Network Science **8** (S1): S43–S64, 2020
doi:10.1017/nws.2019.63

CAMBRIDGE UNIVERSITY PRESS

ORIGINAL ARTICLE

Examining the variability in network populations and its role in generative models

Viplove Arora[†], Dali Guo[†], Katherine D. Dunbar and Mario Ventresca[*]

School of Industrial Engineering, Purdue University, West Lafayette IN, 47907 USA
[*]Corresponding author. Email: mventresca@purdue.edu

Action Editor: Hocine Cherifi

Abstract

A principled approach to understand networks is to formulate generative models and infer their parameters from given network data. Due to the scarcity of data in the form of multiple networks that have evolved from the same process, generative models are typically formulated to learn parameters from a single network observation, hence ignoring the natural variability of the "true" process. In this paper, we highlight the importance of variability in evaluating generative models and present two ways of quantifying the variability for a finite set of networks. The first evaluation scheme compares the statistical properties of networks in a dissimilarity space, while the other relies on data-driven entropy measures to compute variability in network populations. Using these measures, we evaluate the ability of four generative models to synthesize networks that capture the variability of the "true" process. Our empirical analysis suggests that generative models fitted for a single network observation fail to capture the variability in the network population. Our work highlights the need for rethinking the way we evaluate the goodness-of-fit of new and existing network models and devising models that are capable of matching the variability of network populations when available.

Keywords: network models, variability, entropy, generative models

1. Introduction

The representation of various natural and artificial systems in the form of complex networks, where the nodes serve as elements and edges describe interactions between elements, has been tremendously useful in understanding the structure and function of these systems (Watts & Strogatz, 1998; Barabási & Albert, 1999; McPherson et al., 2001; Ladyman et al., 2013). Indeed, the growth of network science has provided new perspectives on complex systems using data-based mathematical models, which have been used to analyze and model these systems at various scales (Barabasi, 2012). In particular, the increasing availability of high throughput data from a wide variety of sources, such as the Internet (Faloutsos et al., 1999), World Wide Web (Huberman, 2003), online social networks (Moreno, 1934), citation and collaboration networks (Price, 1965; Perc et al., 2017), and biological networks (brain connectivity [Fields & Song, 1989], protein-protein interactions [White et al., 1986]), has fueled the rapid development of the field of network science. Topological analysis of real-world networks has revealed a variety of characteristic properties such as small-worldness (Watts & Strogatz, 1998), scale-free degree distributions (Barabási & Albert, 1999), degree correlations (Newman, 2003a; Maslov & Sneppen, 2002; Pastor-Satorras

[†]Both authors contributed equally to this manuscript.

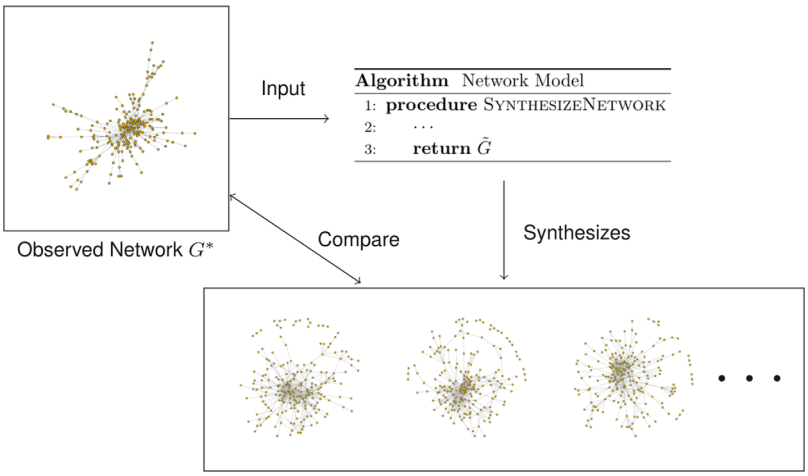

Figure 1. Traditional setup used in network modeling: a network model consists of an algorithmic procedure that is parameterized using the observed network G^* as an input. The parameterized model can then be used to synthesize multiple networks, and the quality of the model is evaluated by comparing the synthesized networks with the original input network.

et al., 2001), motifs (Milo et al., 2002), and communities (Palla et al., 2005; Reichardt & Bornholdt, 2006; Newman, 2008).

A principled approach to understanding these complex networks (and the processes that give rise to them) is to formulate generative models and infer their parameters from given data (Newman, 2003b; Goldenberg et al., 2009). These generative models attempt to identify a common set of laws and principles that can explain the structure and evolution of the networks, and the underlying system it represents (Barabasi, 2002, 2016). The set of laws and principles therefore shed light on predicting implicit characteristics and future development of the system. For instance, a brain network model that explains the development of brain structure can help make early diagnosis of brain disease. Consequently, a goal of network modeling is to solve the problem of decoding how the observed structure of a network supports its perceived/desired function (Alderson, 2008). Due to this, a long-standing question in the network science community has concerned the existence of a general model capable of generating synthetic networks that are statistically representative of real networks. Traditionally, generative models have been formulated to use a single empirical observation G^* of the true system as the input to an algorithmic procedure, whose parameters are best fit to synthesize networks statistically similar to G^* (see Figure 1 for a pictorial representation). Note that it is far from guaranteed that best-fit parameterization of the algorithm will yield a satisfactory generative model. Nevertheless, there are an infinite number of models that can be formulated for a given network observation (Vallès-Català et al., 2018). Unfortunately, most of the existing models either make assumptions biased by system-specific observations that are not plausible across domains, or focus on replicating a few predefined topological features, such as degree distribution and clustering, at the expense of other potentially more important characteristics. Without any indication that they are either necessary or sufficient as descriptors for the actual network data, these summary quantities can often be highly misleading (Goldenberg et al., 2009). Further, even when a model is capable of consistently reproducing a set of target properties, it might fail to capture their naturally occurring stochasticity (Gutfraind et al., 2012).

As depicted in Figure 2, statistical modeling of networks based on a single observation assumes that G^* is somehow representative of the "true" process A^*. Learning a model using a single observation that does not reflect the variability in the process A^* can potentially bias a model to synthesize networks that over-fit G^*. Such a setup could restrict the generalizability of the model

https://doi.org/10.1017/nws.2019.63 Published online by Cambridge University Press

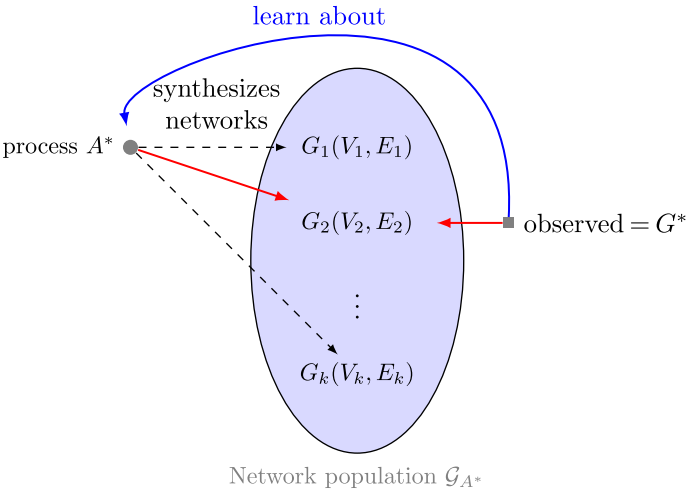

Figure 2. Procedure used for evaluating network models in this paper: Assuming G^* is not an outlier, how well do existing network models approximate the process A*?

by failing to synthesize networks that provide a reliable representation of the "true" process. To understand this, we need to define the concept of a network population.

Definition 1.1 (Network Population). *Let $G_1(V_1, E_1)$ be an arbitrary network that has nonzero probability of being synthesized using the process A^*. A finite set of k such realizations $\mathcal{G}_{A^*} = \{G_i(V_i, E_i)\}\ \forall i = 1, \ldots, k$ is called a network population, where V_i and E_i are the sets of nodes and edges in G_i.*

To illustrate the variability in networks synthesized using a single known process A^*, we use the preferential attachment process of Barabási & Albert (1999) to synthesize a population of 100 networks (labeled as "data" in Figure 3). The stochasticity in the generative process of the BA-model leads to the synthesis of networks that show some variability in their global network properties (degree assortativity and transitivity are used in Figure 3). As depicted in Figure 2, we typically observe a single network from the population and ideally would like a generative model to be capable of learning about the original process A^* using the observation G^*. Consequently, one network from the population was selected at random as input for training two other network models, namely *dk*-random graphs as model 1 and action-based networks as model 2 (see Section 2 for more details), which were then used to synthesize populations of 100 networks each. Finally, we compare the distribution of degree assortativity and transitivity in the three populations in Figure 3. We observe that the networks synthesized by model 1 all have exactly the same global network properties. While there is no variability in these properties, a simple comparison with the observed network G^* might lead to a conclusion that the model aptly describes the underlying process A^*, which can prove to be highly misleading. On the other hand, the population of networks synthesized by model 2 shows more variability, but fails to match the network properties of G^* or the original population.

The inability of network models to synthesize realistic network populations necessitates the evaluation of network models using a well-formulated methodology that treats the observed network as a sample originating from some unknown process A^*. Statistical hypothesis testing for goodness-of-fit typically involves measuring the discrepancy between observed values and the values expected under the model in question. Similarly, in the context of networks, we would like to *evaluate the ability of a candidate model to approximate the network population $\mathcal{G}_{A^*}$ using a single observed network G^** (see Figure 2 for a pictorial representation). Although, in most cases, we do not have a population of independent instances of networks that can be used to draw a set of

https://doi.org/10.1017/nws.2019.63 Published online by Cambridge University Press

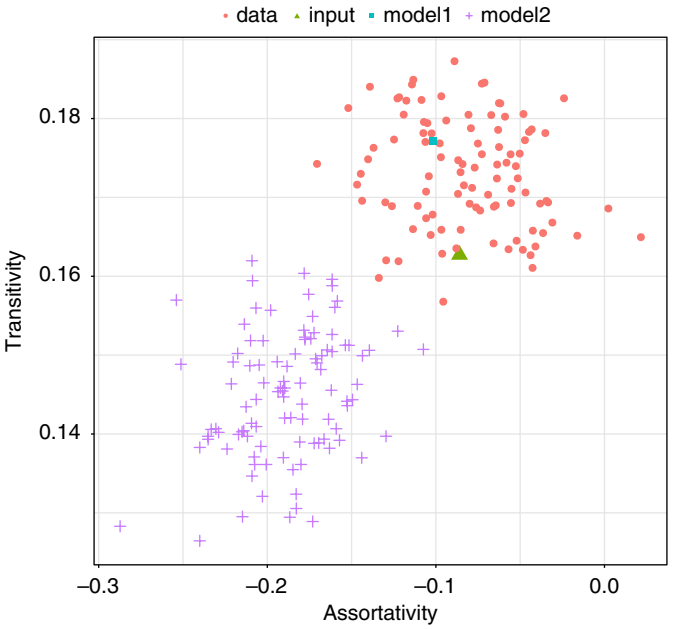

Figure 3. Distribution of global network properties assortativity and transitivity in populations of networks synthesized by the Barabási–Albert model and two other models. The inability of network models to replicate the distributional properties of the original process A^* highlights the need for devising better techniques for training and evaluating models using network populations.

samples (Moreno & Neville, 2009), it has been shown that it is possible to establish a baseline test set to evaluate the ability of a network model to capture the distribution of network populations (Arora & Ventresca, 2019). Emmert-Streib & Dehmer (2012) made a similar observation, where it was shown that evaluating the complexity of a network model using a single network can be biased because of the variability in the samples, and therefore introduced a complexity measure based on network populations.

