PROCEEDINGS A

royalsocietypublishing.org/journal/rspa

Research

Cite this article: Rodrigues FA, Peron T, Connaughton C, Kurths J, Moreno Y. 2025 A machine learning approach to predicting dynamical observables from network structure. *Proc. R. Soc. A* **481**: 20240435. https://doi.org/10.1098/rspa.2024.0435

Received: 16 June 2024
Accepted: 02 October 2024

Subject Category:
Physics

Subject Areas:
complexity

Keywords:
complex networks, machine learning, dynamical processes

Author for correspondence:
Francisco A. Rodrigues
e-mail: francisco@icmc.usp.br

A machine learning approach to predicting dynamical observables from network structure

Francisco A. Rodrigues[1,2], Thomas Peron[1], Colm Connaughton[2,3], Jürgen Kurths[4,5] and Yamir Moreno[6,7,8]

[1]Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, SP, Brazil
[2]Mathematics Institute, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK
[3]London Mathematical Laboratory, London, UK
[4]Potsdam Institute for Climate Impact Research, 14473 Potsdam, Germany
[5]Department of Physics, Humboldt University, 12489 Berlin, Germany
[6]Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, 50018 Zaragoza, Spain
[7]Department of Theoretical Physics, University of Zaragoza, 50018 Zaragoza, Spain
[8]CENTAI Institute, Turin, Italy

FAR, 0000-0002-0145-5571; YM, 0000-0002-0895-1893

Estimating the outcome of a given dynamical process from structural features is a key unsolved challenge in network science. This goal is hampered by difficulties associated with nonlinearities, correlations and feedbacks between the structure and dynamics of complex systems. In this work, we develop an approach based on machine learning algorithms that provides an important step towards understanding the relationship between the structure and dynamics of networks. In particular, it allows us to estimate from the network structure the outbreak size of a disease starting from a single node, as well as the degree of synchronicity of a system made up of Kuramoto oscillators. We show which topological features of the network are key for this estimation and provide a ranking of the importance of network metrics with much higher accuracy than previously done. For epidemic propagation, the k-core plays a fundamental role, while for synchronization, the betweenness centrality and accessibility are the measures most related to the state of an oscillator.

THE ROYAL SOCIETY PUBLISHING

For all the networks, we find that random forests can predict the outbreak size or synchronization state with high accuracy, indicating that the network structure plays a fundamental role in the spreading process. Our approach is general and can be applied to almost any dynamic process running on complex networks. Also, our work is an important step towards applying machine learning methods to unravel dynamical patterns that emerge in complex networked systems.

1. Introduction

Modern network science has been successful at showing that accounting properly for the interaction patterns of a system's components is crucial for describing its functionality [1–5]. Admittedly, tackling the laws that govern the relationship between the structure and function of a system is a formidable challenge. Achieving such a goal would entail not only being able to assess the impact of some structural patterns in the dynamics of networks but also to predict dynamical outcomes from an incomplete, and often noisy, knowledge of the structure of a system. The problem is not a minor one: going from knowing the structure of a system to anticipating its dynamical response implies the need to sort out nonlinearities in the nodes' responses, spatial and temporal correlations due to complex interconnection patterns and feedback resulting from these interactions, among other difficulties.

It is also well known that network properties affect dynamical systems in different ways. In other words, some properties affect a dynamic process more than others. For instance, in the context of disease spreading, the degree distribution is crucial for the critical properties of the system—that is, whether there is a vanishing threshold or not—whereas correlations play a less determinant role [5,6]. Similar conclusions can be extracted for other dynamical processes, such as rumour spreading [7,8] and synchronization phenomena [4,9]. Tracking down which network properties are fundamental for the dynamics of the system will not only allow for more accurate dynamical predictions but also to provide actionable responses to topological changes or control interventions to drive the system to a desired global outcome. Applications range from identifying influential and core spreaders [10] in disease dynamics to tuning the level of synchronization of power grids, electronic circuits and neuronal systems [4].

Here, we address the two challenges described above. First, we develop a methodology that allows the prediction of several macroscopic observables for two paradigmatic dynamics: disease contagion and synchronization. In doing so, we also report on what are the most determinant topological properties for such a prediction to be accurate. In particular, we present a general approach for predicting variables related to dynamical processes in complex networks, namely the degree of synchronization of Kuramoto oscillators and the outbreak size in a susceptible-infected-recovered (SIR) dynamics. We also assess the importance of network properties in predicting such dynamical variables. For example, we verify that the k-core has the strongest relationship with the outbreak size, confirming previous results. However, this contribution alone does not allow an accurate prediction. We show that there is no single measure to identify such main propagators, but a combination of them. In the case of synchronization, betweenness centrality and accessibility are most related to the state of an oscillator, i.e. partially (or globally) phase locked or unlocked. However, the ranking of the metrics' importance depends on the level of synchronization.

Our method is general and could be applied to other scenarios in which the aim is to predict a random dynamic variable from a subset of nodes and their dynamics (e.g. [8,11]). The methodology proposed here opens new paths to investigate the structure and dynamics of complex systems through modern methods of machine learning.

