Nonlinear Characterization of Activity Dynamics in Online Collaboration Websites

Tiago Santos
Know-Center
Graz, Austria
tsantos@know-center.at

Simon Walk
Stanford University
walk@stanford.edu

Denis Helic
Graz University of Technology
dhelic@tugraz.at

ABSTRACT

Modeling activity in online collaboration websites, such as StackExchange Question and Answering portals, is becoming increasingly important, as the success of these websites critically depends on the content contributed by its users. In this paper, we represent user activity as time series and perform an initial analysis of these time series to obtain a better understanding of the underlying mechanisms that govern their creation. In particular, we are interested in identifying latent nonlinear behavior in online user activity as opposed to a simpler linear operating mode. To that end, we apply a set of statistical tests for nonlinearity as a means to characterize activity time series derived from 16 different online collaboration websites. We validate our approach by comparing activity forecast performance from linear and nonlinear models, and study the underlying dynamical systems we derive with nonlinear time series analysis. Our results show that nonlinear characterizations of activity time series help to (i) improve our understanding of activity dynamics in online collaboration websites, and (ii) increase the accuracy of forecasting experiments.

Keywords

Nonlinear time series analysis; Q&A online communities

1. INTRODUCTION

Online Question and Answering portals, such as StackExchange or Quora, are immensely popular and helpful online resources with very large communities, amassing millions of users, questions and answers each[1]. However, while some online portals strive and blossom, the majority fails to attract users and never reaches critical mass, requiring them to shut down due to lack of activity, such as Google's knol project[2]. In this paper, we are motivated by the identification of key deciding features of activity time series, which hopefully will provide the foundation to distinguish successful and failing systems. In a first step towards this ambitious goal, we generalize the problem and apply several nonlinear time series analysis techniques to grasp and characterize hidden nonlinear behavior affecting activity dynamics. Current research on the study of dynamics governing such online collaboration websites focuses on model derivation with nonlinear, differential, parametric dynamical systems to describe observed data [14, 22]. However, such approaches are general purpose approaches, designed to fit observed data while minimizing the model's configuration effort to retain interpretability. In particular, these models do not address specificities of different websites or portals (e.g. StackExchange's Math vs. TeX portals) and do not aim to provide more than a general indicator for trends in activity of collaboration networks.

In this paper, we expand on existing work by conducting the following experiments on 16 randomly picked instances of the StackExchange portal: First, we categorize activity time series derived from online collaboration websites by the time series' likelihood to have stemmed from some hidden, nonlinear dynamical system. To that end, we use 9 statistical tests for nonlinearity to assess the adequateness of a nonlinear dynamical system to model activity. Then, we validate the plausibility of this categorization by comparing forecast performance from 3 standard time series models with nonlinear models, reconstructed from the observed activity time series. Finally, we present an exemplary study of nonlinearity properties of 2 datasets.

We find that activity in online collaboration websites may be modeled accurately by underlying, reconstructed dynamical systems to varying degrees, with some online collaboration websites showing more signs of nonlinear behavior than others. We use these differences to characterize the datasets and show how this knowledge may be used to not only improve activity modeling and forecasting efforts, but also better grasp datasets with nonlinear behavior by using tools from nonlinear time series analysis.

Our main contribution is therefore the improvement of the dynamical system modeling process for activity dynamics in online collaboration websites: Instead of postulating a "one-size-fits-all" dynamical system description via parametrized nonlinear equations, as done e.g. by Ribeiro [14] and Walk et al. [22], we reconstruct dynamical system descriptions directly from observed data and assess the feasibility of such a reconstruction. This allows us to tailor time series models to different data origins and thereby improve activity dynamics forecast quality. Furthermore, the use of nonlinear time

[1]See, for example, `http://stackexchange.com/sites?view=list#traffic`
[2]`http://knol.google.com/`

WWW 2017 Companion, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3051117

series analysis techniques, such as Recurrence Plots analysis, further boosts our understanding of nonlinear activity dynamics, for example through the identification of changes in stationarity or chaotic dynamics, leading to more model fine-tuning possibilities, which incorporate such information.