Even though network science has provided us with numerous methods to compare pairs of networks (see Emmert-Streib et al., 2016; Donnat & Holmes, 2018 for reviews), comparing network populations has been a relatively unexplored area. Given the importance of natural variability in generative processes (Arora & Ventresca, 2019), it is imperative to devise methods that can be used to compare populations and quantify their variability. Such methods can help researchers in the development of models that augment our current understanding of complex networks and the underlying process they represent.

One way to quantify the information content in a finite population of networks is to use entropy-based measures to compute the amount of randomness in the population. Dehmer & Mowshowitz (2011) review the diverse contexts and applications that led to the development of various entropy-based measures for graphs/networks. Information-theoretic metrics have also been used for quantifying the difference between pairs of complex networks (De Domenico & Biamonte, 2016). Methods for quantifying the Shannon entropy of canonical and microcanonical network ensembles have also been proposed (Anand & Bianconi, 2009; Bianconi, 2008), which were subsequently extended to the case of stochastic block models (SBMs) (Peixoto, 2012). While closed-form expressions can be obtained for simple null models, achieving the same for a population of real-world networks is rather unlikely. Alternatively, data-driven techniques to quantify the information content and variability in network populations can aid the rapid development of more explanatory network models that capture the distributional properties of a population.

https://doi.org/10.1017/nws.2019.63 Published online by Cambridge University Press

While the inability of certain network models to reproduce the naturally occurring variability in networks can be attributed to the fact that they sample each edge independently through Bernoulli distributions (Moreno et al., 2018), there could be other factors contributing to the lack of variability. There are at least two ways of overcoming this obstacle: (i) systematic development of models that can replicate the distributions of populations, or (ii) devising techniques that can allow existing models to utilize additional information from the population. There has been recent work on mapping networks into natural Euclidean spaces for conventional hypothesis testing (Ginestet et al., 2017), which can help us determine if two groups of networks are significantly different in statistics. Latent space models (LSMs) (Hoff et al., 2002) were proposed to synthesize networks by mapping the nodes into some low-dimensional Euclidean space while keeping the relationships. Gollini & Murphy (2016) adapted LSMs for multiple input networks and proposed a joint LSM considering multiple networks in the posterior. In Durante et al. (2017), the context was extended to multiple observations drawn from a common population, and a nonparametric model was proposed. Lunagómez et al. (2019) proposed another Bayesian model where the probability of an observation is based on the Hamming distance to the Fréchet mean of the group. Similarly, Sweet et al. (2012) proposed a hierarchical modeling framework to learn better models of network populations. Although these approaches incorporate information about multiple networks to learn better representations for the population, they fail to explicitly account for the natural variability in the population.

To recap, an ideal generative model M would exactly correspond to the true process A^* that defines the dynamical processes responsible for the observed data G^*. That is, if A^* defines a probability distribution $\mathbb{P}_G(A^*)$ $\forall G \in \mathcal{G}_{A^*}$, then $\mathbb{P}_G(M)$ and $\mathbb{P}_G(A^*)$ would be identical. As stated above, A^* is usually unknown and the number of observed networks in the data G^* are usually small (sometimes only one).

1.1 Summary of main contributions

In this work, we expand on previous work (Arora & Ventresca, 2019) exploring the distributional properties of four competing generative models: Chung-Lu model, *dk*-random graphs, exponential random graphs, and action-based network generators (ABNGs) (Section 2 briefly describes each model). In Section 3, we describe the network population data drawn from three known processes and six real-world populations. As described in Figure 2, model parameters are fit using a representative sample G^* chosen at random. In Section 4, we propose the construction of a dissimilarity space that measures the distributional properties of network populations with respect to the observed network G^*. In Section 4.1, the fitted models are used to synthesize networks followed by an investigation of their distributional properties. This evaluation is done by comparing the statistical properties of the synthesized networks with the properties of the corresponding population of networks. In Section 5, we propose empirical measures for quantifying the variability in network populations, and subsequently use it for the evaluating network models that synthesized networks using information extracted from an increasing number of network samples[1].

2. Background

In this section, we briefly introduce the four generative models that are used in the empirical analysis in Section 4.1. Chung-Lu model and exponential random graphs are classical models based on independent edge sampling and conditional sampling on local structure, respectively. Two recent models, *dk*-random graphs and ABNGs, aimed at matching the distribution of graph ensembles and simulating the network formation process are introduced. Where applicable, details regarding user-defined inputs have also been provided. In our experimental analysis in Section 4, each model uses the network $G^* = (V, E)$ to learn a fixed set of parameters. In Section 5, we also test

https://doi.org/10.1017/nws.2019.63 Published online by Cambridge University Press

scenarios when these network models use multiple input networks to learn better models for network populations.

2.1 Chung-Lu model

The Erdös–Rényi model (Erdös & Rényi, 1959, 1960) assumes that nodes are homogeneous with respect to how they connect to other nodes. This assumption fails to produce networks exhibiting properties observed in real-world networks, see Chapter 3 of Cohen & Havlin (2010) for a more comprehensive description of properties of real-world networks that cannot be explained by the random graph model. This inspired a general class of models of inhomogeneous random graphs that can be visualized as attempts at making the random graph model more realistic. The Chung-Lu model Chung & Lu (2002a,b) is one of the simplest models in this category, where a node i is assigned a degree k_i from the degree distribution of G^* and an edge is placed between the node pair (i, j) with probability proportional to $k_i k_j$, that is, the probability that an edge exists between nodes i and j is given by

$$\mathbb{P}_{i,j} = \frac{k_i k_j}{2|E|}. \tag{1}$$

This model has the disadvantage that the final degree sequence is not precisely equal to the desired degree sequence (it matches the degree sequence in expectation), but it has some significant calculational advantages that make the derivation of rigorous results easier (Newman, 2003b). The Chung-Lu model is often used as a null model owing to its simplicity and ability to synthesize fairly realistic networks (Pinar et al., 2012). Unfortunately, the Chung-Lu model synthesizes networks with low clustering coefficients making it unsuitable for most real-world applications.

2.2 Exponential random graphs

One of the most popular statistical network models in the social science literature is the exponential random graph models (ERGMs) (Strauss, 1986; Wasserman & Pattison, 1996; Anderson et al., 1999). These models deal with link formation mechanisms using conditional dependence, which states that the existence of links in a network is shaped by the presence or absence of other links (and possibly node-level attributes) (Robins et al., 2007). These dependence assumptions claim the existence of local processes, that may depend on the local structure, capable of generating node interactions observed in real-world networks. ERGMs represent probability distributions over networks with an exponential linear model that uses feature counts of local graph properties considered relevant by the modeler (e.g., edges, triangles, paths):

$$P(\mathbf{Y} = G^* | \boldsymbol{\theta}) = \frac{1}{Z} \exp\left(\boldsymbol{\theta}^T \boldsymbol{\phi}(G^*)\right), \tag{2}$$

where (i) $\boldsymbol{\phi}(G^*)$ are feature counts of G^*; (ii) $\boldsymbol{\theta}$ are parameters to be learned; and (iii) Z is a normalizing constant. The generality of the exponential distribution makes it an ideal candidate for representing the conditional probability distribution between different local graph properties $\boldsymbol{\phi}(G^*)$. Further, there is theoretical evidence supporting the exponential form as it can be derived from first principles using maximum entropy arguments (Park & Newman, 2004).

Generating an ERGM consists of the following steps (Robins et al., 2007): (i) assume that the existence of each edge is a random variable; (ii) a dependence hypothesis is proposed that embodies the local processes assumed to generate the network; (iii) network configurations (e.g., triangles, 2-stars) get parameter values based on the dependency hypothesis; (iv) parameters are simplified using homogeneity or other constraints to reduce number of parameters; (v) model parameters are estimated and interpreted from the observed network data to get a statistical model for the network. Though ERGMs are the most widely used models for social networks, they are plagued with the degeneracy problem (i.e., the probability distribution is biased toward empty

https://doi.org/10.1017/nws.2019.63 Published online by Cambridge University Press

and complete networks), whereas real-world networks are sparse. In our experimental evaluation, the following feature counts $\boldsymbol{\phi}(G^*)$ were used as they are known to be capable of circumventing the degeneracy problem (see Snijders et al., 2004; Hunter, 2007 for more details): (i) total number of edges, (ii) geometrically weighted degree distribution, (iii) geometrically weighted dyadwise shared partner distribution, and (iv) geometrically weighted edgewise shared partner distribution.

2.3 dk-random graphs

In Orsini et al. (2015), it was observed that fixing some structural properties in a network model to those observed in the given network can lead to the appearance of other statistical properties as a consequence. These observations follow from earlier research on the *dk*-series (Mahadevan et al., 2006), which defines a series of null models or random graph ensembles (Orsini et al., 2015). Consequently, *dk*-random graphs (Orsini et al., 2015) model networks as random ensembles, where ensemble size is controlled using *dk*-distributions. *dk*-random graphs for $d = 0, 1, 2$ correspond to the random graph model (Erdös & Rényi, 1959), configuration model (Bender & Canfield, 1978; Molloy & Reed, 1995), and random graphs with a given joint degree distribution (Stanton & Pinar, 2012), respectively.

