Inferring Exoplanet Disequilibria with Multivariate Information in Atmospheric Reaction Networks

Theresa Fisher1,5, Hyunjoo Kim1,2,3,5, Camerian Millsaps1, Michael Line1, and Sara I. Walker1,2,3,4

1 School of Earth and Space Exploration, Arizona State University, Tempe, AZ 85281, USA; sara.i.walker@asu.edu
2 Beyond Center for Fundamental Concepts in Science, Arizona State University, Tempe, AZ 85281, USA
3 ASU-SFI Center for Biosocial Complex Systems, Arizona State University, Tempe, AZ 85281, USA
4 Santa Fe Institute, Santa Fe, NM 87501, USA

Received 2021 April 20; revised 2022 March 10; accepted 2022 March 26; published 2022 July 15

Abstract

Inferring properties of exoplanets from their atmospheres presents technical challenges in data collection due to low resolution and low signal-to-noise ratio (S/N) and theoretical challenges in the predictions made from forward-modeling due to errors introduced via incomplete or inaccurate assumptions in atmospheric physics and chemistry. The combination of these factors makes developing techniques to identify the most predictive features robust to low S/N and model error an increasingly important challenge for exoplanet science. Here we implement a multivariate approach to identify optimal predictors of the state of disequilibria. As a case study we focus on the prediction of vertical mixing (parameterized as eddy diffusion) in hot Jupiter atmospheres. We use multivariate information contained in molecular abundances, reaction network topology, and Gibbs free energy to demonstrate the variation in prediction efficacy of the vertical mixing coefficient ($K_{zz}$) from different model information. While current approaches target inferring molecular abundances from spectral data, our results indicate that the set of optimal predictors of $K_{zz}$ varies with planetary properties such as irradiation temperature and metallicity. In most cases, multivariate data composed of network topological variables, which capture system-level features, perform as well as the set of optimal predictors and better than any individual variable. We discuss future directions, where identifying the set of optimal predictors should be useful for quantitatively ranking atmospheres in terms of their distance from thermochemical equilibrium, provide target variables for the development of new tools for inverse modeling, and provide applications to the longer-term goal of detection of disequilibria associated with life.

Unified Astronomy Thesaurus concepts: Exoplanet atmospheres (487)

1. Introduction

Exoplanets exhibit a diverse range of physical properties (irradiation temperature, mass, composition, etc.) far exceeding that observed within our own solar system (Seager 2013). Characterizing these unknown worlds based on the limited and low-resolution data available to current and upcoming missions presents a significant challenge. The lack of solar system analogs for the majority of exoplanets discovered means that our knowledge about them will be based entirely on the efficacy of our tools for remote inference. Crucial to this efficacy is the ability to detect—and distinguish between—various drivers of novel or disequilibrium chemistry, including but not limited to alien life (Walker et al. 2018). Near-term technologies will continue to return exclusively low-resolution, low signal-to-noise ratio (S/N) data for the foreseeable future, motivating the development of new methods to meet these compounding challenges (Fuji et al. 2018).

Current state-of-the-art inference approaches include Bayesian retrieval methods, which extract atmospheric information from spectra by forward-modeling atmospheric chemistry under reasonable assumptions and then fitting forward-model predictions to data through a parameterized model of the atmosphere combined with parameter estimation tools like Markov Chain Monte Carlo (Benneke & Seager 2012; Line et al. 2015; Madhusudhan 2019; Madhusudhan et al. 2011a). A limitation of this approach is that our ability to infer exoplanetary properties is heavily dependent on the accuracy of forward-model assumptions in capturing the relevant planetary physics and chemistry—a challenge given the aforementioned lack of solar system analogs for most exoplanets. Recently, there has also been intensifying interest in the application of machine-learning approaches to retrieval methods (Cobb et al. 2019; Hayes et al. 2020). These also enable predictive inferences if trained on reliable model data. However, like Bayesian retrieval methods, this requires first having accurate models to generate the training data sets.

