in the network remains constant.

**Experiments & Results.** We generate synthetic networks with 5 000 nodes over 75 000 iterations with varying values for the maximum peer influence parameter ($q$). To ensure the formation of adequate community structures in the network we set $p_\Delta = 0.9$ and the probability for node-deletion to $p_d = 5{*}10^{-5}$. Further, we fix the parameter for memory strength to favor new ties ($c = 1$) and fix the critical peer influence threshold to $\theta = 0.1$ to reflect the intuition that a small number of active neighbors is sufficient to affect the activity of a node to a large extent. Finally, we run each configuration 40 times to account for statistical fluctuations and report average results.

Figure 1.6 illustrates these results. For the first few hundred iterations, the clustering coefficient ($C(t)$) is low but rapidly increases until it reaches its maximum between iteration $t = 3\,000$ and $t = 5\,000$. After this peak, it slowly declines until the network eventually reaches a stable state. Further, higher values for $q$ increase the speed at which the maximum is reached but also result in a lower average clustering coefficient once the network is stable. As the peer influence mechanism increases the activity in the network, especially in already formed communities, increases and active nodes motivate their neighbors to become more active.

**Findings.** Peer influence is an effective mechanism for the creation of synthetic activity networks. We present a model creating networks that exhibit similar community structures to real-world networks, such as triadic closures (three users all connected with each other) [70]. Further, we show that during the first few iterations the average clustering coefficient increases, indicating that during the early stages of a network, activity is concentrated on a core of highly active users. After reaching a peak activity starts to spread out more evenly throughout the system, indicated by a slow and steady decline of the average clustering coefficient.

For further details and an analysis of inter-event time distributions (*burstiness*), see the full paper on the topic by Wölbitsch et al. [75].

## 1.8 Conclusions

In this chapter, we asked the overarching question what factors govern growth and decline of activity in collaboration websites and how to evaluate their robustness and stability.

To this end, we presented and discussed various approaches to investigate a range of aspects influencing the user dynamics in collaboration websites. First, we used tests to assess the presence of complex user behavior by analyzing the nonlinearity of activity time series. Second, we presented a model based on dynamical systems, incorporating the concepts of loss of motivation (activity decay) and users affecting their neighbors (peer influence) to model and simulate activity in a collaboration website. Third, we introduced a modification to this model to simulate the impact of trolls (spreading negative peer influence). Fourth, we utilized activity potentials and peer influence in a generative model to create synthetic activity networks. Collectively, we summarize our key findings as follows.

**Complex user behavior.** Our results suggest that user activity varies across different collaboration websites with some communities exhibiting more signs of nonlinear behavior than others.

**Activity Decay & Peer influence.** We find that intrinsic activity decay. and peer influence serve as viable mechanics to capture and simulate activity in collaboration websites. Further, we can employ this peer influence to investigate the impact of troll-users on a system.

**Activity Potentials.** Lastly, we can extend the concept of user motivation through the mechanism of activity potentials and utilize this concept in combination with peer influence to generate synthetic activity networks that exhibit structures also present in their real-world counterparts.

The work we present in this chapter extends the body of existing research on dynamics in collaboration websites and may serve as a base for further research to predict the eventual success or failure of a collaboration website at an early stage. Finally, we demonstrated how the viability of an approach to analyze user dynamics in collaboration websites depends on the investigated aspect and the information available in the data.