2. RELATED WORK

We review related work from the following two fields of research: nonlinear time series analysis applications and dynamical systems for networks.

Nonlinear Time Series Analysis and its Applications. Nonlinear time series analysis revolves around reconstructing a high dimensional dynamical system from an univariate time series, and studying the properties of the reconstructed dynamical system to derive knowledge on the original, univariate time series [1]. Nonlinear time series analysis enables studies on the *deterministic and chaotic*, rather than stochastic, nature of time series. Chaos means, in this sense, that small differences in a time series' present lead to great changes in its future, despite the dynamical system governing the time series' evolution being intrinsically deterministic.

Nonlinear time series analysis offers theoretical and practical tools to deal with reconstructed dynamical systems [12, 1], and these tools have found application in numerous areas [16, 7, 17]. In one of the most prominent applications of nonlinear time series analysis, Small and Tse [16] discuss how to predict the outcomes of a roulette wheel. A number of authors have also investigated the presence of chaotic behavior in financial markets, for example, by assessing nonlinearity in stock returns with statistical tests for chaos [7] or identifying events in stock returns with Recurrence Plots (RP) and Recurrence Quantification Analysis (RQA) [17].

For a detailed survey of the theory and application of RP and RQA we refer the interested reader to Bradley and Kantz [1] and Marwan et al. [12].

Dynamical Systems for Networks. Dynamical systems are a well studied topic from the standpoint of mathematics and physics [11, 6]. In general, dynamical systems provide mathematical descriptions on the evolution along the time dimension of a set of numeric quantities. They are employed to describe phenomena like the motion of a mass along some path according to Newton's laws, population growth or even macro-economic systems.

Application categories for dynamical systems on networks include, for example, *activity dynamics*. In the context of collaboration on the Web, activity dynamics apply dynamical systems on network theory to study the evolution of activity in different types of networks. The work by Ribeiro [14] introduces a dynamical system to model activity in membership-based community web pages, where activity is a time series representing the number of daily active users in such web pages. The author's model incorporates two main factors, namely web page users becoming spontaneously active and active users influencing inactive ones to become active. With this model, the author explains and predicts when a web page has reached a self-sustaining level of activity. More recently, Walk et al. [22] also applied dynamical systems theory to study activity dynamics in the context of collaboration networks, such as those arising in Question and Answering portals in the web. Here, the authors directly derive their key contributions from the activity dynamics model they propose, which include the self-sustaining level of activity for that type of collaboration network and the robustness of a collaboration network's activity.

In general, the authors of previously mentioned papers all propose a mathematical model, consisting of parametrized equations for a dynamical system, as a means to describe observed behavior. In contrast, we do not postulate parametrized equations describing a dynamical system on a network. Instead, we interpret the observed activity data, in the form of time series, as one dimensional projections of a hidden, complex and higher dimensional dynamical system. We study the feasibility of reconstructing the dynamical systems underlying the activity time series, characterize these activity time series by their propensity to have originated in such complex dynamical systems, and inspect the reconstructed dynamical system's properties.

3. METHODOLOGY

3.1 Forecasting univariate time series

Time series are sequences of numerical values (or observations), indexed and ordered by time. We consider discrete univariate time series, where each time index is uniquely associated with one observation. Moreover, we assume the time series observations are equally spaced in time.

Assessing nonlinearity in univariate time series. Not all univariate time series are equally suited for the reconstruction of a dynamical system; the presence of e.g. noise or randomness greatly influence the embedding. Therefore, we assess nonlinearity of univariate time series via the 9 following statistical tests: *Broock, Dechert and Scheinkman test* [2]; *Teraesvirta's neural network test* [19]; *White neural network test* [10]; *Keenan's one-degree test for nonlinearity* [9]; *McLeod-Li test* [13]; *Tsay's test for nonlinearity* [21]; *Likelihood ratio test for threshold nonlinearity* [4]; *Wald-Wolfowitz runs test* [4]; *Surrogate test - time asymmetry* [15].