The combination of these factors suggests that developing techniques to identify predictive features in atmospheric models, or developing new frameworks entirely (robust to both low S/N and model error), will become increasingly important as we search to identify atmospheric biosignatures beyond our solar system in the decades to come (Catling et al. 2018; Kiang et al. 2018; Walker et al. 2018). Forward-modeling approaches use thermochemical reactions and their rates to simulate possible exoplanet atmospheres with specified irradiation temperature, composition, and metallicity. Using a grid of models, likelihood distributions of key species abundances can be predicted and used to infer planetary properties, such as vertical mixing strength, $K_{zz}$, from atmospheric data (see Figure 1, left panel). However, other suites of variables can describe the same models and could hold potential to build more accurate inferences from limited data, as outlined in the right panel of Figure 1. This includes system-

---

5 These authors contributed equally to this work.
level properties of collective interactions among molecules within an atmosphere. These can be captured by quantifying the topology of a complex network representing the chemical reactions in a planetary atmosphere (Cenler & Dittrich 2007; Gleiss et al. 2001; Newman et al. 2011), which coarse-grains details of the specific atmospheric chemistry. It also includes calculating thermodynamic quantities, such as the Gibbs free energy, which underlie proposed measures of thermodynamic chemical disequilibria (Krissansen-Totton et al. 2018); (Krissansen-Totton et al. 2016; Line & Yung 2013; Molaverdikhani et al. 2019; Simoncini et al. 2013). Network-theoretic and thermodynamic measures have been proposed as possible candidate measures of global biological activity on terrestrial worlds (Krissansen-Totton et al. 2016) and could provide new targets for direct inference, rather than retrieving just species abundances alone.

Identifying the most informative model features is a first step to the development of next-generation inference tools that build on the most informative and least error-prone model variables. We test the predictive information across a multivariate suite of properties, exploring the chemical disequilibria of hot Jupiters in terms of a relatively well-understood driver of chemical disequilibrium, the vertical mixing coefficient, $K_{zz}$. Disequilibria are a prime target for developing new frameworks for remote inference because they can be diagnostic of planetary processes that are not themselves directly observable and because it is speculated that they can be driven by life (for terrestrial worlds; Krissansen-Totton et al. 2016). Here, however, we focus our analyses on hot Jupiters. Our motivation is threefold: (1) hot Jupiters remain one of the few classes of exoplanets with readily characterizable atmospheres (Seager & Deming 2010; Crossfield 2015; Madhusudhan 2019), (2) many of the chemical kinetics involved have been thoroughly investigated in laboratory and theoretical settings (Venot et al. 2012; Fortney et al. 2021), and (3) a relatively large number of example planets are known. Additionally, chemical kinetics and disequilibria drivers (mixing and photochemistry) have been modeled extensively in the atmosphere of irradiated hot Jupiters (e.g., Zahnle et al. 2009; Moses et al. 2011; Venot et al. 2012; Tsai et al. 2017). In particular, the effect of vertical mixing on the distance from equilibrium tends to weaken as temperatures increase (e.g., Line & Yung 2013). This existing body of research provides a substantial reference point for comparison to the techniques we employ here. We show how some variables of the models or combinations thereof are better predictors of vertical mixing coefficient, $K_{zz}$, in hot Jupiter atmospheres than others, as well as how the most predictive variables change with planetary conditions.

In Section 1, we show how tracking collective patterns of interdependent interactions among species in an atmosphere (e.g., via its network representation) can better constrain the likely $K_{zz}$ value with much greater resolution than with just the observable species data alone. In Section 2, we also show how network properties tend to be more robust to observational uncertainty or perturbations representing reaction data missing from the model. In Section 3, we discuss implications for the development of new inference methods that directly target the most informative atmospheric variables. Overall, our results indicate that a multivariate analysis, treating atmospheres as complex systems and leveraging quantitative frameworks from complexity science, allows new opportunities for understanding exoplanet atmospheric properties from remote data.