2. Machine learning model formulation

To formulate our framework, we depart from traditional methods of nonlinear dynamics and statistical mechanics [12] and instead use a machine learning (ML) approach. To specify the predictive learning model, we associate a dynamic variable Y_i to each node i. This variable is time dependent, but can reach a constant value for stationary processes, $Y_i(t \to \infty) = Y_i$. In the case of synchronization, Y_i can be an indicator variable such that $Y_i = 1$ if oscillator i is synchronized and 0 otherwise. In the case of disease spreading, we define Y_i as the expected fraction of infected nodes when the disease is seeded at node i, i.e.

$$Y_i = \lim_{t\to\infty} \frac{1}{N} \sum_{j=1}^{N} Z_j(t), \tag{2.1}$$

where $Z_j = 1$ if node j is infected and $Z_j = 0$, otherwise. Note that we consider all N nodes in the network, as the disease has the potential to spread to any node $j, j = 1, 2, \ldots, N$. The variable Y_i depends on the structure of the network, since the state of each node i is defined by the interaction with its neighbours, which can be infected or not. We can therefore assume that Y_i is related to a feature vector extracted from the network structure, $\mathbf{X_i} = \{X_{i1}, X_{i2}, \ldots, X_{id}\}$, since

$$Y_i = f(\mathbf{X_i}) + \epsilon, \tag{2.2}$$

where ϵ is a random error term independent of $\mathbf{X_i}$ and normally distributed with mean zero and standard deviation σ. The feature vector $\mathbf{X}$ represents measures that describe structural properties, which should be carefully chosen to predict the variable Y_i with as high an accuracy as possible. Although the selection of the elements in $\mathbf{X_i}$ can be done using methods for model comparison and feature selection algorithms, here we consider basic network metrics. Specifically, we include (i) degree (K), (ii) clustering coefficient (C), (iii) closeness centrality (CC), (iv) betweenness centrality (B), (v) eigenvector centrality (EC), (vi) PageRank (PR), (vii) k-core (KC), and (viii) accessibility (ACC). Note that, except for the clustering coefficient, we consider measures that essentially quantify the centrality of the nodes from different perspectives [13].

The fixed but unknown function $f : \Re^d \to \Re$ in equation (2.2) represents a surface in a d-dimensional space, where d is the number of network measures in $\mathbf{X_i}$. Its estimation is a complex high-dimensional problem that cannot be solved using traditional statistical methods because the observations (extracted from nodes) are not independent. In terms of statistical learning, f can be estimated by inference or prediction [12,14]. In the case of inference, the goal is to understand the relationship between Y_i and $\mathbf{X_i}$, determining, for instance, the contribution of each network measure j, X_{ij}, to the dependent variable Y_i. This kind of analysis has been applied to study synchronization of Kuramoto oscillators using linear regression [8]. Prediction aims at finding an estimation $\widehat{Y}_i$ that can best predict Y_i [14]. In this case, we are not interested in the interpretability of the model, but only in making an accurate prediction. In this work, we are interested in prediction and inference. In the case of epidemic spreading, we have a regression problem since $\{Y_i \in \Re | 0 \le Y_i \le 1\}$, $i = 1, \ldots, N$. On the other hand, for synchronization dynamics, we have a classification problem, given that in such a case the goal is to classify oscillators as partially (or globally) phase-locked or unlocked.

As we are not interested in the exact form of the function f, but in optimum prediction of Y_i, we consider a machine learning method [15] for the estimation of the statistical model given by equation (2.2). To this end, we need to specify a set of training examples [12], so that the machine learning algorithm learns from data [16,17], as well as a cost function, e.g. the minimization of the mean square error, $E = [Y_i - \widehat{f}(X_i)]^2/n$ [16], where $\widehat{f}(X_i) \approx Y_i$, for the regression. In the classification problem, the cost function is given by the fraction of observations classified correctly. In what follows, we show results obtained for the two dynamical processes mentioned before applying two machine learning algorithms, namely, deep neural networks and random forests [14,16]. We stress again that our interest is in proposing a methodology that addresses the structure-dynamics dependency on networks, instead of a comparison of what is the most accurate machine learning algorithm one could adopt.

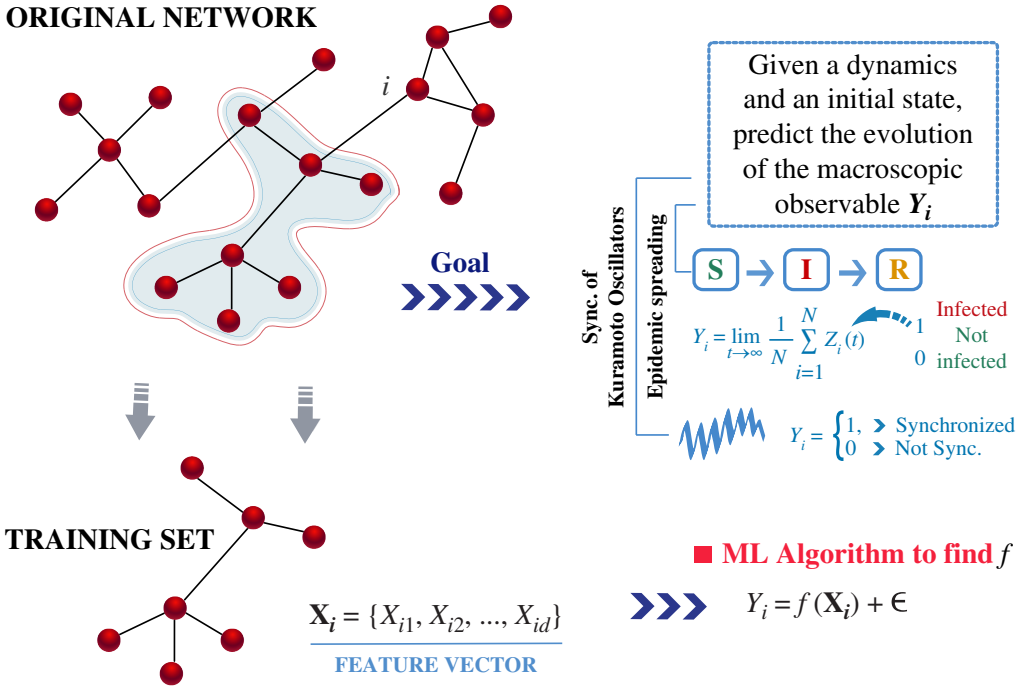

Figure 1. Schematic representation of the methodology employed in this paper. A subset of nodes of a given network is used to train an ML algorithm that makes predictions about the dynamics of the entire system, for disease spreading and synchronization.