The sampling algorithm of *dk*-random graphs relies on ergodic edge-swapping operations to sample networks from an ensemble defined using the chosen *dk*-distributions. The lack of an edge-swapping operation that is ergodic for 3*k*-distributions leads to the creation of 2.1*k* and 2.5*k*-targeting rewiring, where the moves preserve the 2*k*-distribution, but each move is accepted with probability p designed to drive the graph closer to a target value of average clustering $\bar{c}(2.1k)$ or degree-dependent clustering $\bar{c}(k)(2.5k)$. Experimental results (Orsini et al., 2015; Schieber et al., 2017; Arora & Ventresca, 2017) have shown that *dk*-random graphs can synthesize networks resembling the input network G^* across a wide range of network properties and for most real-world networks. Despite this fact, the limited inferential capabilities and inability to perform tasks such as compression, extrapolation limit the utility of *dk*-random graphs. In our empirical analysis, we used the *dk*2.5 variant as it is known to outperform other network models on a variety of measures (Schieber et al., 2017; Arora & Ventresca, 2017).

2.4 Action-based network generators

A variety of generative models have been designed to synthesize networks having specific properties (such as power law degree distributions, small-worldness), leaving the modeler with the nontrivial task of choosing an appropriate one. To alleviate the burden of choosing an appropriate model, machine learning and evolutionary algorithms can be used to automatically infer appropriate network generation mechanisms from the observed network structure (Menezes & Roth, 2019). Along these lines, a novel action-based framework was proposed in Arora & Ventresca (2016, 2017) that builds on the common assumption that complex networks naturally form through local node interactions, where the nodes themselves are typically oblivious of the global topology.

The action-based approach of Arora & Ventresca (2017) models networks using local node interactions based on simple link creation processes known as actions. An action is a decision process a node uses to form a link with another node. Given a predefined set of actions, the aim of action-based networks is to learn a probability distribution over these actions, such that the resultant model can synthesize networks statistically similar to a given network observation. A synthesis algorithm $f(\mathbf{M}, n)$ can then be used to synthesize networks containing n nodes using the learned action-based model $\mathbf{M}$, leading to ABNGs. The fundamental idea behind action-based networks is to define a unifying network generative process, which follows from observations by Zheng et al. (2014), who note that there must exist an assembling algorithm to combine local mechanisms for emergence of different complex network structures.

https://doi.org/10.1017/nws.2019.63 Published online by Cambridge University Press

For an observed network G^*, the action-based model $\mathbf{M}$ is determined by solving a multiobjective optimization problem. While choosing an appropriate set of user-defined network characteristics Y and action set A is a challenge, some prior information about network characteristics and plausible actions can yield efficient models (Arora & Ventresca, 2017). In our empirical analysis, we used degree distribution, local assortativity (Piraveenan et al., 2008), and local transitivity of the observed network as the set of network properties in the objective function.

3. Network Population Data

- Barabási–Albert: The Barabási–Albert model (Barabási & Albert, 1999) was used to synthesize networks with each arriving node adding five edges using the linear preferential attachment mechanism.
- Forest Fire: Network populations synthesized using the Forest Fire model (Leskovec et al., 2007) used a forward burning probability $p = 0.38$.
- Stochastic Block Model: The SBM was used to synthesize networks with three assortative communities of sizes 30, 70, and 50.
- Brain Networks: We used diffusion-weighted imaging (DWI) data from the 100 unrelated subjects of the HCP 900 subjects data release (Van Essen et al., 2013) to get the structural brain networks. The preprocessing of the DWI data to get the corresponding networks is described in Amico & Goni (2018). One network represents the abstracted brain structure of one subject. Nodes in network represent regions of interest (ROIs) in brain and edges represent the density of connecting fibers. All networks share the same set of nodes since brain images of different subjects are regularized into a common template of ROIs.
- Contact Networks: Sixty-nine daily cumulated networks where nodes represent visitors of the Science Gallery while the edges represent close-range face-to-face proximity between the concerned persons (Isella et al., 2011). Since visitors showing up on different days are different, the node sets are not fixed for the 69 networks.
- Social Networks in Indian Villages: Data from a survey of social networks in 75 villages in rural southern Karnataka, a state in India (Banerjee et al., 2013). One network represents the social network of one village. Nodes in the network represent individuals and edges represent different social interactions.
- Travian Network Datasets: Data collected over 30 days for real-time strategy game Travian. The message network contains links for messages sent between players, while the trade network represents trading relations (Hajibagheri et al., 2015).
- Autonomous Systems: The graph of routers comprising the Internet can be organized into subgraphs called Autonomous Systems (AS). The dataset (Views, 2000) contains 733 daily instances spanning an interval of 785 days from November 8, 1997 to January 2, 2000. The first 100 networks were used in this study.

Table 1 shows the statistics and some common network metrics of listed datasets. First three populations are generated from parameterized network models, and the the rest six are networks obtained from real-world interactions. Among the real-world populations, the Travian and contact networks are created from interactions among different sets of individuals across different days, but the underlying systems that support these interactions remain the same. The networks are thus different instances of interaction processes happening on a fixed system and can thus be hypothesized to have a common generative process. The structural organization of the human brain is controlled by the human genome, and it is safe to assume that the network representation of different individuals belongs to a population. Similarly, it is reasonable to assume that social interactions between individuals in different villages arise from similar generative mechanisms. Daily instances of subgraphs of AS can again be assumed to have a common underlying generative process.

https://doi.org/10.1017/nws.2019.63 Published online by Cambridge University Press

Table 1. Statistics and network metrics of network populations (standard deviation in parentheses), APL is the average path length

Name	Sample Size	# of nodes	Density	Transitivity	Assortativity	APL
Barabási–Albert	100	100	0.098	0.173	−0.082	2.22
		(0)	(0)	(0.0069)	(0.036)	(0.016)
Forest	100	200	0.058	0.406	−0.032	3.102
Fire		(0)	(0.017)	(0.043)	(0.098)	(0.401)
SBM	100	150	0.072	0.177	−0.512	2.72
		(0)	(0.0025)	(0.0097)	(0.054)	(0.0527)
Brain	100	360	0.032	0.422	0.141	3.73
Networks		(0)	(0.001)	(0.011)	(0.042)	(0.129)
Contact	69	167.21	0.045	0.470	0.362	4.104
Networks		(64.71)	(0.012)	(0.143)	(0.215)	(1.091)
Social	43	212.23	0.048	0.198	−0.078	2.77
Networks		(53.54)	(0.013)	(0.037)	(0.054)	(0.0207)
Travian	30	1144.5	0.0039	0.019	−0.055	4.35
Trades		(123.22)	(0.00055)	(0.003)	(0.036)	(0.134)
Travian	30	1722.2	0.0026	0.108	−0.513	2.72
Messages		(180.76)	(0.00023)	(0.022)	(0.034)	(0.244)
Autonomous	100	3196.3	0.001099	0.015	−0.221	3.77
Systems		(101.7)	(0.000027)	(0.0013)	(0.004)	(0.014)

4. Pairwise evaluation

Evaluation of the distributional properties of a generative model requires a well-defined methodology that correctly represents the distribution over networks. Although model-based techniques for hypothesis testing of networks have been proposed in the literature (Moreno & Neville, 2013; Peixoto, 2015; Casiraghi et al., 2016), they heavily rely on the choice of a baseline model. Alternatively, one could build on the concept of a network morphospace (Avena-Koenigsberger et al., 2014), which provides a coarse-grained approach for classifying and mapping network architectures according to a set of network-level structural characteristics. The network morphospace can be transformed to a network dissimilarity space ($\mathfrak{D}_{G^*} \subset \mathbb{R}^d$), where networks are placed based on their dissimilarity to the single observed network $G^* \in \mathscr{G}_{A^*}$ with respect to a variety of dissimilarity measures (see Definition 4.1). The true process and network models also have counterpart distributions $\mathbb{P}_{\mathfrak{D}_{G^*}}(A^*)$ and $\mathbb{P}_{\mathfrak{D}_{G^*}}(M)$ in the network dissimilarity space. In an appropriately defined dissimilarity space, if $\mathbb{P}_{\mathfrak{D}_{G^*}}(M)$ sufficiently approximates $\mathbb{P}_{\mathfrak{D}_{G^*}}(A^*)$, we might be able to conclude that model M can synthesize networks that belong to the same population as the observed network G^*.

The utility of such a network dissimilarity space relies heavily on the choice of dissimilarity measures used for network comparison. Network science provides numerous quantitative tools to measure and classify different patterns of local and global network architectures across disparate types of systems. The development of methods for the pairwise comparison of networks is an active area of research, and in recent years many new methods have been introduced (see Soundarajan et al., 2014; Emmert-Streib et al., 2016; Donnat & Holmes, 2018 for reviews), which generally take the following form:

Definition 4.1 (Dissimilarity Measure)**.** *Given two networks $G_1 \in \mathscr{G}_1$ and $G_2 \in \mathscr{G}_2$, a (bivariate) network dissimilarity measure $d(G_1, G_2)$ is a mapping from $\mathscr{G}_1 \times \mathscr{G}_2 \to \mathbb{R}$.*

https://doi.org/10.1017/nws.2019.63 Published online by Cambridge University Press

Any dissimilarity measure that defines a real-valued distance akin to the one in Definition 4.1, which goes to zero for a pair of isomorphic networks, can be used in the dissimilarity space. A set of node-level measures that could prove particularly useful for the network dissimilarity space is provided by the *dk*-series (Orsini et al., 2015), which is a systematic series of properties $(Y_0, Y_1, \dots)$ of network structure defined in a way such that each Y_i provides more detailed information about the network structure and Y_n fully characterizes a network with n nodes. Orsini et al. (2015) have shown that the first three terms in the *dk*-series (Y = degrees + correlations + clustering/transitivity) are capable of almost fully defining local and global organization of most real-world networks that do not exhibit community structure.

As shown in Sections 1 and 2, most generative models are aimed at inferring the generative process using a single network observation. In our experiments with the dissimilarity space proposed above, we assume a single network randomly drawn from the network population serves as the input network G^* for the generative models. The rest of the networks in the population are treated as unobserved samples and are used to evaluate the performance of models on matching the variability in the network population. Although Section 4.1 only shows the result for one randomly drawn G^*, the analysis and conclusion are consistent for different G^* samples (See Supplementary Information).

4.1 Experimental results

In our experiments to evaluate the distributional properties of generative models, we propose to use the Kolmogorov–Smirnov (KS) statistic for evaluating the dissimilarity between networks based on node-level properties of degree, correlations, and clustering. To examine the ability of existing generative models to approximate the ground truth process using a single network observation (assuming it is representative of the true process with respect to the measures of interest), we propose two different experiments: (i) a controlled experiment where a known process is used to create a population of networks, and (ii) set of real-world networks that have most likely evolved from a common generative process (e.g., social interaction networks of different villages) or generative processes that share the same mechanisms.