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 [23, p. 652], we compute the test statistic of *Broock, Dechert and Scheinkman* on the residuals of an autoregressive integrated moving average (ARIMA) model, a class of linear models basing on auto regression, to check for nonlinearity not captured by the ARIMA model. For the *Wald-Wolfowitz runs* test, since a run represents a series of similar responses, we define a positive run as the amount of times the time series value was greater than the previous one [20].

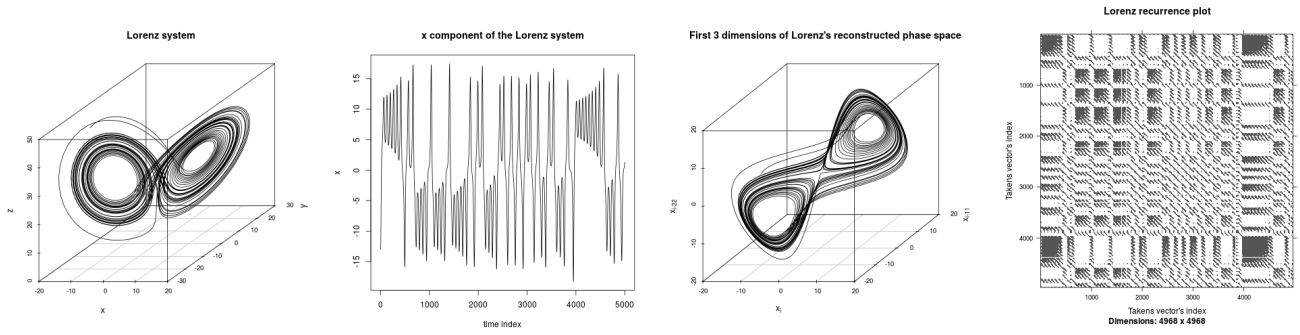
Reconstructing state space from univariate time series. Nonlinear time series analysis studies dynamical systems reconstructed from univariate time series. Takens [18] presents an embedding function, which, under certain conditions, maps an univariate time series to the higher dimensional phase space the reconstructed dynamical system lives in, and restores the topological characteristics of the dynamical system's reconstructed state space.

We briefly present theory on the embedding map required to reconstruct the state space of a dynamical system.

If x_t denotes the value of a time series x at time t, then an embedding of x can be obtained with a reconstruction vector of the form

$$R_t = (x_t, x_{t-\tau}, x_{t-2\tau}, \ldots, x_{t-(m-1)\tau}) \in \mathbb{R}^m. \qquad (1)$$

There are two free parameters in equation 1: τ and m. τ is the time lag, representing a distance in time between time



(a) Lorenz system has two attractors

(b) Time series of the Lorenz system's first component

(c) The reconstructed state space plot shows the attractors

(d) The recurrence plot of the Lorenz system reflects overall dynamics of the system

Figure 1: **Illustration of nonlinear time series analysis with the Lorenz dynamical system.** Figure 1a depicts the Lorenz system with parameters $\sigma = 10$, $\beta = 8/3$ and $\rho = 28$. We extract the Lorenz system's first component, shown in Figure 1b, and then reconstruct its state space with the embedding described in the embedding theory part of Section 3.1. The results of that embedding, with parameters $\tau = 11$ and $m = 4$, are the subject of picture 1c (showing only 3 dimensions). The reconstructed state space captures the original structure of the Lorenz system and its two attractors remarkably well. The structure of the Lorenz system can also be observed in the corresponding recurrence plot (RP) (see 1d). The RP shows the Lorenz attractors prominently around time indexes 1600 and 4200. Note also the large number of short diagonals around the main diagonal of the plot: These reflect the chaotic behavior of this Lorenz system.

series observations. m is the embedding dimension, i.e. the size of the vectors R_t in the space of the reconstructed dynamical system.