3. Results and discussion

To perform the classification and regression, we define a set of nodes in the training set, used to infer the function f in equation (2.2). The remaining vertices that do not belong to the training set are used to evaluate the accuracy of the prediction. Since the results depend on the choice of the test and training sets, we consider the nested cross-validation approach [15]. In this case, the data are divided into $k_{out} = 10$ equally sized sets, called folds, such that one fold is used for testing and the remaining ones for training. The training set is divided again into $k_{in} = 10$ folds, where one fold is used for validation and the remaining ones for training. By changing the roles (testing or training) of the folds, we evaluate the model performance by averaging all possible configurations. Particularly, for the synchronization dynamics, it being a classification problem, we implement a single stratified k-folds cross-validator with $k = 10$ folds. The stratified k-folds cross-validator returns stratified folds, preserving the percentage of samples for each class. The general approach developed here is summarized in figure 1. Next, we present our results for the two dynamical processes here considered.

(a) Epidemic spreading

The prediction of when, how and to which extent an epidemic outbreak will take place is one of the most important challenges of modern science with fundamental implications for public health [18]. For instance, the prediction of the expected number of cases once a disease is seeded in a given individual or group of subjects would enable the identification of the most influential spreaders [7,10,19] and contribute to develop methods for an efficient disease control via vaccination or other procedures. Specifically, we consider an SIR model whose dynamics has an absorbing state, i.e. the number of infected nodes goes to zero in a finite time [3,18]. We

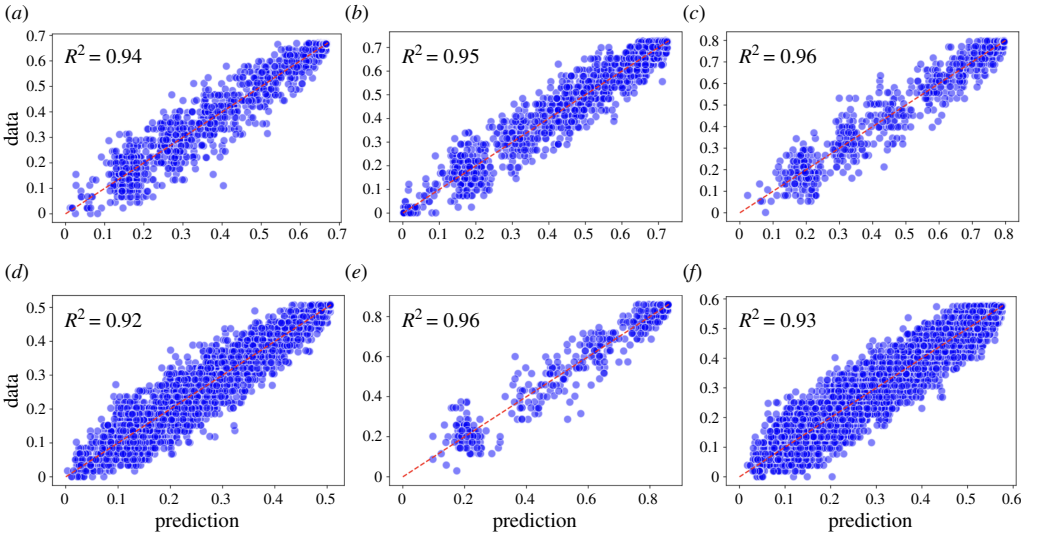

Figure 2. Prediction of the outbreak size in terms of network structure for the (*a*) US air transportation network, (*b*) Hamsterster social network, (*c*) political blogs network, (*d*) Bitcoin Alpha network, (*e*) email data from a large European research institution and (*f*) Gnutella peer-to-peer file sharing network. The SIR model is simulated using $\beta = 0.25$ and $\mu = 1$. The dashed line represents $x = y$. The associated coefficient of determination is included in each plot.

conjecture that the outbreak size, Y_i (equation (2.1)), depends on the topological properties as given by equation (2.2), where we consider that the feature vector $\mathbf{X_i}$ contains the structural metrics listed before (i.e. properties (i) through (vii)). To infer f in equation (2.2), we use both deep neural networks and random forests and predict the average size of outbreaks starting from each node i, $i = 1, 2, \ldots, N$.

To validate our approach, we use spreading data obtained from Monte Carlo simulations of the model with discrete time [20]. The process starts at one seeded node i, $i = 1, \ldots, N$, and at each time step, each infected node contacts all its neighbours and transmits the disease with probability β. Infected nodes recover with probability μ and the dynamics stops when there are no infected nodes in the network. In our simulations, we consider $\mu = 1$ and $\beta = 0.25$. The selection of the training and testing sets is done by the nested cross-validations approach, as described before. Note that the observations used for testing are not considered in the training phase, so that the prediction is performed in unknown examples.

Figure 2 shows the outbreak size predicted by the model, equation (2.2), for the (i) US air transportation network (only the main airports, $N = 500$ airports in 2002, and $\langle k \rangle = 11.9$ edges per airport) [21], (ii) Hamsterster social network ($N = 1788$ users and $\langle k \rangle = 13.95$ edges per user) [22], (iii) political blogs network ($N = 1222$ blogs and $\langle k \rangle = 27.4$ hyperlinks per blog) [23], (iv) Bitcoin Alpha network ($N = 3775$ users and $\langle k \rangle = 7.5$ connections per user) [24], (v) email data from a large European research institution ($N = 986$ users and $\langle k \rangle = 32.6$ connections per user) [25], and (vi) Gnutella peer-to-peer file sharing network ($N = 6299$ computers and $\langle k \rangle = 6.6$ links per computer) [25]. For all the networks, we find that random forests can predict the outbreak size for each node with high accuracy, indicating that the network structure plays a fundamental role in the spreading process.