4.1.1 Networks without community structure

Figure 4 shows the results for the first set of experiments when the Barabási & Albert (1999) and Forest Fire models (Leskovec et al., 2007) are used as the true processes A^*. For the second experiment, we consider the five real-world network populations described above, with results presented in Figures 5, 6, and 8. Results presented in Figures 4, 5, 6, and 8 are composed of three different plots:

1. Scatter plots below the diagonal show each synthesized/real network as a point in the network dissimilarity space, where the coordinates are computed using the Kolmogorov–Smirnov distance of the associated properties when the network is compared to the observed network G^* (the observed network itself is at the (0,0) position). Network models (colored triangles) showing higher overlap with networks originating from the true process (black dots) are better.
2. In the blocks above the diagonal, we evaluate the amount of overlap between $\mathbb{P}_{\mathfrak{D}_{G^*}}(A^*)$ and $\mathbb{P}_{\mathfrak{D}_{G^*}}(M)$ using the 2-D KS distance (Peacock, 1983) (lower the better). This quantifies the extent to which a given generative model is able to reproduce the distributional properties of the population representing the true process.
3. Plots along the diagonal show the density distributions of the Kolmogorov–Smirnov distance of the associated properties when the network is compared to the observed network G^*. Similar density distribution to the ground truth (black curves) implies good match in the properties.

https://doi.org/10.1017/nws.2019.63 Published online by Cambridge University Press

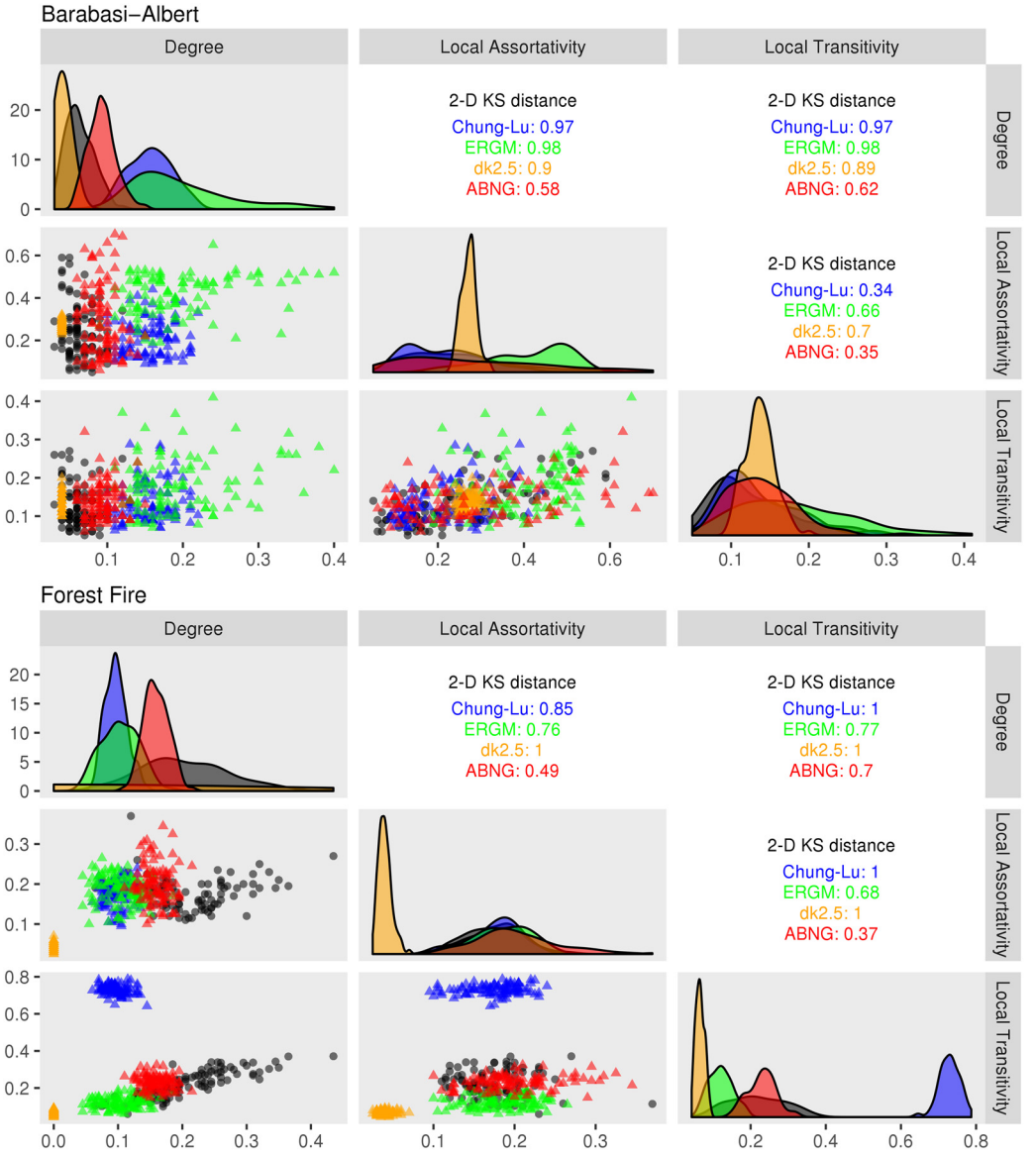

Figure 4. Empirical evaluation of the ability of network models to approximate the ground truth system based on observation of a single network. The Barabási–Albert and Forest Fire models are used as the true generators.

Based on Figures 4 and 5, we can easily conclude that ABNG consistently outperforms the other models considered here by replicating the natural variability of network populations when computed in the dissimilarity space for both the experimental settings. In social networks in Indian villages and Travian Trades networks, ERGM generates dense graphs (causing spikes to the right in degree KS plot) because of model degeneracy. The plots also show that *dk*-random graphs, which are considered to be the state of the art, fail to capture the variability of the true generative process and potentially over-fit the observed network. This leads us to question the fundamental idea behind *dk*-random graphs, that is, whether exactly preserving the distribution of differently sized subgraphs of a given network leads to a good model for real-world networks. In fact, in most cases we see that the Chung-Lu model, by matching the degree distribution in

https://doi.org/10.1017/nws.2019.63 Published online by Cambridge University Press

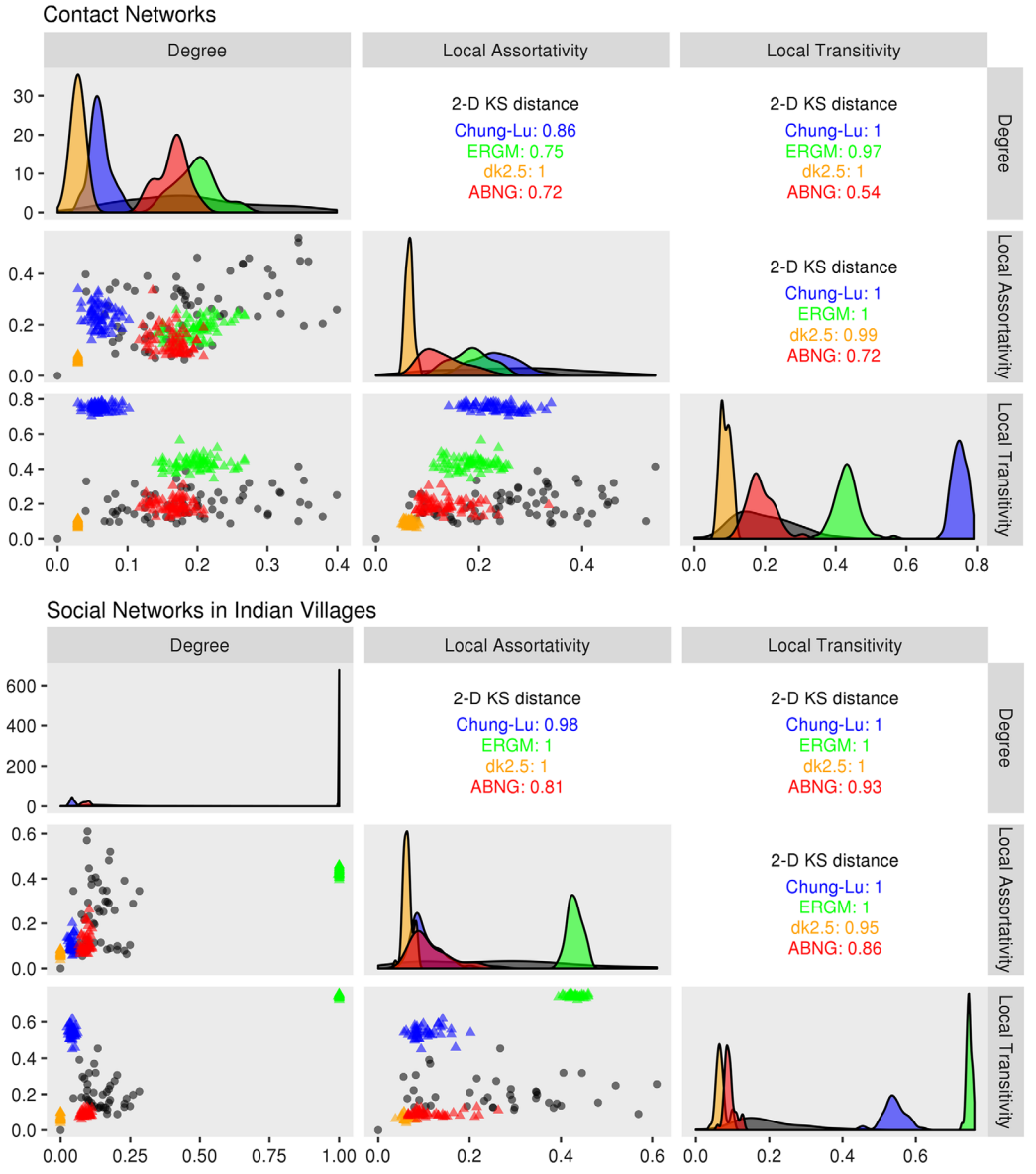

Figure 5. Empirical evaluation of the ability of network models to approximate the ground truth system based on observation of a single network. Two real-world datasets were considered: contact networks, and social networks in Indian villages.

expectation, outperforms *dk*-random graphs by synthesizing networks with more variability. These results highlight the need for evaluating the ability of a generative model to capture the distributional properties of a network population as comparing only with the observed network might produce misleading results.