To estimate the embedding parameters, we start with the time lag τ. Bradley and Kantz [1] stress that τ should be large enough to encompass one full cycle of a time series' periodic dynamics. To estimate such a cycle's length (and thus τ too), the same authors propose different measures of independence between time series observations. We use the first minimum of average mutual information between observations as a measure of independence to estimate τ. The estimation of the embedding dimension m is an iterative process, which consists of computing some invariant of the reconstructed dynamical system for $m = 1, 2, \ldots$. We stop the process when the value of the invariant stabilizes, which indicates that the reconstructed dynamical system has been properly unfolded. We employ the commonly used iterative procedure [3] for the estimation of the embedding dimension.

Forecasts from linear models. To forecast an univariate time series, often used models include linear, ARIMA and ETS models.

In a linear model, a target variable is expressed as a linear combination of explanatory variables. We choose Fourier coefficients as explanatory variables, to account for seasonality effects of the type we encounter in the data described in 4.1.

The ARIMA class of models comprises auto-regressor models, which express the target univariate time series as a linear combination of its own past values and some lagged moving average error terms as well. This class of models assumes weak stationarity of the time series, so differencing—a technique to make a time series stationary—may be applied.

The ETS class of models includes exponential smoothing models, which—similarly to ARIMA—define the value of the target time series as a linear combination of lagged terms, such as level, trend, seasonality and error.

There are many variations of ARIMA and ETS time series models, and we automate the choice of model parameters and configurations with the algorithm devised by Hyndman and Khandakar [8].

Forecasts nonlinear models. Forecasts from nonlinear models require, first, the embedding map to reconstruct state space dynamics from the target time series, as described in the embedding theory part of Section 3.1. Given an embedding, nonlinear models forecast a target time series first by searching for nearest neighbor (with respect to the target time series) trajectories in the reconstructed state space. Then, the forecast from the nonlinear model is the arithmetic mean of future values of those near trajectories.

3.2 Recurrence Analysis

We analyze recurrences in reconstructed state space trajectories with Recurrence Plots (RPs), which give insights into both the behavior (e.g. stationary or drifting) and type (e.g. periodic, deterministic chaotic or random) of reconstructed dynamical systems, so we aim to use RPs to help with the nonlinear characterization of our data.

The RP is associated with a recurrence matrix—a square matrix which shows reconstructed state space trajectories $\vec{x}_i$ close to each other:

$$R_{i,j}(\epsilon) = \Theta(\epsilon - \|\vec{x}_i - \vec{x}_j\|), \tag{2}$$

where Θ is the Heaviside function and ϵ is the recurrence threshold establishing closeness between reconstructed state space trajectories $\vec{x}_i$. Thus, the RP is a scatter plot, simply showing points where the recurrence matrix is equal to 1. In Equation 2, we use the Euclidean norm and for ϵ we take the standard deviation of the distance matrix of all reconstructed state space trajectories.

Figure 1 shows an example of nonlinear time series analysis, complete with an RP characterization of the reconstructed state space. Starting with the standard chaotic Lorenz system, we extract its first component to reconstruct its state space and we analyze the reconstruction. We observe that the reconstructed system's topology accurately resembles the original system's one, and that the RP, with

Table 1: The table shows, per dataset, activity time series length in weeks, embedding parameters τ and m, nonlinearity test results (nonlin. tests), i.e. number and reference of statistical tests indicating nonlinearity with a significance level of 95% out of the 9 applied tests, normalized root mean squared error (RMSE) of a 1 year forecast per model. We also show the ranking the Friedman test assigns to the models' forecast RMSE for datasets with 5 or more tests indicating nonlinearity and the rest. Nonlinear models show best prediction performance on datasets with more than five statistical tests indicating nonlinearity.

