COMMENTARY

One model to rule them all in network science?

Roger Guimerà[a,b,1]

If you have ever used a social network platform, you know that you are regularly prompted about people you may know in the network. Sometimes these recommendations are striking—we get a suggestion for a person we have met only once, or an old acquaintance that we have not seen in years. How could anyone (let alone a computer) possibly guess? Predicting acquaintances in a social network is just one example of the general problem of link prediction (1–5), which consists of predicting connections (links) in a network (6) from the observation of other connections (Fig. 1). Besides social networking, the problem of link prediction occurs in many contexts, from recommender systems, where customers are recommended (linked to) items based on their previous ratings or purchases (7), to the prediction of unknown harmful (or perhaps synergistic) interactions between drugs (8).

Besides its practical importance, link prediction is also relevant from a fundamental point of view—just as our ability to predict astronomical phenomena reflects our understanding of gravitation and the physical and chemical processes of celestial bodies, our ability to predict links reflects our understanding of the processes responsible for the structure of complex networks. Indeed, if we knew exactly the mechanisms that generated a given network, and could translate those mechanisms into a fully specified (deterministic or, more likely, probabilistic) generative model with precise parameter values, then we would be able to make optimal link predictions for that network. In the real world, things are more complicated because we never have such exact models, but in general it is still true that more plausible models tend to make better predictions of links (9).

Despite the fact that hundreds of link prediction algorithms have been proposed in recent years, no large-scale systematic comparison existed to this day. In PNAS, Ghasemian et al. (10) consider 203 link prediction models and apply them to 550 real-world social, biological, economic, technological, information, and transportation networks. They find wide disparities in performance; models often perform well in some groups of networks but poorly in others, and no model is superior for all networks. Also, some groups of networks (biological and technological) are considerably harder to predict than others (social). Finally, and perhaps most importantly, Ghasemian et al. show that an ensemble method known as model stacking (11), in which all 203 models are combined (or "stacked"), can predict links in most networks more accurately than any of the 203 models on their own (Fig. 1).

Free Lunches in Link Prediction

Ghasemian et al. (10) frame their contribution in terms of the so-called no-free-lunch (NFL) theorems (12, 13). Generally speaking, NFL theorems state that, for certain classes of mathematical problems, no single model or algorithm performs better (or worse) than any other when applied to all possible problems in the class; if a model is better than another at solving one particular problem, then it must be worse at others. In network science, an NFL theorem has been proved for the problem of identifying natural groups (or communities or modules) of nodes in the network (13). According to the theorem, all algorithms perform identically when applied to all possible networks and all possible groupings of nodes in each network. So Ghasemian et al. (10) hypothesize that link prediction may be of the NFL class and, therefore, argue that no algorithm may be consistently better at predicting links across all networks.

With this rationale they use model stacking (11), that is, a metamodel that works by combining all other models; if each model is good at some link prediction problems and bad at others, then such a metamodel may be able to get the best of each one and be good in all cases. Indeed, they find that model stacking performs better than all individual models in most networks considered. And here is the paradox—the stacked model is, after all, another model, so this looks exactly like a free lunch. In this regard, the results of Ghasemian et al. (10) highlight the theoretical value of NFL theorems, as well as their limitations from a

[a]Department of Chemical Engineering, Universitat Rovira i Virgili, 43007 Tarragona, Catalonia, Spain; and [b]Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Catalonia, Spain

Author contributions: R.G. wrote the paper.

The author declares no competing interest.

See companion article, "Stacking models for nearly optimal link prediction in complex networks," 10.1073/pnas.1914950117.

[1]Email: roger.guimera@urv.cat.

First published September 28, 2020.