We also compare the random forest approach with the traditional multivariate linear regression model. As we can see in figure 3, for different values of transmission probability β, the random forest provides more accurate results, in terms of the coefficient of determination R^2, which quantifies how well the independent variables predict the variance in the dependent variable. It is given by Pearson's square coefficient, ranging from 0 to 1, where higher values indicate a better fit. The deep neural network provides similar results to the random forest

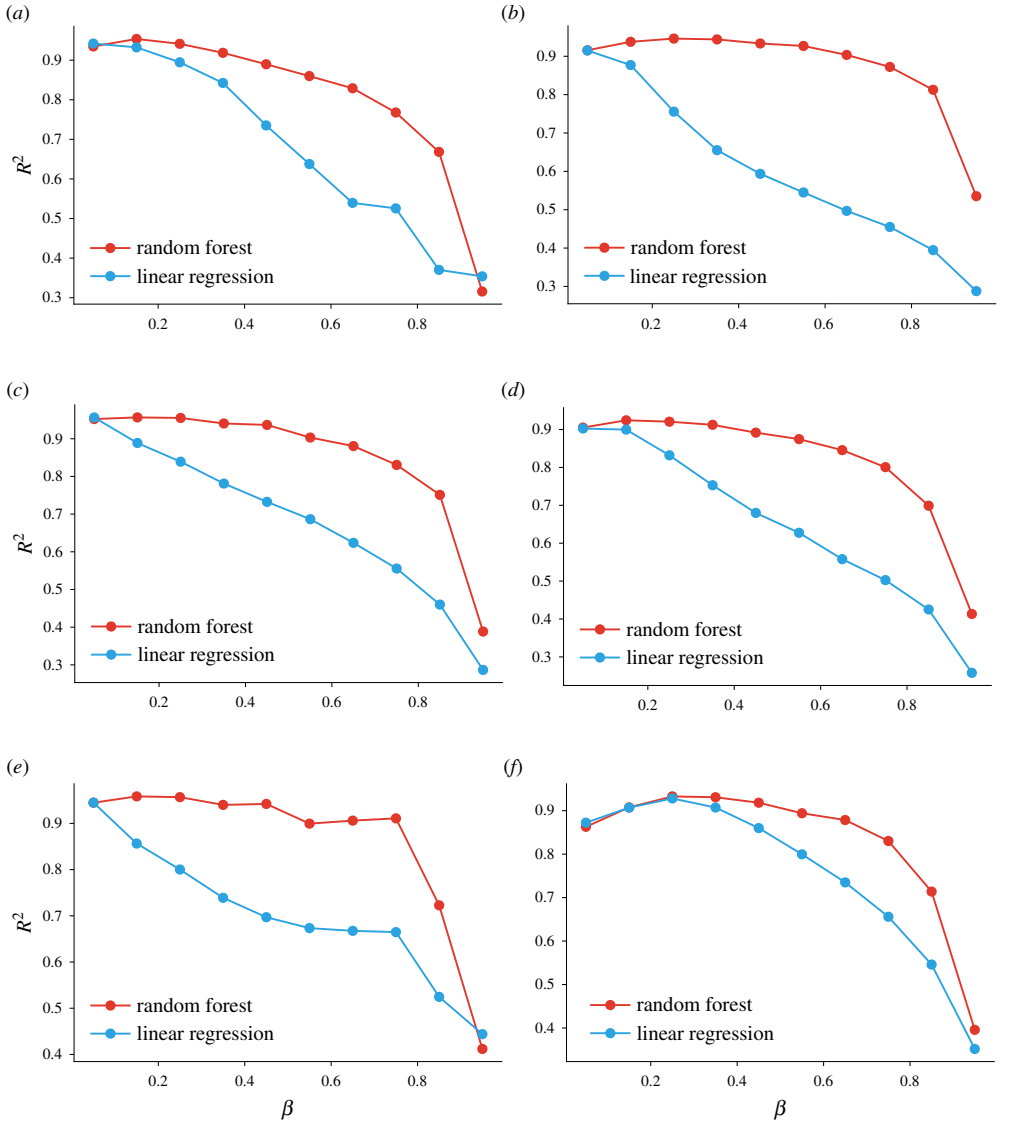

Figure 3. Comparison of the random forest and linear regression in terms of the coefficient of determination (R^2) for the (*a*) US air transportation network, (*b*) Hamsterster social network, (*c*) political blogs network, (*d*) Bitcoin Alpha network, (*e*) email data from a large European research institution and (*f*) Gnutella peer-to-peer file sharing network.

algorithm (data not shown here). Therefore, our findings reveal that the information contained in the vector $\mathbf{X_i}$ is suitable to predict Y_i with high accuracy and that machine learning models are necessary for an accurate prediction.

We also observe in figure 3 that the forecasting is more accurate for small values of infections probability β, which suggests that it is easier to predict the spreading capacity of each node for diseases with small spreading probabilities. This is expected, since if β is high, then most of the nodes tend to infect all their neighbours, which reduces the influence of the network topology.