Dataset	Weeks	τ	m	Nonlin. test score	Positive nonlin. tests	ARIMA	ETS	Linear	Nonlin.
english[b]	240	2	9	2/9	[2] [13]	0.6794	0.4452	0.3329	0.3080
unix[b]	239	1	7	2/9	[2] [13]	0.2091	0.2092	0.2418	0.2074
chemistry[b]	158	2	7	3/9	[2] [13] [4]	0.4982	0.2539	0.3247	0.4610
webmasters	244	1	8	3/9	[9] [13] [15]	0.2313	0.2528	0.3341	0.2346
chess	148	2	8	4/9	[2] [9] [13] [15]	0.2545	-[a]	0.5622	0.5110
history	177	1	9	4/9	[2] [9] [13] [4]	0.3503	0.2368	0.3044	0.4052
linguistics	181	2	6	4/9	[2] [9] [13] [15]	0.2512	0.2704	0.3009	0.3280
sqa	200	3	9	4/9	[2] [9] [13] [15]	1.8136	0.2531	0.6549	0.3903
tex[b]	241	1	7	4/9	[13] [21] [4] [15]	0.1589	0.1580	0.2767	0.2751
tridion	107	1	7	4/9	[19] [10] [9] [13]	0.2717	-[a]	0.6144	-[a]
Friedman test rank of models' forecast RMSE on datasets with nonlin. test score $< 5/9$						2	1	4	3
arduino	56	1	10	5/9	[2] [19] [10] [9] [13]	0.3489	-[a]	-[a]	-[a]
sports	159	1	7	5/9	[2] [9] [13] [4] [15]	0.2442	0.3348	0.4019	0.3323
ux	239	2	8	5/9	[2] [10] [9] [13] [21]	0.3479	0.1743	0.3491	0.1374
bitcoin	182	4	11	6/9	[2] [19] [10] [9] [13] [15]	0.6099	0.5549	0.5938	0.5781
math[b]	242	2	8	6/9	[2] [19] [13] [21] [4] [15]	0.1327	0.2314	0.3521	0.2912
bicycles	235	2	7	7/9	[2] [19] [10] [9] [13] [4] [15]	0.2971	0.3097	0.3252	0.2805
Friedman test rank of models' forecast RMSE 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.

its large number of small diagonals and its clusters of points depicting the Lorenz attractors, reflects the overall chaotic behavior of the Lorenz system.

4. EXPERIMENTAL SETUP

4.1 Datasets

For our analysis, we gathered data from 16 randomly picked StackExchange[3] questions and answers portals.

We follow the procedure described by Walk et al. [22] to derive univariate time series describing activity in these online collaboration websites: First, we measure a user's activity in such online collaboration websites as the user's number of questions, answers and comments per day. Then, we smooth these daily activities with a rolling mean over a 7 day window, to account for and remove outliers, and aggregate activity over all users per week. Finally, we require the weekly activity time series to have at least one unit of activity (i.e. one post, reply or comment) per day. This implied a burn-in of initial phases of inactivity or very low activity from the activity time series. For more details, see Table 1.

4.2 Predicting Activity

To assess if the activity time series show signs of nonlinear behavior, we apply all 9 statistical tests (described in the nonlinearity assessment part of Section 3.1) on the datasets (see 4.1), with a significance level of 95%. We then build a ratio, per dataset, of the number of tests indicating nonlinearity out of all 9 tests. That ratio serves as an indicator for hidden nonlinear dynamics, or not, in a given activity time series: We conjecture that higher values of that ratio will likely indicate hidden nonlinear dynamics, while datasets, which score lower on that ratio, are less likely to have such dynamics.

To test this approach for distinguishing time series with nonlinear behavior, we benchmark the performance of nonlinear forecasts on all datasets against those from other models. For datasets, characterized as nonlinear by the nonlinearity tests, we expect time series forecasts from nonlinear models to compare favorably against other models. The other models we benchmark nonlinear models against are linear, ARIMA and ETS models. For each of the datasets, we train those four models on a shorter version of the activity time series, excluding the last year of activity. We predict that last year with each of those 4 models and, finally, we compare the models' forecast results with the empirically observed values. We use the root mean squared error, normalized by the range of the activity time series, for the forecast performance comparison.