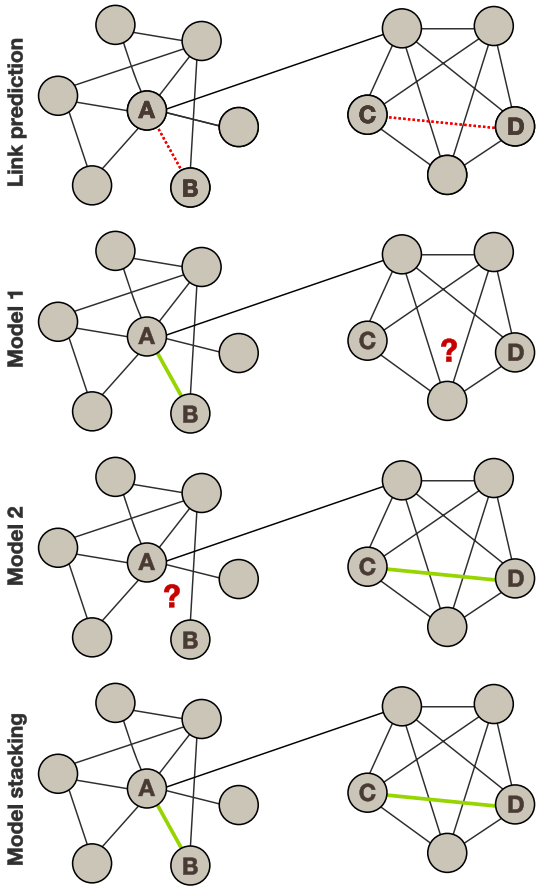

Fig. 1. In the problem of link prediction, we are asked to identify which unobserved links in a network are more likely to exist. Nodes could represent individuals or drugs, and links could represent, respectively, friendship relationships in a social network or harmful drug–drug interactions. In this example, links *AB* and *CD* exist but have not been observed, so we aim to predict them. Model 1 pays attention only to the connectivity of nodes, and it captures that many nodes are connected to *A*, so it correctly predicts the *AB* link. However, since there is nothing especial about the connectivity of *C* and *D*, it misses the *CD* link. Conversely, model 2 pays attention only to group structure, so it realizes that all nodes in the group at *Right* are connected to each other, and it predicts the *CD* link. However, since in the group at *Left* many pairs of nodes are not connected to each other, it misses link *AB*. Model stacking, in which models 1 and 2 are combined, may be able to predict both links.

practical point of view. Theoretically, they compel us to think about ensembles of models, each model having its own virtues and limitations. In practice, we are rarely (if ever) interested in solving all problems of a given class, and some models are more appropriate than others because they are good at solving the interesting problems and bad at solving the uninteresting ones. So, in practice, there are free lunches. In the particular case of link prediction, we are not interested in predicting links in any network, but only real-world networks. And real-world networks have some general properties that make some specific models more appropriate than others, which network science has been uncovering in the last 20 y—broad degree distributions or the related properties of group structure and "small worldiness," to name the most prominent ones.

By bringing forth this apparent paradox of NFL theorems, the article by Ghasemian et al. (10) invites us to think more deeply about two keys aspects of network science (and, in a way, science more broadly): ensemble methods and model expressiveness.

Ensemble Methods

With regard to the first issue, Ghasemian et al. (10) convincingly argue for the need to combine different models because of their error diversity, that is, because each model makes certain errors and, conversely, is able to correctly capture certain unique features. By combining models we can get the best of each. But, no less important and perhaps more fundamental, ensemble methods are necessary because in (network) science we are seldom certain about the exact process that generated whatever empirical observations we may have, and this uncertainty must be taken into consideration using probability theory (14). In particular, given a set $\mathcal{M}=\{M_1, M_2, \dots\}$ of candidate models that could potentially explain a certain dataset D, each model M_i has a probability $p(M_i|D)$ of being the true generating model given the data. Given this posterior distribution over models, and assuming that the true generating model is one of the models in $\mathcal{M}$, the consistent estimate f^* of a given property $f(M)$ (for example, the existence of a link in a network in the link prediction problem) is

$$f^* = \sum_i f(M_i)p(M_i|D). \quad [1]$$

This is sometimes referred to as Bayesian model averaging (15), although it is nothing more than the direct application of the laws of probability. When we consider only one model for estimating f^* (typically the most plausible model, arg $\max_M p(M|D)$), we are making an approximation to the consistent estimator, and this approximation will be poor unless one model is overwhelmingly more plausible than all others. Although model stacking and Bayesian model averaging are not the same, they can be interpreted under unifying frameworks, along with other ensemble methods. Bayesian model averaging provides the consistent estimation provided that the true generating model is considered [if not, more models should be added to the ensemble to have increasingly plausible theories (14)], but model stacking may be more useful to solve specific problems in practice. Given their success, understanding ensemble methods and their connections better is one important problem ahead of us.

Model Expressiveness

Ghasemian et al. (10) also shed light on the second issue, namely model expressiveness in network science. Model expressiveness is the ability of a model to represent distinct relevant situations; in network science, a model is expressive to the extent that it can represent many different features of real-world networks (for example, broad connectivity distributions, connectivity correlations, network motifs, or group structure). The proposed model stacking is explicitly built to be very expressive by incorporating the expressiveness of each of the 203 constituent models, but not all constituent models are equally expressive. Indeed, given a network and the stacked model, a very interesting question is, Which constituent model contributes most to the description of the network? It is not difficult to see how answering this question could illuminate the mechanisms responsible for the structure of the network and how the process of mechanism discovery could be set on solid grounds by extending this scheme—adding new models to the stacked model, checking how they improve predictive power, and iterating the process.