## References

1. C. Zang, P. Cui, C. Faloutsos, W. Zhu, in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, 2017), pp. 565–574. DOI 10.1145/3097983.3098055
2. A. Baronchelli, M. Felici, V. Loreto, E. Caglioti, L. Steels, Journal of Statistical Mechanics: Theory and Experiment **2006**(06), P06014 (2006). DOI 10.1088/1742-5468/2006/06/P06014
3. K. Sznajd-Weron, J. Sznajd, International Journal of Modern Physics C **11**(06), 1157 (2000). DOI 10.1142/S0129183100000936
4. P.L. Krapivsky, S. Redner, Physical Review Letters **90**(23), 238701 (2003). DOI 10.1103/PhysRevLett.90.238701
5. A.C. Martins, International Journal of Modern Physics C **19**(04), 617 (2008). DOI 10.1142/S0129183108012339
6. A. De, I. Valera, N. Ganguly, S. Bhattacharya, M. Gomez-Rodriguez, in *Proceedings of the 30th International Conference on Neural Information Processing Systems* (Curran Associates Inc., 2016), pp. 397–405
7. K. Fan, W. Pedrycz, Physica A: Statistical Mechanics and its Applications **462**, 431 (2016). DOI 10.1016/j.physa.2016.06.110
8. C. Wang, arXiv preprint arXiv:1609.05732 (2016)
9. J. Leskovec, L.A. Adamic, B.A. Huberman, ACM Transactions on the Web (TWEB) **1**(1), 5 (2007). DOI 10.1145/1232722.1232727
10. J. Leskovec, L. Backstrom, J. Kleinberg, in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, 2009), pp. 497–506. DOI 10.1145/1557019.1557077
11. S.A. Myers, C. Zhu, J. Leskovec, in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, 2012), pp. 33–41. DOI 10.1145/2339530.2339540
12. A. Vespignani, Nature Physics **8**(1), 32 (2012). DOI 10.1038/nphys2160
13. J.L. Iribarren, E. Moro, Phys. Rev. Lett. **103**, 038702 (2009). DOI 10.1103/PhysRevLett.103.038702
14. P. Oliver, G. Marwell, R. Teixeira, American journal of Sociology pp. 522–556 (1985). DOI 10.1086/228313

15. P.E. Oliver, G. Marwell, American Sociological Review pp. 1–8 (1988). DOI 10.2307/2095728
16. G. Marwell, P.E. Oliver, R. Prahl, American Journal of Sociology **94**(3), 502 (1988). DOI 10.1086/229028
17. J. Solomon, R. Wash, Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM) (2014)
18. B. Suh, G. Convertino, E.H. Chi, P. Pirolli, in *WikiSym '09: Proceedings of the 5th International Symposium on Wikis and Open Collaboration* (ACM, Orlando, Florida, 2009), pp. 1–10. DOI 10.1145/1641309.1641322
19. A. Halfaker, R.S. Geiger, J.T. Morgan, J. Riedl, American Behavioral Scientist **57**(5), 664 (2013). DOI https://doi.org/10.1177/0002764212469365
20. J. Shin, in *Society for Information Technology & Teacher Education International Conference* (Association for the Advancement of Computing in Education (AACE), 2008), pp. 2834–2840
21. T. Schoberth, J. Preece, A. Heinzl, in *System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference On* (IEEE, 2003), pp. 10–pp. DOI 10.1109/HICSS.2003.1174576
22. D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, S. Suri, in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, 2008), pp. 160–168. DOI 10.1145/1401890.1401914
23. L. Mamykina, B. Manoim, M. Mittal, G. Hripcsak, B. Hartmann, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (ACM, 2011), pp. 2857–2866. DOI 10.1145/1978942.1979366
24. A. Furtado, N. Andrade, N. Oliveira, F. Brasileiro, in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (ACM, 2013), pp. 1237–1252. DOI 10.1145/2441776.2441916
25. C. Danescu-Niculescu-Mizil, R. West, D. Jurafsky, J. Leskovec, C. Potts, in *Proceedings of the 22Nd International Conference on World Wide Web* (Rio de Janeiro, Brazil, 2013), WWW '13, pp. 307–318. DOI 10.1145/2488388.2488416
26. J. Yang, K. Tao, A. Bozzon, G.J. Houben, in *User Modeling, Adaptation, and Personalization* (Springer, 2014), pp. 266–277. DOI 10.1007/978-3-319-08786-3_23
27. J. Zhang, M.S. Ackerman, L. Adamic, in *Proceedings of the 16th International Conference on World Wide Web* (ACM, 2007), pp. 221–230. DOI 10.1145/1242572.1242603
28. G. Wang, K. Gill, M. Mohanlal, H. Zheng, B.Y. Zhao, in *Proceedings of the 22nd International Conference on World Wide Web* (ACM, 2013), pp. 1341–1352. DOI 10.1145/2488388.2488506
29. R.M. Anderson, R.M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, USA, 1991)
30. Y. Matsubara, Y. Sakurai, B.A. Prakash, L. Li, C. Faloutsos, in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, 2012), pp. 6–14. DOI 10.1145/2339530.2339537
31. B. Ribeiro, in *Proceedings of the 23rd International Conference on World Wide Web* (Seoul, Korea, 2014), WWW '14, pp. 653–664. DOI 10.1145/2566486.2567984
32. S. Walk, D. Helic, F. Geigl, M. Strohmaier, ACM Transactions on the Web (TWEB) **10**(2), 11 (2016). DOI 10.1145/2873060
33. C. Zang, P. Cui, C. Faloutsos, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, 2016), pp. 2015–2024. DOI 10.1145/2939672.2939825
34. E. Bradley, H. Kantz, Chaos: An Interdisciplinary Journal of Nonlinear Science **25**(9), 097610 (2015). DOI 10.1063/1.4917289
35. J.P. Eckmann, S.O. Kamphorst, D. Ruelle, EPL (Europhysics Letters) **4**(9), 973 (1987). DOI 10.1209/0295-5075/4/9/004
36. N. Marwan, M.C. Romano, M. Thiel, J. Kurths, Physics reports **438**(5), 237 (2007). DOI 0.1016/j.physrep.2006.11.001