Figure 1. Left: “standard” analysis pipeline used to infer atmospheric $K_{zz}$ by generating a likelihood distribution from the abundance of key species. Right: schematic of the analysis pipeline used for the multivariate approach to identify optimal predictors. Inverse modeling, based on observational data, is used to create a model of a planet’s atmospheric chemistry. From this model, the list of thermodynamically favorable reactions and their rates is used to construct the chemical reaction network, for which topology, key species abundances, and Gibbs free energy are measured. The data then form the basis of an interpolation function, which is fed a selection of initial physical conditions, drawn from a Gaussian distribution, to determine what the corresponding network parameters are if a network was to be constructed using those initial physical conditions. The resulting distributions can then be used to quantitatively rank any given atmosphere with another in terms of distance from thermochemical disequilibrium and can provide target variables for development in future inverse modeling tools.
2. Methods

2.1. Atmospheric Modeling with VULCAN

We used the VULCAN modeling package (Tsai et al. 2017) to simulate the H\textsubscript{2}–C\textsubscript{2}–N–O atmospheric chemistry of hot Jupiters over a grid of input parameters: irradiation temperatures (centered on 400–3000 K, in ~100 K increments; this represents the equilibrium temperature at the substellar point; after Heng & Showman 2015), metallicities (incremented as 0.1, 1, 3, 10, 30, 100, and 300 times that of the solar system, evaluated on a log scale), pressures (from 50 to 150 mb, reflecting the likely depth in the atmosphere that will be observable, in increments of ~50 mb), and, finally, vertical mixing coefficients, with values of $K_{zz} = 0$ (thermochemical equilibrium), $10^6$, $10^8$, and $10^{10}$ s\(^{-1}\) cm\(^{-2}\) (in this paper, the units of $K_{zz}$ are assumed to be s\(^{-1}\) cm\(^{-2}\) unless mentioned otherwise). This made for a total of 1882 model runs simulated over the grid. For all simulations, we used VULCAN’s default temperature–pressure profile, which is based on the hot Jupiter HD 189733b (Moses et al. 2011; Tsai et al. 2017).

VULCAN computes the abundances of atmospheric species by solving a set of mass continuity equations written as

$$\frac{\partial n_i}{\partial t} = P_i - L_i - \frac{\partial f}{\partial z}, \tag{1}$$

where $n_i$ equals the number density of the $i$th species, $t$ denotes time, and $z$ is the spatial coordinate in the vertical/radial direction. $P_i$ and $L_i$ are the production and loss rates of the $i$th species (cm\(^{-3}\) s\(^{-1}\)), respectively. The transport flux, $f$, can be written as

$$f = -K_{zz} n_{total} \frac{\partial X_i}{\partial z}, \tag{2}$$

where $X_i$ represents the mixing ratio of the $i$th species defined as $X_i = n_i / n_{total}$ and $n_{total}$ is the number density of all species. The vertical mixing coefficient, $K_{zz}$, as the primary driver of the transport flux, represents how much mixing is occurring between the lower layers of the atmosphere, which are at higher temperature and pressure (in the absence of photochemical inputs), and the upper layers, which are at lower temperature and pressure.

Thus, $K_{zz}$ can influence the mixing ratio of a given species in an atmospheric layer, which in turn affects the reaction rate of the reactions that include that species. Since the edges of the network we construct are weighted by the respective reaction rates, $K_{zz}$ can impact the topology of the resulting weighted chemical reaction network. Thus, we might expect network representations, which are a coarse-graining of system-level features of atmospheric chemistry, to vary in a manner dependent on $K_{zz}$—if these variations change in a systematic, predictable way, it indicates that topology could be a useful correlate of vertical-mixing-induced disequilibria.