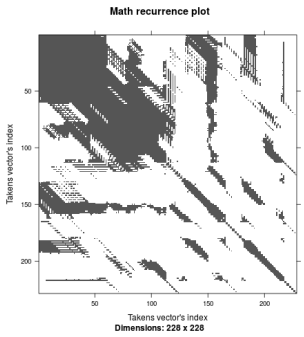
Since the nonlinearity tests (see nonlinearity assessment in Section 3.1) focus on the distinction between possibly chaotic determinism and randomness, activity time series with a strongly increasing (or decreasing) linear trend will be recognized as non-random. Strong linear trends may mask hidden nonlinear dynamics, which we aim to inspect.

Therefore, we first assess the strength of the trend of an activity time series by inspection of both the time series' plot and relative weight of a LOESS decomposition assigns to the trend component of that time series. For time series with a strong linear component, we estimate the linear trend with a simple linear regression, minimizing the weighted least squared error. Finally, we subtract that fitted linear trend from the time series, and perform nonlinearity tests and forecast computations on the detrended activity time series.

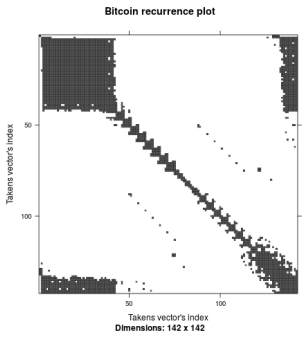
5. RESULTS & DISCUSSION

Findings on nonlinearity assessment and activity forecast models. We have listed all results of the nonlinear characterization via nonlinearity tests and activity forecast benchmarks in Table 1. The nonlinearity test scores indicate some disparity in the presence of nonlinear dynamics for the activity time series. Out of the 16 datasets we analyze, 6 datasets test five or more times positive for nonlinearity, and the other 10 datasets below five times. We interpret this split as an hint at some differences in nonlinear behavior of these datasets, and compare modeling and forecasting performance of the nonlinear, linear, ARIMA and ETS models for each of those two groups of activity time series.

[3] http://stackexchange.com/



(a) Recurrence Plot for the math activity time series



(b) Recurrence Plot for the bitcoin activity time series

Figure 2: **Recurrence Plots (RPs) give insights into activity time series dynamics.** Although both datasets, "Math" and "Bitcoin", have the same amount of statistical tests indicating nonlinearity, their RPs look quite different. The "Math" RP in figure 2a shows a higher density of recurrence points in the upper left corner, which gradually diminishes towards the lower right corner; this is a sign of a drift in the activity time series, still present after linear detrending [12]. Note both the diagonal as well as vertical structures present in Math's RP. The former, prominent around time indexes 100 to 175, could be a sign of chaotic dynamics, while the latter points towards states in the reconstructed state space which are (very) slowly changing. In contrast to Math, Bitcoin's RP in Figure 2b prominently features one strong main diagonal, with some remarkable periodicity around it. Another interesting aspect of Bitcoin's RP are the white bands around the main diagonal and the cluster of recurrence points in the lower left (and by symmetry of the RP also upper right) corner. These both hint at non-stationary transitions in the activity time series [12].

We assess the performance of these four models by calculating the normalized root mean squared error of a one year activity forecast with the Friedman test, as described by Demšar [5]. The Friedman test ranks nonlinear model performance highest for the group of datasets with more than five statistical tests indicating nonlinearity. In contrast, the nonlinear models only rank third for the other datasets, where less than five tests indicated nonlinear behavior. This result suggests a distinction in the degree of hidden nonlinear behavior in these activity time series.

We reason that activity time series, which were characterized as less likely to be driven by hidden nonlinear dynamics, were also better modeled by approaches other than nonlinear models due to their strong stochastic behavior. In such cases, we believe the role noise and external factors such as events play should not be underestimated.