The results in Figures 4 and 5 suggest that network models, when carefully designed, can potentially capture the structural variability in network populations (when evaluated in the dissimilarity space) using a single network as input. But when this analysis was extended to more network populations, all network models failed to synthesize populations that resemble the original one, and the results are presented in Figure 6.

https://doi.org/10.1017/nws.2019.63 Published online by Cambridge University Press

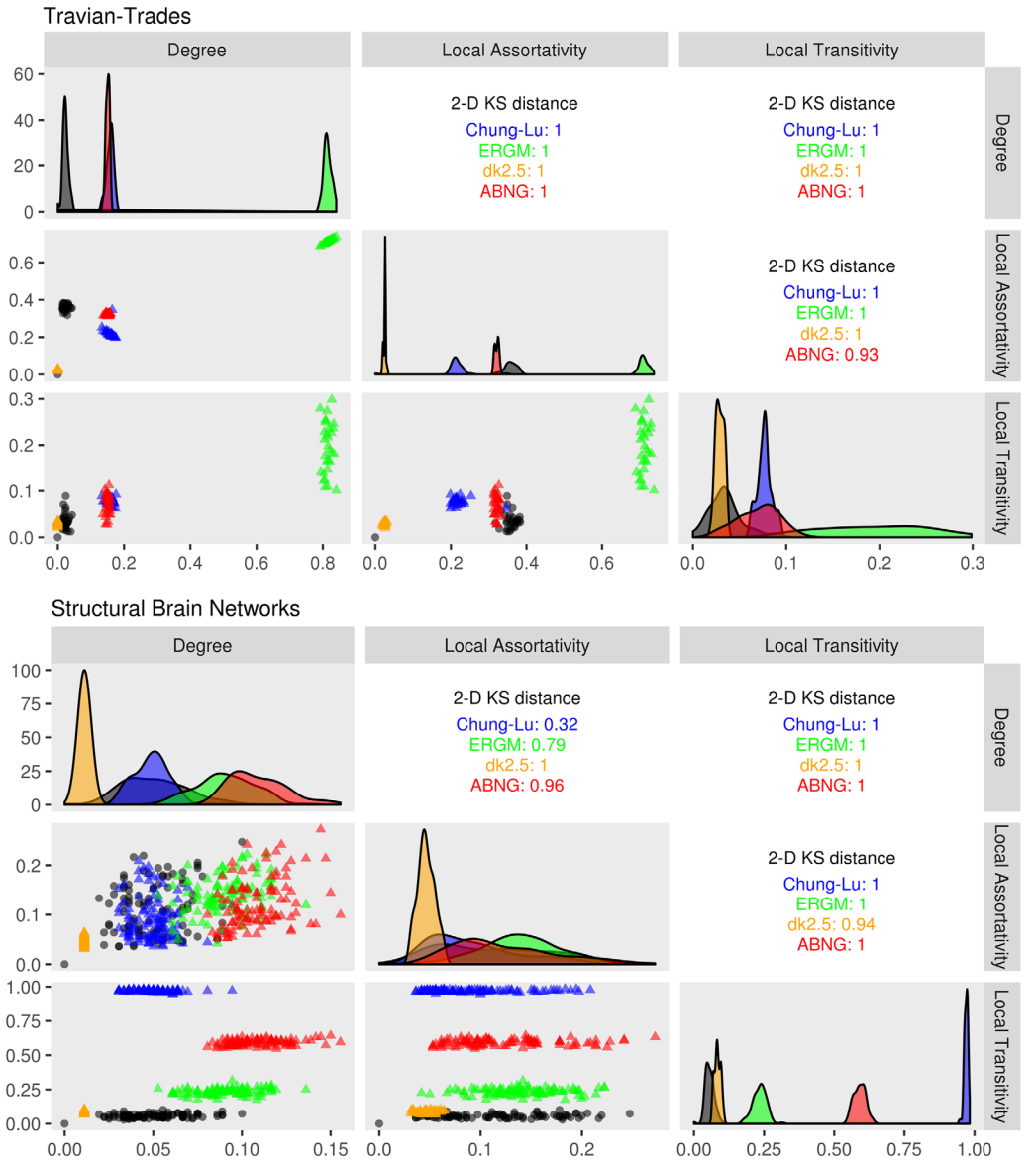

Figure 6. Empirical evaluation of the ability of network models to approximate the ground truth system based on observation of a single network. Two real-world datasets were considered: Travian trades and structural brain networks.

4.1.2 Networks with community structure

While the network dissimilarity space defined in Section 4 works well for networks without communities, it will prove ineffective for networks with community structures, which is a property seen in most real-world networks (Fortunato & Hric, 2016). In this section, we extend the network dissimilarity space by adding a fourth dimension to compare the community structures of two networks. The following procedure was used for comparing community structures in the extended network dissimilarity space:

1. Compute node memberships using a community detection algorithm (Infomap community detection algorithm (Rosvall & Bergstrom, 2008) was used in our experiments).

https://doi.org/10.1017/nws.2019.63 Published online by Cambridge University Press

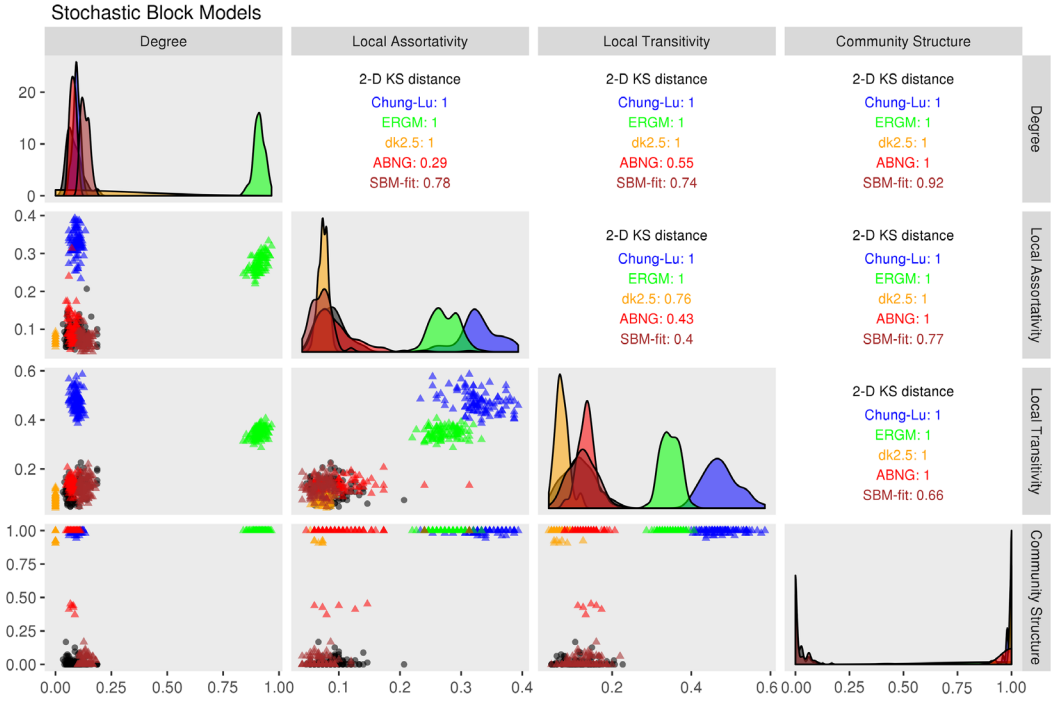

Figure 7. Empirical evaluation of the ability of network models to approximate the ground truth system based on observation of a single network. The stochastic block model is used as the true generator, and the ability of different models to replicate the community structure is tested.

2. Sort the communities based on sizes, that is, community 1 is the largest community.
3. Compare the sorted memberships using the normalized mutual information measure (Danon et al., 2005).

We also add the microcanonical SBM (Peixoto, 2017) (referred to as SBM-fit in the plots) to our set of generative models and evaluate its ability to replicate the community structure of these networks.

Again, we performed two different experiments to test the validity of our extended network dissimilarity space: (i) a controlled experiment where the true process is known, and (ii) set of real-world networks (with communities) that have most likely evolved from a common generative process. For the first case, we used the standard version of the SBM (Wasserman & Anderson, 1987; Faust & Wasserman, 1992) with three communities of different sizes, and the results can be seen in Figure 7. As expected, ABNG performs well on the original measures, but fails to reproduce the community structure, while the fitted SBM is the most likely candidate capable of replicating the true process. This is an expected result as the four original models are not designed to create networks with communities. Figure 8 shows the results for the networks of AS and Travian messages, where only the fitted SBM was able to capture some of the features of the true process. Results presented in Figure 8 show the inability of the microcanonical block model to reproduce the local transitivity of the true generative process, thus creating an exciting direction for future research.

In summary, our empirical analysis in the dissimilarity space has highlighted the discrepancy between observed network population and synthesized network as well as the importance of considering distributional properties of network populations for evaluating generative models

https://doi.org/10.1017/nws.2019.63 Published online by Cambridge University Press