37. J.P. Zbilut, C.L. Webber Jr, Physics letters A **171**(3-4), 199 (1992). DOI 10.1016/0375-9601(92)90426-M
38. C.L. Webber Jr, J.P. Zbilut, Journal of applied physiology **76**(2), 965 (1994). DOI 10.1152/jappl.1994.76.2.965
39. D.A. Hsieh, The journal of finance **46**(5), 1839 (1991). DOI 10.1111/j.1540-6261.1991.tb04646.x
40. F. Strozzi, J.M. Zaldívar, J.P. Zbilut, Physica A: Statistical Mechanics and its Applications **312**(3), 520 (2002). DOI 10.1016/S0378-4371(02)00846-4
41. M. Small, C.K. Tse, Chaos: an interdisciplinary journal of nonlinear science **22**(3), 033150 (2012). DOI 10.1063/1.4753920
42. 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)
43. A. Barrat, M. Barthelemy, A. Vespignani, *Dynamical Processes on Complex Networks*, vol. 1 (Cambridge university press, 2008)
44. M. Allen, Higher Education, Sydney **336**, 108 (1999)
45. C. Hardaker, Journal of Politeness Research (2010). DOI https://doi.org/10.1515/jplr.2010.011
46. K. Bergstrom, First Monday **16**(8) (2011). DOI 10.5210/fm.v16i8.3498
47. E.E. Buckels, P.D. Trapnell, D.L. Paulhus, Personality and individual Differences **67**, 97 (2014). DOI 10.1016/j.paid.2014.01.016
48. F. Pasquale, The Chronicle of Higher Education (2015)
49. W. Broock, J.A. Scheinkman, W.D. Dechert, B. LeBaron, Econometric reviews **15**(3), 197 (1996). DOI 10.1080/07474939608800353
50. T. Teräsvirta, C.F. Lin, C.W. Granger, Journal of Time Series Analysis **14**(2), 209 (1993). DOI 10.1111/j.1467-9892.1993.tb00139.x
51. T.H. Lee, H. White, C.W. Granger, Journal of Econometrics **56**(3), 269 (1993). DOI 10.1016/0304-4076(93)90122-L
52. D.M. Keenan, Biometrika **72**(1), 39 (1985). DOI 10.1093/biomet/72.1.39
53. A.I. McLeod, W.K. Li, Journal of Time Series Analysis **4**(4), 269 (1983). DOI 10.1111/j.1467-9892.1983.tb00373.x
54. R.S. Tsay, Biometrika **73**(2), 461 (1986). DOI 10.1093/biomet/73.2.461
55. K.S. Chan, Journal of the Royal Statistical Society. Series B (Methodological) pp. 691–696 (1991)
56. A. Wald, J. Wolfowitz, The Annals of Mathematical Statistics **11**(2), 147 (1940). DOI 10.1214/aoms/1177731909
57. T. Schreiber, A. Schmitz, PhysicaD:NonlinearPhenomena **142**(3), 346 (2000). DOI 10.1016/S0167-2789(00)00043-9
58. E. Zivot, J. Wang, *Modeling Financial Time Series with S-Plus®*, vol. 191 (Springer Science & Business Media, 2007)
59. A. Trapletti, K. Hornik, *Tseries: Time Series Analysis and Computational Finance* (2016). R package version 0.10-35.
60. M. Friedman, The Annals of Mathematical Statistics **11**(1), 86 (1940). DOI 10.1214/aoms/1177731944
61. T. Santos, S. Walk, D. Helic, in *Proceedings of the 26th International Conference on World Wide Web Companion* (2017), pp. 1567–1572. DOI 10.1145/3041021.3051117
62. N.A. Christakis, J.H. Fowler, New England journal of medicine **358**(21), 2249 (2008). DOI 10.1056/NEJMsa0706154
63. S. Aral, D. Walker, Science **337**(6092), 337 (2012). DOI 10.1126/science.1215842
64. C. Wagner, S. Mitter, C. Körner, M. Strohmaier, Making Sense of Microposts (#MSM2012) p. 2 (2012)
65. K.E. Atkinson, *An Introduction to Numerical Analysis* (John Wiley & Sons, 2008)
66. P. Galán-García, J.G. de la Puerta, C.L. Gómez, I. Santos, P.G. Bringas, in *International Joint Conference SOCO'13-CISIS'13-ICEUTE'13* (Springer, 2014), pp. 419–428. DOI 10.1007/978-3-319-01854-6_43