It is also important to point out that although random forests and neural networks provide similar prediction errors, the random forest presents some advantages compared with neural networks. In particular, it has only a few parameters to be tuned, like the number of trees (we used 100 trees), contrariwise to the neural network, in which we have to set the number of neurons in each layer, the number of layers, the activation function and the solver for weight

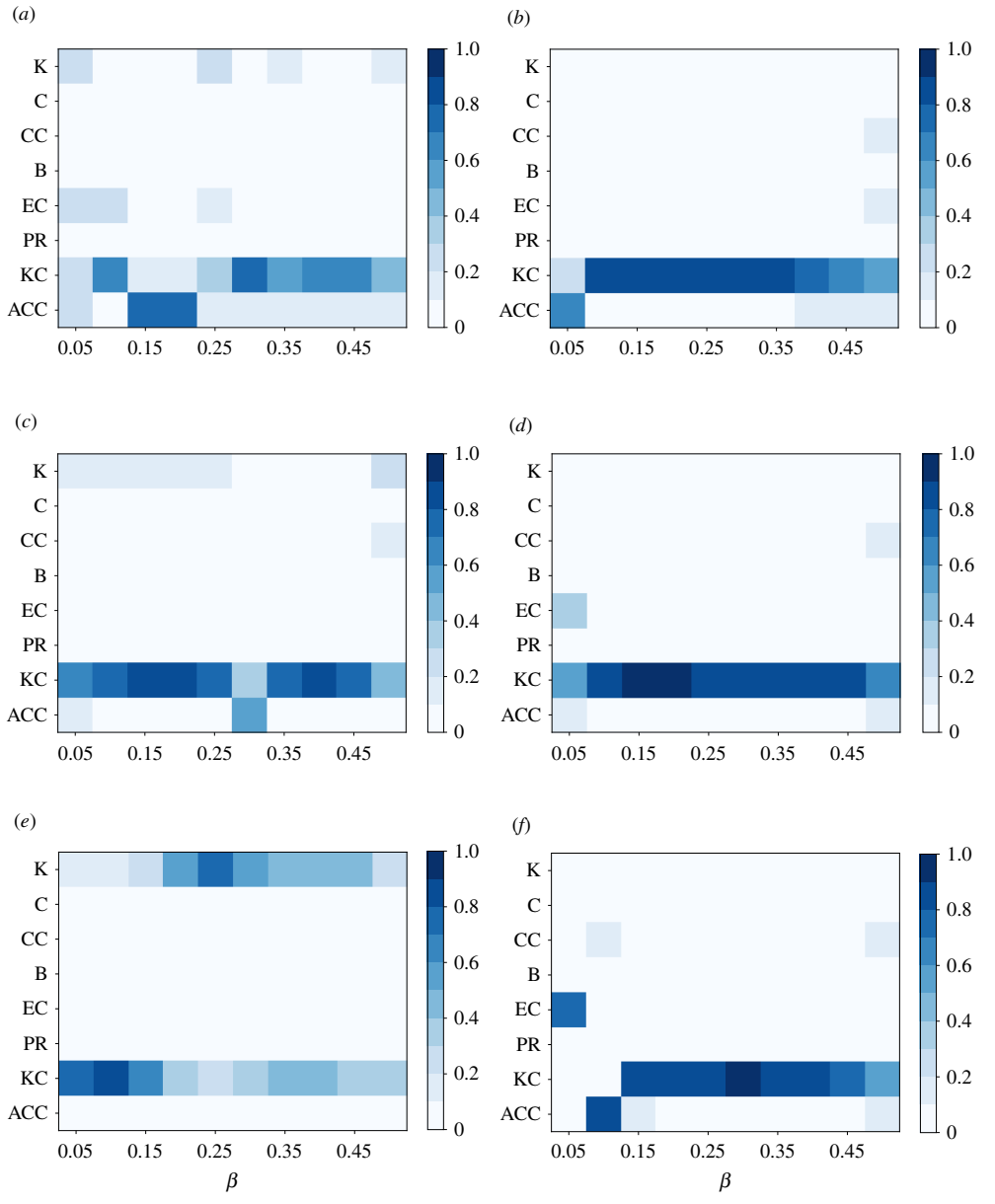

Figure 4. Importance of network measures on the prediction of the outbreak size for the (*a*) US air transportation network, (*b*) Hamsterster social network, (*c*) political blogs network, (*d*) Bitcoin Alpha network, (*e*) email data from a large European research institution and (*f*) Gnutella peer-to-peer file sharing network. The measures are the degree (*K*), clustering coefficient (*C*), closeness centrality (*CC*), betweenness centrality (*B*), eigenvector centrality (*EC*), PageRank (*PR*), k-core (*KC*) and accessibility (*ACC*).

optimization [26]. Importantly, a random forest algorithm also enables the quantification of the features' importance. Basically, the importance is computed by measuring how effective the feature is at reducing uncertainty when creating the decision trees. All the values provided by the random forest sum to one and give the percentage of contribution of each measure on the prediction of the estimated quantity.

Figure 4 displays the results obtained when the estimation of the outbreak size is broken down as a function of the contribution of each of the network metrics used in the feature vector. We can

see that the importance of each measure depends on the infectious probability β. Moreover, for most of the networks, the k-core is the most determinant feature to predict the outbreak size. This result agrees with previous analysis of influential spreaders in networks [7]. Also, the accessibility plays an important role in the prediction of spreading power for the US air transportation network and Gnutella network, as verified in [10]. The contribution (influence) of each metric strongly depends on the kind of network. For instance, for the email network, the degree is as relevant as the k-core, which is not the case for the Hamsterster and political blog networks, in which the k-core plays a fundamental role.

Therefore, there is no unique measure to identify such main propagators, but a combination of them. This influence of network properties also depends on the spreading probability β. Thus, our analysis is more general than those presented in previous works (e.g. [7,10]) and substantially improves the identification of the most influential spreaders. Our approach allows for a comparison of the influence of as many network properties as desired concurrently, i.e. when their possible role is considered altogether and they are combined to make the prediction better.