In this sense, the work of Ghasemian et al. (10) is also seminal as a large-scale comparison of the predictive ability of hundreds of models on hundreds of networks. And the results are telling—stochastic block models (3, 16–18) and, in particular, the degree-corrected

stochastic block model (19) are, on average, the most explanatory individual models for biological, economic, technological, information, and transportation networks, that is, for all types of networks except social networks (figure 1 in ref. 10). Considering, again, the NFL theorems, the general goodness of stochastic block models is remarkable and suggests that they are, themselves, extremely expressive models. In fact, Bayesian model averaging over the whole family of stochastic block models (3, 9, 20) is almost as predictive of links as the whole stacked model including 203 diverse constituent models (figure 3 in ref. 10)—except, of course, that when the Bayesian model averaging of the stochastic block models is added to the stack, the stacked model comes out ahead by a fair amount again. So stochastic block models are very expressive, but can still benefit for features that are present in other existing models. Can we find ways to incorporate such features into mathematically tractable model families?

In very remarkable ways, then, and besides link prediction, Ghasemian et al. (10) contribute significantly to a program for solid progress in network modeling and for understanding the generative mechanisms of complex networks. In a way, it seems, progress in network science is about finding free lunches—model families that can express the features of real-world networks, that we can explore systematically, and that we can refine as new patterns are uncovered to improve their predictive performance.

Acknowledgments

R.G.'s research is supported by the Spanish Ministerio de Economia y Competitividad, Grant FIS2016-78904-C3-P-1, and by the Government of Catalonia, Grant 2017SGR-896.

1 D. Liben-Nowell, J. Kleinberg, The link-prediction problem for social networks. *J. Assoc. Inf. Sci. Technol.* **58**, 1019–1031 (2007).
2 A. Clauset, C. Moore, M. E. J. Newman, Hierarchical structure and the prediction of missing links in networks. *Nature* **453**, 98–101 (2008).
3 R. Guimerà, M. Sales-Pardo, Missing and spurious interactions and the reconstruction of complex networks. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 22073–22078 (2009).
4 L. Lü, T. Zhou, Link prediction in complex networks: A survey. *Physica A* **390**, 1150–1170 (2011).
5 L. Lü, L. Pan, T. Zhou, Y. C. Zhang, H. E. Stanley, Toward link predictability of complex networks. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 2325–2330 (2015).
6 M. E. J. Newman, *Networks* (Oxford University Press, ed. 2, 2018).
7 A. Godoy-Lorite, R. Guimerà, C. Moore, M. Sales-Pardo, Accurate and scalable social recommendation using mixed-membership stochastic block models. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14207–14212 (2016).
8 R. Guimerà, M. Sales-Pardo, A network inference method for large-scale unsupervised identification of novel drug-drug interactions. *PLoS Comput. Biol.* **9**, e1003374 (2013).
9 T. Vallès-Català, T. P. Peixoto, R. Guimerà, M. Sales-Pardo, Consistencies and inconsistencies between model selection and link prediction in networks. *Phys. Rev. E* **97**, 062316 (2018).
10 A. Ghasemian, H. Hosseinmardi, A. Galstyan, E. M. Airoldi, A. Clauset, Stacking models for nearly optimal link prediction in complex networks. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 23393–23400 (2020).
11 D. H. Wolpert, Stacked generalization. *Neural Network* **5**, 241–259 (1992).
12 D. H. Wolpert, W. G. Macready, No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* **1**, 67–82 (1997).
13 L. Peel, D. B. Larremore, A. Clauset, The ground truth about metadata and community detection in networks. *Sci. Adv.* **3**, e1602548 (2017).
14 E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, 2003).
15 D. Madigan, J. A. Hoeting, A. E. Raftery, C. T. Volinsky, Bayesian model averaging: A tutorial. *Stat. Sci.* **14**, 382–417 (1999).
16 H. C. White, S. A. Boorman, R. L. Breiger, Social structure from multiple networks. I. Blockmodels of roles and positions. *Am. J. Sociol.* **81**, 730–780 (1976).
17 K. Nowicki, T. A. B. Snijders, Estimation and prediction for stochastic blockstructures. *J. Am. Stat. Assoc.* **96**, 1077–1087 (2001).
18 T. P. Peixoto, Hierarchical block structures and high-resolution model selection in large networks. *Phys. Rev. X* **4**, 011047 (2014).
19 B. Karrer, M. E. J. Newman, Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **83**, 016107 (2011).
20 T. P. Peixoto, Merge-split Markov chain Monte Carlo for community detection. *Phys. Rev. E* **102**, 012305 (2020).