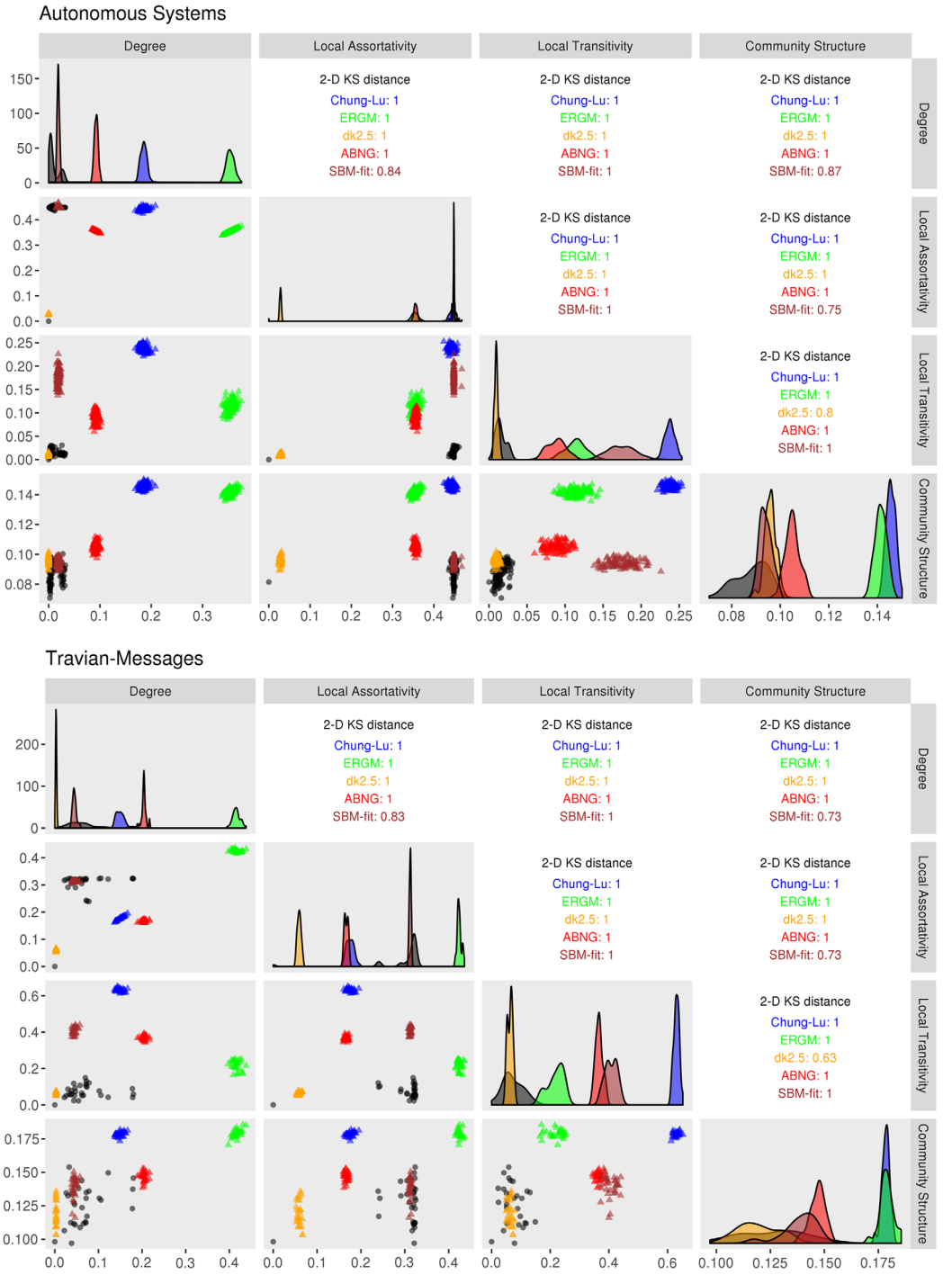

Figure 8. Empirical evaluation of the ability of network models to approximate the ground truth system based on observation of a single network. The ability of different models to replicate the community structure of networks of autonomous systems and Travian messages networks is tested.

https://doi.org/10.1017/nws.2019.63 Published online by Cambridge University Press

of complex networks. This shows that there is an urgent need to rethink the network modeling problem and create new models that are explicitly designed to reproduce the variability in the structural properties of network populations.

5. Evaluating variability in network populations

In previous sections, we empirically verified that current state-of-the-art models are ineffective in reproducing topological variability when the input is a single network. More specifically, while synthesized networks can match metrics to one network from a population, the dissimilarity distribution of synthesized networks $\mathbb{P}_{\mathfrak{D}_{G^*}}(M)$, and of the original population, $\mathbb{P}_{\mathfrak{D}_{G^*}}(A^*)$ do not match. Quantifying such inherent variability can help assess the performance of models in matching the distributional properties of the populations. Shannon entropy is a measure of the inherent uncertainty or variability of a single variable. Therefore, we can measure the Shannon entropy of a network population using a network as the random variable. For a given network population $\mathscr{G}_{A^*} = \{G_1(V_1, E_1), \ldots, G_k(V_k, E_k)\}$, assuming that each network G_i is observed with probability P_i, the entropy of the population $\mathscr{G}_{A^*}$ can be calculated using the following formula:

$$H = -\sum_{i=1}^{k} P_i \log P_i. \tag{3}$$

While this provides a convenient way of measuring the information content in a given network population, insufficient information about the process A^* means that we do not have explicit values for the P_i's. Using the formula in Equation (3) becomes even more complicated for real-world populations as the probabilities (P_i's) are unknown, and assigning equal probabilities to each network in the population can inadvertently lead to inappropriate conclusions by maximizing the entropy of the population.

An alternative method for quantifying the information content is to use some discrete random variable $X \in B$ as an event that is observed in every network in the population, where B is the set of all possible events. Then, the probability mass function of the random variable X in a network population given by $p(x_i) = Pr(X = x_i)$ can be used to evaluate the entropy of the population:

$$H_x = -\sum_{x_i \in B} p(x_i) \log (p(x_i)). \tag{4}$$

For a network population that contains networks defined on the same set of nodes, that is, $V_i = V\ \forall i$, the random variable X can be defined as events that are related to topological structure of subsets of nodes, such as density, clustering coefficient. For simplicity, we assume all networks in one population share the same set of nodes to avoid the additional topological variability caused by variability in nodes. Otherwise, the definition and computation of node variability are challenging, for example, the study of graph automorphism which is a relabeling of nodes that does not change the structure of graph. In fact, this assumption is satisfied in real-world populations created from brain networks of different individuals, traffic flow in a city across different days, social networks of the same group of people over time, etc. In this paper, we start our analysis with entropy measures defined on pairs of nodes in a network population. For example, we can measure the variability of an edge between any pair of nodes $v_i, v_j \in V$, or the *edge existence entropy* as

$$H_e(e_{ij}) = -(p_1 \log_2 p_1 + p_0 \log_2 p_0), \tag{5}$$

where $p_1 = Pr(e_{ij} = 1)$ and $p_0 = Pr(e_{ij} = 0)$ are the empirical distributions of edge e_{ij} in the network population. H_e thus measures the average uncertainty in the existence of an edge in a network population. Another measure of variability for a pair of nodes can be defined using

https://doi.org/10.1017/nws.2019.63 Published online by Cambridge University Press

the hop distance δ_{ij} between nodes $v_i, v_j \in V, i \neq j$. In this case, the random variable $D \in \Delta \subseteq \{1, \ldots, |V| - 1, \infty\}$ and its variability in the population can be used to define the *geodesic entropy*

$$H_d(\delta_{ij}) = -\sum_{d \in \Delta} p_d \log_{|\Delta|} p_d, \tag{6}$$

where p_d is the proportion of network samples in which the distance δ_{ij} between node v_i and v_j is d, and the set Δ is determined empirically. Using Equations (5) and (6), we can compute the edge existence and geodesic entropies for any pairs of nodes. To obtain variability of these measures over the entire network, mean edge existence and geodesic entropies of all pairs of nodes can be used to quantify the average variability of these measures in the network population.

As defined in Equation (4), these entropy measures can easily be extended to account for larger subsets of nodes, such as graphlets of varying sizes (Przulj et al., 2004), to quantify their variability in the network population. As a first step, we restricted our analysis to measure the entropy of edge existence and geodesic distances between pairs of nodes to account for variability in local density, strength of connection, and communication efficiency in a network population. As a result, these entropy measures can allow us to gauge the degree of uncertainty associated with topological structures in a network population.

5.1 Experimental results

In this section, we use the definitions of edge and geodesic entropies provided in Equations (5) and (6), respectively, to evaluate the variability of network population of 100 structural brain networks (described in Section 3). Among the real-world network populations described in Section 3, only the structural brain networks satisfy the assumption of a fixed node set, which is central to the measures proposed for evaluating variability of network populations in Section 5. Additionally, we also synthesize network populations using three models, namely configuration model, *dk*-random graphs, and action-based networks, and compare their variability with the real-world dataset. Instead of randomly picking one network from the network population as input, we parameterize each model using multiple input networks assuming that the additional information will enhance the ability of models to synthesize populations that are representative of the "true" distribution. In addition to the model training and synthesis on the complete dataset, network population of 100 networks, partial input is considered by bootstrapping to verify the modeling performance and the effect of sample size on entropy measures.

To verify the effect of sample size to the entropy methods, 20 replications are done for each sample size $s_i \in S = \{2, 3, 5, 10, 20, 50, 80, 100\}$. In each replication, s_i networks are randomly drawn from the 100 networks without replacement, thus creating a network population with sample size s_i, which is then used to compute edge existence entropy and geodesic entropy. Given a sampled network population, a mean degree sequence is estimated by averaging the degree sequences of the s_i networks in the population. This mean degree sequence is then used by the configuration model to synthesize a network population. Similarly, for ABNG, mean value of the user-defined characteristics in the network population can be used for learning action matrices for the population. For the *dk*-random graph model, a probability matrix is created by averaging the adjacency matrices of the s_i networks in the population. A representative network is then sampled from the probability matrix, which is repeatedly rewired using the 2.5*k*-rewiring scheme to synthesize network populations.

Figure 9 shows the edge and geodesic entropies of structural brain networks and networks synthesized using the configuration model, *dk*-random graphs, and ABNG. The real network populations lie on the left side of the space, and the synthesized network populations of three models are positioned right of them. The edge and geodesic entropy of the full dataset (100 structural brain networks) and population synthesized by different models are annotated in the plot. Partial population with bootstrapping are colored by their population sizes. For the three models

https://doi.org/10.1017/nws.2019.63 Published online by Cambridge University Press

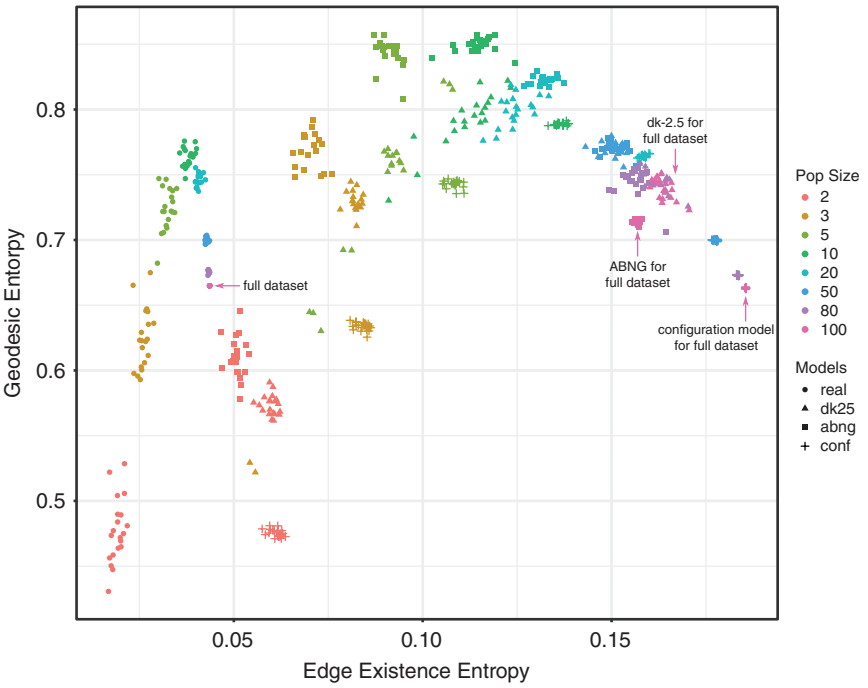

Figure 9. Edge existence entropy and geodesic entropy distributions for brain networks and synthetic networks with different sample sizes. Each point represents the edge and geodesic entropy of one network population containing multiple networks. With increasing population size, the geodesic entropy increases and gets peaked at around population size of 10 and then decreases. That is caused by insufficient sample size and network sparsity. Models that are better on matching the variability will have scatters closer to "full dataset".

and input, the edge existence entropy and geodesic entropy both increase with increasing sample size before 10 samples despite the geodesic entropy decreases when the population size is larger than 10. Both entropy measures converge for more than 50 samples. Generally, limited population size will reduce the accuracy of entropy measures.