At $K_{zz} = 0$, there is no vertical mixing between layers, and thus we take this to be the equilibrium scenario. As the atmosphere becomes more mixed, it moves away from equilibrium because the species thermochemically abundant within the hotter, deeper layers are dredged up to the shallower, cooler, low-pressure areas of the atmosphere (where they are not thermochemically favorable) at a rate faster than that in which the kinetics can reestablish thermochemical equilibrium. Thus, $K_{zz}$ is often used as a proxy for an atmosphere’s distance from chemical equilibrium when it is the dominant driver of nonequilibrium conditions (e.g., in the absence of photochemistry).

2.2. Quantifying the Characteristic Topology of Chemical Reaction Networks

We constructed chemical reaction networks (CRNs) from the simulated atmospheric data of hot Jupiters for each model on our grid via VULCAN. Systems of interacting components can be projected onto network representations to study system-level characteristics emerging from the collective interdependent interactions of these components. While network analyses are widespread in their application in the study of complex systems, such as biological or technological systems, they have so far seen less application to astronomical systems, although a few notable examples do exist. Solé and Munteanu (2004) compared the topology of CRNs of the atmosphere of Earth to those of Titan, Mars, and Jupiter and found that Earth’s exhibited unique hierarchical and modular properties. The CRNs of the interstellar medium have also been analyzed (Jolley & Douglas 2010) and are distinguishable from the CRNs of biological systems (Jolley & Douglas 2012). Estrada (2012) further evaluated network topology as a potential biosignature and also introduced criteria of quantifying a network’s ability to return to thermodynamic equilibrium. Otherwise, however, most atmospheric modeling has focused strictly on recovering a few key species abundances (see the left panel of Figure 1).

We introduce a network-theoretic approach here as a method for statistical analysis of the collective interdependent interactions among chemical species in exoplanetary atmospheres. This is a different subset of information available to remote inference, to contrast with those strictly focused on individual species abundance data alone, as it deals with interactions between species. We expect that this will provide a pathway toward the development of new systems-level tools for atmospheric reaction network inference in exoplanet atmospheres, useful across different planet types. We are particularly motivated to develop these approaches for their longer-term applicability to terrestrial world atmospheres (Walker et al. 2018). In recent years it has become increasingly recognized that current biosignature gases such as CH\textsubscript{4} and O\textsubscript{2} can only be interpreted as biosignatures when considering their planetary context (Meadows 2017; Schwieterman et al. 2018). Inferring the global structure of a network from atmospheric data affords the possibility of tracking not just individual species abundances but also systems-level properties emerging from collective chemical reactions in planetary atmospheres. Thus, understanding the properties of network representations of atmospheric chemistry may enable us to better characterize the contextual measurements required to probe how life can drive features of atmospheric chemistry by quantifying them in terms of reaction network topology (Walker et al. 2018). However, to even approach developing this as a method for biosignature detection, we must first investigate how a network-theoretic approach can help us understand planetary atmospheres in the absence of life, when physical and chemical conditions are relatively well constrained, motivating our work on hot Jupiter atmospheres.

We constructed network representations, also called graphs, of the chemical reactions for hot Jupiter atmosphere models by first assigning each species present in the model as a node and
linking two nodes by an edge when two species participate in
the same chemical reaction as a reactant and a product,
respectively. Note that projecting real systems into abstract,
graphical representations as complex networks is nontrivial and
many choices are possible. For example, one could choose a
bipartite representation where both chemical species and
reactions are treated as two distinct classes of nodes
(see, e.g., Montañez et al. 2010 for discussion of motivations behind
different graph representations of the same system, in the
context of biochemical networks). We choose here to use a
unipartite representation, where only chemical species are
represented as nodes. This allows us to study interaction
patterns among species within an atmosphere and focus on
identifying its statistical characteristics from a relatively simple
network projection. The abundance, reaction list, and reaction
rate constants used for network construction were all taken
directly from the VULCAN model specifications and model
outputs at steady state.