The nonlinearity tests [10] and [19] appear to be more sensitive to the presence of nonlinear dynamics than other tests, since they test positive for nonlinearity 4 times more often in the dataset group with 5 or more tests indicating nonlinearity than in the other dataset group. Since [10] and [19] apply neural networks to assess linearity in mean, we attribute the usefulness of these two tests to the well-studied ability of neural networks to model nonlinear behavior.

We observe that the choice of appropriate models for activity dynamics should incorporate this characterization of activity time series according to evidence found for nonlinear behavior. Therefore, we find that a set of parametrized dynamical system equations to describe activity dynamics for all these StackExchange datasets at once, while easier to grasp and interpret, will likely fail to accurately reflect dataset specificities and thus perform poorer overall than the tailoring of time series models and reconstruction, where appropriate, of nonlinear dynamical system descriptions of the observed data.

Recurrence Plot analysis. Due to limitations in space, we perform an exemplary RP analysis on two activity time series. In Figure 2, the RPs of the datasets "Math" and "Bitcoin", two datasets with 6 statistical tests indicating nonlinearity, suggest differences in their underlying nonlinear dynamical systems, despite the apparent resemblance afforded by similar nonlinearity test results.

Math's RP shows, even after linear detrending, a drift pattern, which is conveyed by the reduction in recurrence point density from the RP's top-left to its bottom-right. We can observe other properties in Math's RP: There are some signs of chaotic behavior, apparent by the numerous short diagonals towards the lower-right corner and alongside the RP's main diagonal, and there are also some signs of slowly changing states in activity, as the long vertical line along time index 150 indicates. Armed with this knowledge we could tailor any type of time series model better to the data: The knowledge of drift enables us to introduce some parameter describing it. Slowly changing states transitioning to chaotic behavior suggest the choice of some threshold model, addressing those characteristics separately.

The main features of Bitcoin's RP are the periodically repeating structures around the main diagonal, the prominent white bands around the main diagonal and the point cluster in the lower left corner (and, by symmetry of the RP, in the upper right corner too). The latter two features indicate strong stationarity changes, while the regularity along the main diagonal hints at deterministic behavior. Again, these observations help with activity dynamics model design: We could introduce some periodic component to address the observed regularities, and we could include some exogenous variable to deal with the stationarity affecting events indicated by the RP's point clusters and white bands.

6. CONCLUSIONS & FUTURE WORK

We set out to explore a new and important issue on modeling activity dynamics: to recognize and characterize different online collaboration websites by the plausibility of

hidden nonlinear dynamical systems governing them, and thereby understand, model and forecast them better.

To address these open issues, we proposed using 9 different statistical tests for the nonlinear characterization of activity time series, and to validate this characterization with a comparison of the performances of different forecasting models. We also provided a sample RP analysis of activity time series characterized as nonlinear, to showcase the utility of these methods.

Our results can be summarized as follows. Firstly, a characterization of nonlinearity in activity time series by statistical tests gauges the plausibility of an activity time series being accurately described by dynamical systems (in contrast to, for example, some stochastic process), thus influencing model choice and helping discern driving forces of activity in our datasets. Secondly, nonlinear models seem adequate for forecasting activity time series, deemed nonlinear by statistical tests, more so than classical forecasting models (and vice-versa), a distinction which improves overall activity dynamic forecast quality. Thirdly, nonlinear modeling enables, via Recurrence Plots, a more granular study and deeper understanding of nonlinear dynamics governing activity time series, allowing for finer customization of time series models to explain activity in online collaboration websites.

This paper's limitations are a direct consequence of those of nonlinear time series analysis and the Friedman test's conservative estimations: Less noise, longer time series and more datasets should make results more conclusive.

With the hope of understanding *why* we see the observed activity dynamics, we believe that one of the most promising avenues for future work on nonlinear analyses of activity dynamics to be the connection between network science and the reconstructed dynamical systems. We speculate that hidden connections between statistics on these reconstructed dynamical systems, given for example by Recurrence Quantification Analysis, and properties of the underlying collaboration networks of websites will deliver further insights into the dynamic processes driving activity.