(b) Synchronization

The Kuramoto model (KM) is a paradigmatic model to study synchronization phenomena [4]. In complex networks, the KM is described by the set of equations,

$$\frac{d\theta_i}{dt} = \omega_i + \lambda \sum_{i=1}^{N} A_{ij} \sin(\theta_j(t) - \theta_i(t)), \tag{3.1}$$

where the oscillator i has the phase angle θ_i and the natural frequency ω_i, λ represents the overall coupling strength between the oscillators and A_{ij}, $i, j \in \{1, \ldots, N\}$, are the elements of the adjacency matrix, i.e. $A_{ij} = 1$ if oscillators i and j are connected and $A_{ij} = 0$, otherwise. The degree of coherence in the system is quantified by the order parameter, $Re^{i\psi(t)} = \frac{1}{N}\sum_{j=1}^{N} e^{i\theta_j}$ and synchronization emerges when $\lambda > \lambda_c$. For uncorrelated networks, the critical coupling has been shown to be $\lambda_c = \frac{2\langle k \rangle}{\pi g(\bar{\omega}) \langle k^2 \rangle}$ [27]. Thus, synchronization is expected to depend on the distribution of the natural frequencies, $g(\omega)$, and the network structure [4]. We define the state of an oscillator by considering a dummy variable as $Y_i = 1$ if $\lim_{t\to\infty} |\dot{\theta}_i(t) - \Omega(t)| < 1/\sqrt{N}$ and 0 otherwise, where $\Omega(t) = \dot{\psi}(t)$ is the mean-field frequency [4].

We predict the variable Y_i using the neural network of the worm *C. elegans* ($N = 297$, $\langle k \rangle = 14.4$) [28]. The value of the coupling strength determines the fraction of synchronized oscillators. Note that in this case, for very small values of λ, most oscillators are drifting, whereas, for large λ, almost all of them are synchronized. Thus, depending on the value of λ, we can have imbalanced classes and therefore the accuracy of the classifier cannot be measured by the fraction of correct classifications. We have calculated the receiver operating characteristic (ROC) curve and the corresponding area under the curve (AUC) measure [12]. For a classifier with no predictive power, i.e. random guessing, $AUC = 0.5$ and the ROC curve follows the diagonal. Figure 5 shows the AUC curve for the *C. elegans* network considering different coupling strengths, where λ_c is the critical coupling for the emergence of the synchronous state. We consider the random forests algorithm to perform the classification. We specifically use 100 trees, with squared error as the loss function and a minimum sample split criterion of two. The AUC is larger than 0.6 for all cases, reaching the maximum 0.99 for $3.5\lambda_c$. This confirms that the synchronization level of the whole system is predictable if we know the topology and the degree of synchronization of a fraction of oscillators. Moreover, as previously done, we identify quantitatively the importance of each network measure on the synchronization dynamics. In figure 6, we present the results of this analysis. The higher the weight of a given feature in the random forest approach, the more important is the feature to predict the local synchronization. Interestingly enough, the ranking of the

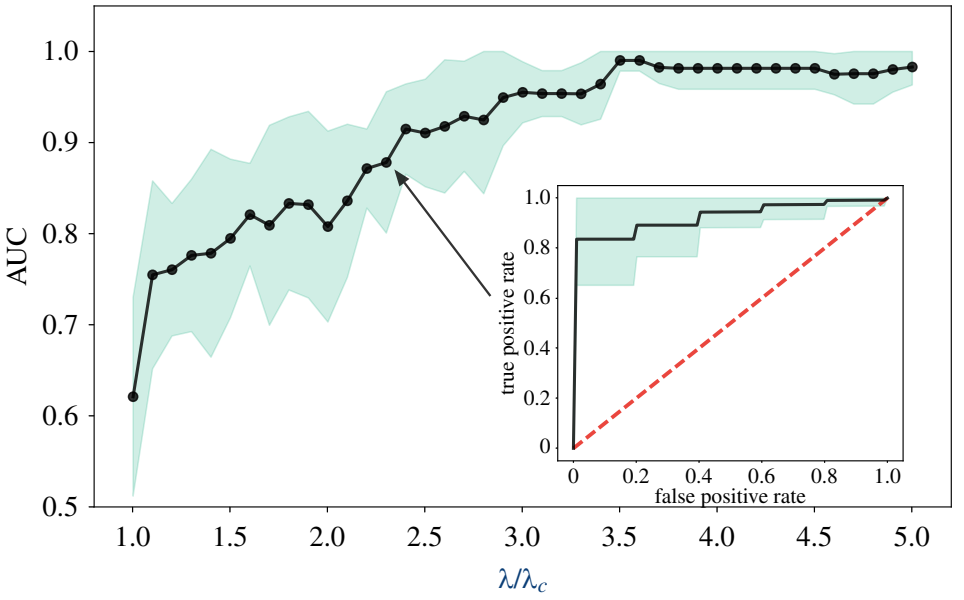

Figure 5. AUC in function of the coupling strength for the neural network of *C. elegans*. The shadow region shows the 95% interval that contains the average value. The random forests algorithm is considered to predict the state of the oscillators. The inside plot shows the ROC curve for $2.5\lambda_c$.

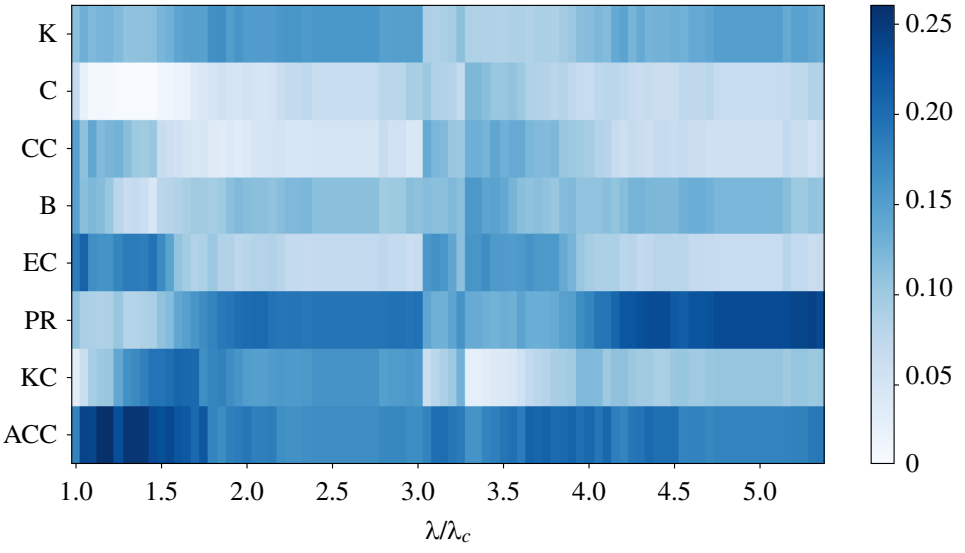

Figure 6. Measures' importance quantified by the random forest algorithm in the neural network of the worm *C. elegans*. The intensity of colour represents the importance evaluated through the random forest algorithm.