67. C.W. Seah, H.L. Chieu, K.M.A. Chai, L.N. Teow, L.W. Yeong, in *Information Fusion (Fusion), 2015 18th International Conference on* (IEEE, 2015), pp. 792–799
68. E. Cambria, P. Chandra, A. Sharma, A. Hussain, (2010)
69. P. Koncar, S. Walk, D. Helic, M. Strohmaier, in *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017, Companion Volume* (2017). DOI 10.1145/3041021.3051116
70. G. Laurent, J. Saramäki, M. Karsai, The European Physical Journal B **88**(11), 301 (2015). DOI 10.1140/epjb/e2015-60481-x
71. A. Moinet, M. Starnini, R. Pastor-Satorras, Physical review letters **114**(10), 108701 (2015). DOI 10.1103/PhysRevLett.114.108701
72. M. Karsai, N. Perra, A. Vespignani, Scientific reports **4**, 4001 (2014). DOI 10.1038/srep04001
73. G. Bianconi, R.K. Darst, J. Iacovacci, S. Fortunato, Physical Review E **90**(4), 042806 (2014). DOI 10.1103/PhysRevE.90.042806
74. J.M. Kumpula, J.P. Onnela, J. Saramäki, K. Kaski, J. Kertész, Phys. Rev. Lett. **99**, 228701 (2007). DOI 10.1103/PhysRevLett.99.228701
75. M. Wölbitsch, S. Walk, D. Helic, in *International Workshop on Complex Networks and Their Applications* (Springer, 2017), pp. 353–364. DOI 10.1007/978-3-319-72150-7_29