7. ACKNOWLEDGMENTS

The work has been performed in the project Power Semiconductor and Electronics Manufacturing 4.0 (SemI40), under grant agreement No 692466. The project is co-funded by grants from Austria, Germany, Italy, France, Portugal and - Electronic Component Systems for European Leadership Joint Undertaking (ECSEL JU).

8. REFERENCES

[1] E. Bradley and H. Kantz. Nonlinear time-series analysis revisited. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(9):097610, 2015.

[2] W. Broock, J. A. Scheinkman, W. D. Dechert, and B. LeBaron. A test for independence based on the correlation dimension. *Econometric reviews*, 15(3):197–235, 1996.

[3] L. Cao. Practical method for determining the minimum embedding dimension of a scalar time series. *Physica D: Nonlinear Phenomena*, 110(1):43–50, 1997.

[4] K. S. Chan. Percentage points of likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 691–696, 1991.

[5] J. Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine learning research*, 7(Jan):1–30, 2006.

[6] J. Guckenheimer and P. J. Holmes. *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, volume 42. Springer Science & Business Media, 2013.

[7] D. A. Hsieh. Chaos and nonlinear dynamics: application to financial markets. *The journal of finance*, 46(5):1839–1877, 1991.

[8] R. Hyndman and Y. Khandakar. Automatic time series forecasting: The forecast package for r 7. *URL: https://www.jstatsoft.org/article/view/v027i03 [accessed 2016-02-24]*, 2007.

[9] D. M. Keenan. A tukey nonadditivity-type test for time series nonlinearity. *Biometrika*, 72(1):39–44, 1985.

[10] T.-H. Lee, H. White, and C. W. Granger. Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests. *Journal of Econometrics*, 56(3):269–290, 1993.

[11] D. G. D. G. Luenberger. Introduction to dynamic systems; theory, models, and applications. Technical report, 1979.

[12] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths. Recurrence plots for the analysis of complex systems. *Physics reports*, 438(5):237–329, 2007.

[13] A. I. McLeod and W. K. Li. Diagnostic checking arma time series models using squared-residual autocorrelations. *Journal of Time Series Analysis*, 4(4):269–273, 1983.

[14] B. Ribeiro. Modeling and predicting the growth and death of membership-based websites. In *Proceedings of the 23rd international conference on World Wide Web*, pages 653–664. ACM, 2014.

[15] T. Schreiber and A. Schmitz. Surrogate time series. *PhysicaD:NonlinearPhenomena*, 142(3):346–382, 2000.

[16] M. Small and C. K. Tse. Predicting the outcome of roulette. *Chaos: an interdisciplinary journal of nonlinear science*, 22(3):033150, 2012.

[17] F. Strozzi, J.-M. Zaldívar, and J. P. Zbilut. Application of nonlinear time series analysis techniques to high-frequency currency exchange data. *Physica A: Statistical Mechanics and its Applications*, 312(3):520–538, 2002.

[18] F. Takens. Detecting strange attractors in turbulence. In *Dynamical systems and turbulence, Warwick 1980*, pages 366–381. Springer, 1981.

[19] T. Teräsvirta, C.-F. Lin, and C. W. Granger. Power of the neural network linearity test. *Journal of Time Series Analysis*, 14(2):209–220, 1993.

[20] A. Trapletti and K. Hornik. *tseries: Time Series Analysis and Computational Finance*, 2016. R package version 0.10-35.

[21] R. S. Tsay. Nonlinearity tests for time series. *Biometrika*, 73(2):461–466, 1986.

[22] S. Walk, D. Helic, F. Geigl, and M. Strohmaier. Activity dynamics in collaboration networks. *ACM Transactions on the Web (TWEB)*, 10(2):11, 2016.

[23] E. Zivot and J. Wang. *Modeling financial time series with S-Plus®*, volume 191. Springer Science & Business Media, 2007.