The performance of network generative models on local variability can be evaluated by comparing the entropy of input network population and model synthesized network populations. All three models overestimate edge existence entropy which means in this scenario, true structural brain networks have lower variability in local connectivity than the models express overall. ABNG gets a better estimate than the other two models with respect to edge existence entropy. As for geodesic entropy, both ABNG and *dk*-random graph model overestimate it but configuration model estimates the geodesic entropy accurately. In conclusion, none of the three models can match the variability of input network population. They either overestimate the edge existence entropy or overestimate geodesic entropy.

Figure 10 shows the assortativity–transitivity space, similar to the one in Figure 3, for the 100 structural brain networks and network populations synthesized in the results presented in Figure 9. Instead of 20 replications for each population size setting, Figure 10 only shows 1 replication. Each point in the plot represents the assortativity and transitivity of one network. Among the three models, configuration model, *dk*-random graphs, and ABNG, *dk*-random graph model performs the best on matching the centroid which represents the average metrics. However, it has the lowest generalized variance implying it fails at matching the variability of the group. ABNG outperforms the other two models at matching the generalized variance implying its capability of capturing the variability of input networks. Furthermore, there is no significant relation between population size annotated by color and the goodness-of-fit in the space for all three models.

https://doi.org/10.1017/nws.2019.63 Published online by Cambridge University Press

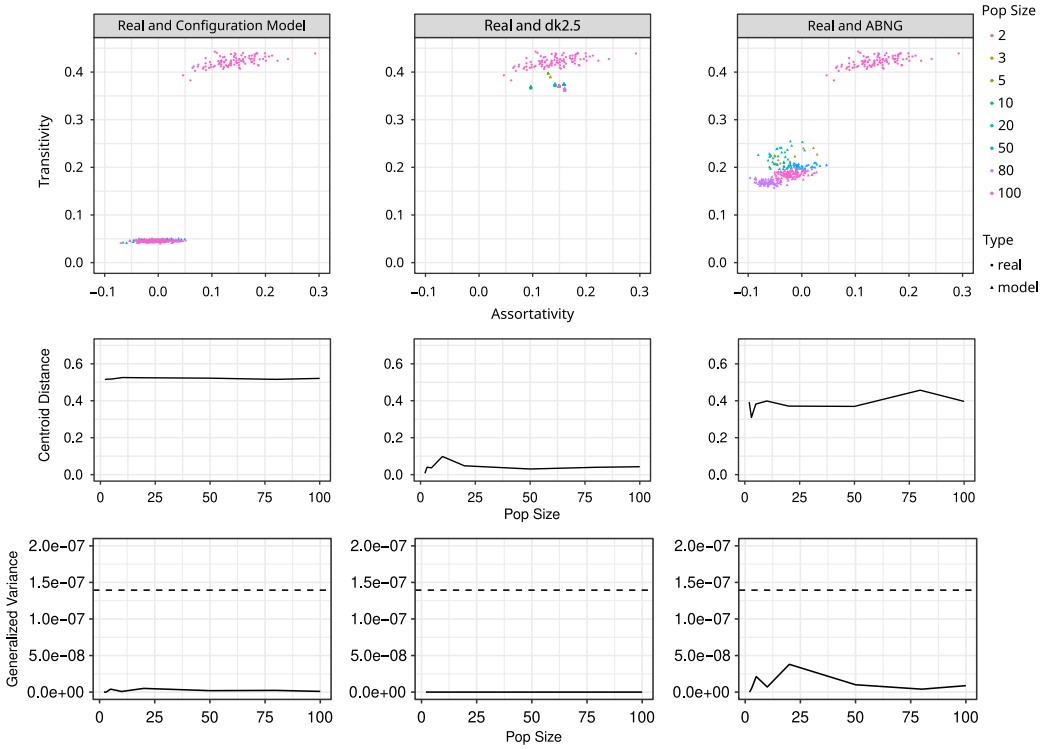

Figure 10. First row shows distribution of assortativity of vertex degree and transitivity in network populations of structural brain networks and networks synthesized from them by configuration model, *dk*-random graphs and ABNG. Second row shows the centroid distance between synthesized networks and input networks. Lower distance means better estimate on mean value. Last row shows the generalized variance of assortativity and transitivity of synthesized networks while the dashed line represents the level of input networks. Being closer to dashed line indicates better estimate on the variance on both metrics.

6. Conclusions

Traditional approaches for evaluating the ability of a network model to synthesize networks exhibiting real-world characteristics have compared the similarity of the synthesized networks with the observed network. While this approach assumes that the particular observation is representative of the underlying process that created the observation, it does not account for the natural variability of the population from which it is sampled. Our experiments have highlighted the importance of considering network populations for evaluating generative models. Although it is difficult to obtain data corresponding to network populations, we have shown that it is possible to establish a baseline test set to evaluate the ability of a network model to capture the distribution of network populations. This test set can then be used for preliminary validation of a network model before it is used for drawing conclusions about real-world networks. Moreover, a quantitative evaluation of the distributional goodness-of-fit has been left for the future.

The need for devising generative models for network population is raised in this paper given limited models that can learn the properties of network populations. More specifically, the devised models should be capable of matching the inherent variability of real-world network populations when available. Instead of modifying and transforming existing network generative models to fit the representative/average network of a network population, we need to develop models that can extract and quantify the variability existing in network populations. For example, Bayesian modeling frameworks are being successfully used to learn the joint distribution of edges in network

https://doi.org/10.1017/nws.2019.63 Published online by Cambridge University Press

population that shares the same set of vertices (Gollini & Murphy, 2016; Durante et al., 2017; Lunagómez et al., 2019), though the topology of networks is not explicitly fitted.

As discussed in Section 5, population size plays a crucial role in a data-driven evaluation of the variability in network populations. Preliminary results show that both entropy measures proposed in Section 5 are biased when evaluated for small population size. This can be a serious limitation when the population sizes are small. Future work on the development of an unbiased measure of variability accompanied with a quantitative analysis of the effect of network sample size and power can guide the sampling process of networks and the development of improved generative network models.

Acknowledgments. We thank Enrico Amico for assistance with the Human Connectome Project data. We thank Bryan Chong and the anonymous reviewers who provided insights and comments that greatly improved this paper. This research was supported by the National Science Foundation under Grant No. 1762633.

Conflict of interest. The authors declare no conflicts of interest.

Supplementary materials. For supplementary material for this article, please visit http://doi.org/10.1017/nws.2019.63

Note

1 Code available at https://github.com/dlguo/network-variability.

References

Alderson, David L. (2008). OR FORUM—Catching the 'Network Science' Bug: Insight and Opportunity for the Operations Researcher. *Operations Research*, *56*(5), 1047–1065.

Amico, E., & Goni, J. (2018). Mapping hybrid functional-structural connectivity traits in the human connectome. *Network Neuroscience*, *2*(3), 306–322.

Anand, K., & Bianconi, G. (2009). Entropy measures for networks: Toward an information theory of complex topologies. *Physical Review E – Statistical, Nonlinear, and Soft Matter Physics*, *80*(4), 1–4.

Anderson, C. J., Wasserman, S., & Crouch, B. (1999). A p* primer: Logit models for social networks. *Social Networks*, *21*(1), 37–66.

Arora, V., & Ventresca, M. (2016). A Multi-objective optimization approach for generating complex networks. *Proceedings of the 2016 on genetic and evolutionary computation conference companion – gecco 2016 companion* (pp. 15–16). New York, NY, USA: ACM Press.

Arora, V., & Ventresca, M. (2017). Action-based modeling of complex networks. *Scientific Reports*, *7*(1), 66–73.

Arora, V., & Ventresca, M. (2019). Evaluating the natural variability in generative models for complex networks. *Studies in computational intelligence* (pp. 743–754). Springer International Publishing.

Avena-Koenigsberger, A., Goni, J., Sole, R., & Sporns, O. (2014). Network morphospace. *Journal of the Royal Society Interface*, *12*(103), 20140881–20140881.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013). The diffusion of microfinance. *Science*, *341*(6144), 1236498.

Barabasi, A.-L. (2002). *Linked: The new science of networks* (1st ed.) Basic Books.

Barabasi, A.-L. (2012). The network takeover. *Nature Physics*, *8*(1), 14–16.

Barabasi, A.-L. (2016). *Network science*. Cambridge University Press.

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512.

Bender, E. A., & Canfield, E. R. (1978). The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory, Series A*, *24*(3), 296–307.

Bianconi, G. (2008). The entropy of randomized network ensembles. *Epl (Europhysics Letters)*, *81*(2), 28005.

Casiraghi, G., Nanumyan, V., Scholtes, I., & Schweitzer, F. (2016). Generalized hypergeometric ensembles: Statistical hypothesis testing in complex networks. *7*, *arXiv*, 1–5.

Chung, F., & Lu, L. (2002a). Connected components in random graphs with given expected degree sequences. *Annals of Combinatorics*, *6*(2), 125–145.

Chung, F., & Lu, L. (2002b). The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, *99*(25), 15879–15882.

Cohen, R., & Havlin, S. (2010). *Complex networks: Structure, robustness, and function*. Cambridge University Press.

Danon, L., Díaz-Guilera, A., Duch, J., & Arenas, A. (2005). Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, *2005*(09), P09008–P09008.