metrics' importance depends on the level of synchronization. For small coupling strength, the accessibility measure plays a key role in the prediction. However, as we increase λ, the PageRank centrality plays a more important role in the prediction of the nodes' states. Indeed, the PageRank and accessibility are the most important measures associated with the state of the oscillators.

4. Conclusion

In summary, we have shown that it is possible to estimate dynamical outcomes in complex networks by extracting features from a small number of nodes. The methodology presented here is general enough as to be accurate for two different dynamics, the spreading of diseases and the synchronization of coupled oscillators. Our results are relevant for the analysis of dynamical processes on networks and pave the way for predicting dynamics from structural knowledge.

Our work has important implications for the study of dynamic processes in complex networks and can be extended in a number of ways. For example, we can apply this methodology to the study of cooperation [29], rumour propagation or other epidemic models [30]. Furthermore, in temporal networks [31], this methodology can be used to predict future connections and the evolution of dynamic processes. In adaptive dynamic networks [32], it is possible to predict how the network will change according to the underlying dynamic process. It is also possible to explore cases where there is more than one dynamic process in the network while these processes interact [33,34]. For example, we can try to predict the effect of a rumour on the spread of epidemics, or of cooperation on the evolution of opinions. Finally, it is possible to use the methodology proposed here to study networks with higher order structures [35,36]. In this case, we can determine the influence of simplicial complexes on the evolution of dynamic processes.

In conclusion, using machine learning to study dynamic processes in complex networks can help to understand the relationship between structure and dynamics, so that changes in network structure can be used to control a dynamic process. This study is general and can be applied to any type of dynamic process or complex network, including temporal, multi-layer, adaptive and high-order structured networks.

Data accessibility. The data and codes are available in this repository: https://github.com/franciscorodrigues-usp/Predicting-epidemic-sync.

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. F.A.R.: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing—original draft, Writing—review & editing; T.P.: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing—original draft, Writing—review & editing; C.C.: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing—original draft, Writing—review & editing; J.K.: Conceptualization, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing—review & editing; Y.M.: Conceptualization, Formal analysis, Investigation, Project administration, Supervision, Validation, Visualization, Writing—original draft, Writing—review & editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed herein.

Conflict of interest declaration. At the time of article submission and acceptance, Yamir Moreno was a member of the Proceedings A Editorial Board. They were not involved in the editorial process for this manuscript. The authors declare no other competing interests.

Funding. No funding has been received for this article.

Acknowledgements. F.A.R. acknowledges CNPq (grant 308162/2023-4) and FAPESP (grants 24/02322-0, 20/09835-1 and 13/07375-0) for the financial support given for his research. F.A.R. gratefully acknowledges support from the Leverhulme Trust for the Visiting Professorship provided. This research is also supported by FAPESP (grant 2015/50122-0) and DFG-GRTK (grant 1740/2). T.P. acknowledges FAPESP (Grant No. 2023/07481-6) and CNPq/Brazil (Grant No. 310248/2023-0). Y.M. was partially supported by the Government of Aragón, Spain and 'ERDF A way of making Europe' through grant E36-23R (FENOL) and by Ministerio de Ciencia e Innovación, Agencia Española de Investigación (MCIN/AEI/10.13039/501100011033), Grant No. PID2020-115800GB-I00.

References

1. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU. 2006 Complex networks: structure and dynamics. *Phys. Rep.* **424**, 175–308. (doi:10.1016/j.physrep.2005.10.009)
2. Barrat A, Barthelemy M, Vespignani A. 2008 *Dynamical processes on complex networks*. New York, NY: Cambridge University Press.

3. Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A. 2015 Epidemic processes in complex networks. *RMP* **87**, 925. (doi:10.1103/revmodphys.87.925)
4. Rodrigues FA, Peron TKD, Ji P, Kurths J. 2016 The Kuramoto model in complex networks. *Phys. Rep.* **610**, 1–98. (doi:10.1016/j.physrep.2015.10.008)
5. de Arruda GF, Rodrigues FA, Moreno Y. 2018 Fundamentals of spreading processes in single and multilayer complex networks. *Preprint arXiv*.1804.08777. (doi:10.48550/arXiv.1804.08777)
6. Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A. 2015 Epidemic processes in complex networks. *RMP* **87**, 925. (doi:10.1103/revmodphys.87.925)
7. Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, Makse HA. 2010 Identification of influential spreaders in complex networks. *Nat. Phys.* **6**, 888. (doi:10.1038/nphys1746)
8. de Arruda GF, Peron TKD, de Andrade MG, Achcar JA, Rodrigues FA. 2013 The influence of network properties on the synchronization of Kuramoto oscillators quantified by a bayesian regression analysis. *J. Stat. Phys.* **152**, 519–533. (doi:10.1007/s10955-013-0775-z)
9. Arenas A, Díaz-Guilera A, Kurths J, Moreno Y, Zhou C. 2008 Synchronization in complex networks. *Phys. Rep.* **469**, 93–153. (doi:10.1016/j.physrep.2008.09.002)
10. De Arruda GF, Barbieri AL, Rodríguez PM, Rodrigues FA, Moreno Y, da Fontoura Costa L. 2014 Role of centrality for the identification of influential spreaders in complex networks. *Phys. Rev. E* **90**, 032812. (doi:10.1103/physreve.90.032812)
11. Pineda AM, Kent P, Connaughton C, Rodrigues FA. 2023 Machine learning-based prediction of Q-voter model in complex networks. *J. Stat. Mech.: Theory Exp.* **2023**, 123402. (doi:10.1088/1742-5468/ad06a6)
12. Mehta P, Bukov M, Wang CH, Day AG, Richardson C, Fisher CK, Schwab DJ. 2018 A high-bias, low-variance introduction to machine learning for physicists. *Preprint arXiv*.1803.08823. (doi:10.48550/arXiv.1803.08823)
13. Rodrigues FA. 2018 Network centrality: an introduction. In *From nonlinear dynamics to complex systems: a mathematical modeling approach*. New York, NY: Springer.
14. James G, Witten D, Hastie T, Tibshirani R. 2013 *An introduction to statistical learning* vol. 112. New York, NY: Springer.
15. Rodrigues FA. 2023 Machine learning in physics: a short guide. *EPL* **144**, 22001. (doi:10.1209/0295-5075/ad0575)
16. Theodoridis S, Koutroumbas K. 2008 *Pattern recognition, Fourth Edition*. Orlando, FL: Academic Press, Inc. 4th edition. (doi:10.1016/B978-1-59749-272-0.X0001-2)
17. Bishop CM. 2006 *Pattern recognition and machine learning*. New York, NY: Springer. (doi:10.1117/1.2819119)
18. Keeling MJ, Rohani P. 2008 *Modeling infectious diseases in humans and animals*. Princeton, NJ: Princeton University Press. (doi:10.2307/j.ctvcm4gk0)
19. Lü L, Chen D, Ren XL, Zhang QM, Zhang YC, Zhou T. 2016 Vital nodes identification in complex networks. *Phys. Rep.* **650**, 1–63. (doi:10.1016/j.physrep.2016.06.007)
20. Gómez S, Arenas A, Borge-Holthoefer J, Meloni S, Moreno Y. 2010 Discrete-time Markov chain approach to contact-based disease spreading in complex networks. *EPL* **89**, 38009. (doi:10.1209/0295-5075/89/38009)
21. Colizza V, Pastor-Satorras R, Vespignani A. 2007 Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nat. Phys.* **3**, 276–282. (doi:10.1038/nphys560)
22. Dünker D, Jérôme K. 2015 Social networking by proxy: analysis of Dogster, Catster and Hamsterster. In *Proc. of the 24th Int. Conf. on World Wide Web*, pp. 361–362. (doi:10.1145/2740908.2745936)
23. Adamic LA, Glance N. 2005 The political blogosphere and the 2004 US election: divided they blog. In *Proc. of the 3rd Int. Workshop on Link Discovery*, pp. 36–43. (doi:10.1145/1134271.1134277)
24. Kumar S, Spezzano F, Subrahmanian V, Faloutsos C. 2016 Edge weight prediction in weighted signed networks. In *Data Mining (ICDM), 2016 IEEE 16th Int. Conf. on Data Mining*, pp. 221–230. IEEE. (doi:10.1109/icdm.2016.0033)
25. Leskovec J, Kleinberg J, Faloutsos C. 2007 Graph evolution: densification and shrinking diameters. *ACM Trans. Knowl. Discov. Data.* **1**, 2–es. (doi:10.1145/1217299.1217301)
26. Goodfellow I, Bengio Y, Courville A, Bengio Y. 2016 *Deep learning* vol. 1. Cambridge, MA: MIT Press Cambridge.
27. Ichinomiya T. 2004 Frequency synchronization in a random oscillator network. *Phys. Rev. E* **70**, 026116. (doi:10.1103/PhysRevE.70.026116)

28. Watts DJ, Strogatz SH. 1998 Collective dynamics of 'small-world' networks. *Nat.* **393**, 440.
29. Perc M, Gómez-Gardenes J, Szolnoki A, Floría LM, Moreno Y. 2013 Evolutionary dynamics of group interactions on structured populations: a review. *J. R. Soc. Interface.* **10**, 20120997. (doi:10.1098/rsif.2012.0997)
30. de Arruda GF, Rodrigues FA, Moreno Y. 2018 Fundamentals of spreading processes in single and multilayer complex networks. *Phys. Rep.* **756**, 1–59. (doi:10.1016/j.physrep.2018.06.007)
31. Holme P, Saramäki J. 2012 Temporal networks. *Phys. Rep.* **519**, 97–125. (doi:10.1016/j.physrep.2012.03.001)
32. Berner R, Gross T, Kuehn C, Kurths J, Yanchuk S. 2023 Adaptive dynamical networks. *Phys. Rep.* **1031**, 1–59. (doi:10.1016/j.physrep.2023.08.001)
33. Velásquez-Rojas F, Ventura PC, Connaughton C, Moreno Y, Rodrigues FA, Vazquez F. 2020 Disease and information spreading at different speeds in multiplex networks. *Phys. Rev. E* **102**, 022312. (doi:10.1103/PhysRevE.102.022312)
34. Ventura PC, Moreno Y, Rodrigues FA. 2021 Role of time scale in the spreading of asymmetrically interacting diseases. *Phys. Rev. Res.* **3**, 013146. (doi:10.1103/PhysRevResearch.3.013146)
35. Majhi S, Perc M, Ghosh D. 2022 Dynamics on higher-order networks: a review. *J. R. Soc. Interface.* **19**, 20220043. (doi:10.1098/rsif.2022.0043)
36. Ferraz de Arruda G, Aleta A, Moreno Y. 2024 Contagion dynamics on higher-order networks. *Nat. Rev. Phys.* pp. 1–15. (doi:10.1038/s42254-024-00733-0)