De Domenico, M., & Biamonte, J. (2016). Spectral entropies as information-theoretic tools for complex network comparison. *Physical Review X*, *6*(4), 041062.
Dehmer, M., & Mowshowitz, A. (2011). A history of graph entropy measures. *Information Sciences*, *181*(1), 57–78.
Donnat, C., & Holmes, S. (2018). Tracking network dynamics: A survey of distances and similarity metrics. *arxiv preprint*, *1*.
Durante, D., Dunson, D. B., & Vogelstein, J. T. (2017). Nonparametric Bayes Modeling of Populations of Networks. *Journal of the American Statistical Association*, *112*(520), 1516–1530.
Emmert-Streib, F., & Dehmer, M.. (2012). Exploring statistical and population aspects of network complexity. *Plos One*, *7*(5), e34523.
Emmert-Streib, F., Dehmer, M., & Shi, Y. (2016). Fifty years of graph matching, network alignment and network comparison. *Information Sciences*, *346–347*(6), 180–197.
Erdös, P., & Rényi, A. (1959). On random graphs. *Publicationes Mathematicae Debrecen*, *6*, 290–297.
Erdös, P., & Rényi, A. (1960). On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, *5*(1), 17–61.
Faloutsos, M., Faloutsos, P., & Faloutsos, C. (1999). On power-law relationships of the Internet topology. *Proceedings of the conference on applications, technologies, architectures, and protocols for computer communication - sigcomm '99*, vol. 50 (251–262). New York, NY, USA: ACM Press.
Faust, K., & Wasserman, S. (1992). Blockmodels: Interpretation and evaluation. *Social Networks*, *14*(1–2), 5–61.
Fields, S., & Song, O.-K. (1989). A novel genetic system to detect protein–protein interactions. *Nature*, *340*(6230), 245–246.
Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports*, *659*(11), 1–44.
Ginestet, C. E., Li, J., Balachandran, P., Rosenberg, S., & Kolaczyk, E. D. (2017). Hypothesis testing for network data in functional neuroimaging. *The Annals of Applied Statistics*, *11*(2), 725–750.
Goldenberg, A., Zheng, A. X., Fienberg, S. E., & Airoldi, E. M. (2009). A survey of statistical network models. *Foundations and Trends® in Machine Learning*, *2*(3), 129–233.
Gollini, I., & Murphy, T. B. (2016). Joint Modeling of multiple network views. *Journal of Computational and Graphical Statistics*, *25*(1), 246–265.
Gutfraind, A., Meyers, L. A., & Safro, I. (2012). Multiscale Network Generation. *arXiv:1207.4266*, *7*, 28.
Hajibagheri, A., Lakkaraju, K., Sukthankar, G., Wigand, R. T., & Agarwal, N. (2015). Conflict and communication in massively-multiplayer online games. *International conference on social computing, behavioral-cultural modeling, and prediction* (pp. 65–74). Springer.
Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, *97*(460), 1090–1098.
Huberman, B. A. (2003). *The laws of the Web: Patterns in the ecology of information*. MIT Press.
Hunter, D. R. (2007). Curved exponential family models for social networks. *Social Networks*, *29*(2), 216–230.
Isella, L., Stehlé, J., Barrat, A., Cattuto, C., Pinton, J.-F., & den Broeck, W. (2011). What's in a crowd? Analysis of face-to-face behavioral networks. *Journal of Theoretical Biology*, *271*(1), 166–180.
Ladyman, J., Lambert, J., & Wiesner, K. (2013). What is a complex system? *European Journal for Philosophy of Science*, *3*(1), 33–67.
Leskovec, J., Kleinberg, J., & Faloutsos, C. (2007). Graph evolution: Densification and shrinking diameters. *Acm Transactions on Knowledge Discovery from Data*, *1*(1), 2–es.
Lunagómez, S., Olhede, S. C., & Wolfe, P. J. (2019). Modeling network populations via graph distances. *arxiv preprint*, 1–26.
Mahadevan, P., Krioukov, D., Fall, K., & Vahdat, A. (2006). Systematic topology analysis and generation using degree correlations. *Acm Sigcomm Computer Communication Review*, *36*(4), 135.
Maslov, S., & Sneppen, K. (2002). Specificity and stability in topology of protein networks. *Science*, *296*(5569), 910–913.
McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444.
Menezes, T., & Roth, C. (2019). Automatic discovery of families of network generative processes. *Dynamics on and of complex networks iii* (pp. 83–111).
Milo, R., Shen-Orr, S., Itzkovitz, S., & Kashtan, N. (2002). Network Motif: Simple building blocks of complex networks. *Science*, *298*(5594), 298.
Molloy, M., & Reed, B. (1995). A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms*, *6*(2–3), 161–180.
Moreno, J. L. (1934). *Who shall survive? A new approach to the problem of human interrelations*. Nervous and mental disease monograph series, no. 58. Washington: Nervous and Mental Disease Publishing Co.
Moreno, S., & Neville, J. (2013). Network hypothesis testing using mixed Kronecker product graph models. In H. Xiong, G. Karypis, B. M. Thuraisingham, D. J. Cook, & X. Wu (eds), *13th international conference on data mining* (pp. 1163–1168).
Moreno, S., & Neville, J. (2009). An Investigation of the distributional characteristics of generative graph models. *Proceedings of the 1st workshop on information in networks*.
Moreno, S., Neville, J., & Kirshner, S. (2018). Tied Kronecker product graph models to capture variance in network populations. *Acm Transactions on Knowledge Discovery from Data*, *20*(3), 1–40.
Newman, M. E. J. (2003a). Mixing patterns in networks. *Physical Review E*, *67*(2), 026126.
Newman, M. E. J. (2003b). The structure and function of complex networks. *Siam Review*, *45*(2), 167–256.

https://doi.org/10.1017/nws.2019.63 Published online by Cambridge University Press

Newman, M. (2008). The physics of networks. *Physics Today*, *61*(11), 33–38.
Orsini, C., Dankulov, M. M., Colomer-de Simón, P., Jamakovic, A., Mahadevan, P., Vahdat, A., Bassler, K. E., Toroczkai, Z., Boguñá, M., Caldarelli, G., Fortunato, S., & Krioukov, D. (2015). Quantifying randomness in real networks. *Nature communications*, **6**(May), 8627.
Palla, G., Derényi, I., Farkas, I., & Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, *435*(7043), 814–818.
Park, J., & Newman, M. E. J. (2004). Statistical mechanics of networks. *Physical Review E*, *70*(6), 066117.
Pastor-Satorras, R., Vázquez, A., & Vespignani, A. (2001). Dynamical and correlation properties of the internet. *Physical Review Letters*, *87*(25), 258701–1.
Peacock, J. A. (1983). Two-dimensional goodness-of-fit testing in astronomy. *Monthly Notices of the Royal Astronomical Society*, *202*(3), 615–627.
Peixoto, T. P. (2012). Entropy of stochastic blockmodel ensembles. *Physical Review E*, *85*(5), 056122.
Peixoto, T. P. (2015). Model selection and hypothesis testing for large-scale network models with overlapping groups. *Physical Review X*, *5*(1), 1–20.
Peixoto, T. P. (2017). Nonparametric Bayesian inference of the microcanonical stochastic block model. *Physical Review E*, *95*(1), 1–21.
Perc, M., Jordan, J. J., Rand, D. G., Wang, Z., Boccaletti, S., & Szolnoki, A. (2017). Statistical physics of human cooperation. *Physics Reports*, *687*(5), 1–51.
Pinar, A., Seshadhri, C., Kolda, T. G., Pinar, A., & Kolda, T. G. (2012). The similarity between stochastic Kronecker and Chung-Lu graph models. M. Zaki, Z. Obradovic, P. N. Tan, A. Banerjee, C. Kamath & S. Parthasarathy (Eds.), *Twelfth Siam international conference on data mining* (pp. 1071–1082). Philadelphia, PA: Society for Industrial and Applied Mathematics.
Piraveenan, M., Prokopenko, M., & Zomaya, A. Y. (2008). Local assortativeness in scale-free networks. *Epl (Europhysics Letters)*, *84*(2), 28002.
Price, D. J. D. S. (1965). Networks of scientific papers. *Science*, *149*(3683), 510–515.
Przulj, N., Corneil, D. G., & Jurisica, I. (2004). Modeling interactome: Scale-free or geometric? *Bioinformatics*, *20*(18), 3508–3515.
Reichardt, J., & Bornholdt, S. (2006). Statistical mechanics of community detection. *Physical Review E*, *74*(1), 016110.
Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social Networks*, *29*(2), 173–191.
Rosvall, M., & Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, *105*(4), 1118–1123.
Schieber, T. A., Carpi, L., Díaz-Guilera, A., Pardalos, P. M., Masoller, C., & Ravetti, M. G. (2017). Quantification of network structural dissimilarities. *Nature Communications*, *8*(May 2016), 13928.
Snijders, T. A. B., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2004). New specifications for exponential random graph models. *Sociological Methodology*, *44*.
Soundarajan, S., Eliassi-Rad, T., & Gallagher, B. (2014). A guide to selecting a network similarity method. *Proceedings of the 2014 Siam international conference on data mining* (pp. 1037–1045). Philadelphia, PA: Society for Industrial and Applied Mathematics.
Stanton, I., & Pinar, A. (2012). Constructing and sampling graphs with a prescribed joint degree distribution. *Journal of Experimental Algorithmics*, *17*(1), 3.1.
Strauss, D. (1986). On a general class of models for interaction. *Siam Review*, *28*(4), 513–527.
Sweet, T. M., Thomas, A. C., & Junker, B. W. (2012). Hierarchical network models for education research. *Journal of Educational and Behavioral Statistics*, *38*(3), 295–318.
Vallès-Català, T., Peixoto, T. P., Sales-Pardo, M., & Guimerà, R. (2018). Consistencies and inconsistencies between model selection and link prediction in networks. *Physical Review E*, *97*(6), 062316.
Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., & Ugurbil, K. (2013). The WU-Minn human connectome project: An overview. *Neuroimage*, *80*(10), 62–79.
Views, R. (2000). *University of oregon route views project*.
Wasserman, S., & Pattison, P. (1996). Logit models and logistic regressions for social networks. *Psychometrika*, *60*, 401–425.
Wasserman, S., & Anderson, C. (1987). Stochastic a posteriori blockmodels: Construction and assessment. *Social Networks*, *9*(1), 1–36.
Watts, D. J, & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442.
White, J. G., Southgate, E., Thomson, J. N., & Brenner, S. (1986). The structure of the nervous system of the Nematode Caenorhabditis elegans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *314*(1165), 1–340.
Zheng, B., Wu, H., Kuang, L., Qin, J., Du, W., Wang, J., & Li, D. (2014). A simple model clarifies the complicated relationships of complex networks. *Science Report*, *4*, 6197.

Cite this article: Arora V., Guo D., Dunbar K. D., and Ventresca M. (2020). Examining the variability in network populations and its role in generative models. *Network Science* **8**, S43–S64. https://doi.org/10.1017/nws.2019.63