The result of forward models with vertical mixing is
thermochemical steady-state values for the abundances (in
units of mole fraction or mixing ratio) of all species included in
the model over a range of pressures (see Figure 2, top panel).
While such measurements are familiar to planetary and
atmospheric scientists, some of the multivariate chemical
network measures we track are less familiar, but they also
vary in the model over a range of pressures (see Figure 2,
bottom panel). Such network measures include average
clustering coefficient \( C \) (the average density of edges between
neighbors of any given node) and average neighbor degree
\( \langle K_{nn} \rangle \) (the average degree of the nodes connected to a given

Figure 2. Solid lines represent the atmosphere at \( K_{zz} \) of zero, dashed lines an atmosphere at \( K_{zz} \) of \( 10^{10} \). Top: the abundance of key molecular species as a function of
pressure, at 1200 K and 1 solar metallicity. Bottom: network topological measures—average clustering coefficient, and average neighbor degree (see Table 1 and
Section 2.2)—as a function of pressure, at an irradiation temperature of 1200 K and 1 solar metallicity. Network topological measures vary as physical conditions,
such as pressure, change and capture bulk features of planetary atmospheric chemistry.
Notation Network Measure Description in the Context of Chemical Reaction Networks

\((k)\) Mean degree The average of the total flux driven by a chemical compound

\((k_{\text{nn}})\) Mean of average neighbor degree The average of the total flux driven by neighbors of a chemical compound

\((\ell)\) Average shortest path length The average of the minimal amounts of flux along dependency pathways between every pair of compounds

\((C)\) Average clustering coefficient The average tendency of a chemical compound to have neighbor compounds that participate in the same chemical reaction

\((g(v))\) Average node betweenness centrality The average tendency of a compound to be part of a shortest (flux-weighted reaction) path between every pair of compounds

\((g(e))\) Average edge betweenness centrality The average tendency of a reaction to be part of shortest (reaction flux-weighted) path between every pair of compounds

node; see Table 1 and Section 2.2). As we will show in more detail, these other ways of quantifying atmospheric interactions can offer additional information over the same range of physical values.

Visualization of network projections for hot Jupiters at different temperatures and \(K_{\text{eq}}\) is shown in Figure 3. Some of the edges in the network appear thicker than others because the edges in the network representations are “weighted”—that is, a value is assigned to each connection between two nodes based on the rates of reactions associated with the edges (because more than two species typically participate in a given reaction, a single reaction can correspond to more than one edge in this network projection, where each edge associated with that reaction carries the same weight). The weight of an edge connecting two nodes \(i\) and \(j\), \(w_{ij}\), is determined by the reaction rate, i.e., the reaction rate constant multiplied by the abundance of the reactants involved in the reaction. Hence, the weighted edges connected to nodes indicate the influence of the corresponding chemical species in driving the chemical flux through the network.

For each network generated on our model grid, we calculated average values of several standard network-theoretic measures over each atmospheric network in our grid data set: degree, shortest path length, clustering coefficient, average neighbor degree, node betweenness centrality, and edge betweenness centrality. The average value of each of these measures can quantify global topology emerging from collective interactions among chemical species in each atmosphere. We provide a summary of each measure in what follows. More detailed descriptions of these measures can be found in Albert & Barabási (2002) and Newman (2003, 2010) for general cases, and in Antoniou & Tsompa (2008), Barrat et al. (2004), and Onnela et al. (2005) with a particular focus on weighted networks.

Degree is one of the most frequently used network measures to characterize the structure of a network (Jeong et al. 2000; Smith et al. 2021). For an unweighted network, the degree of node \(i\), \(d_i\), is simply the number of edges connecting the nodes to the rest of the network. Thus, in a chemical network, the degree quantifies the number of chemical species that a given species shares a reaction with. However, for a weighted network like the atmospheric network models we study here, the degree of node \(i\), \(k_i\), is the sum of weights of the edges connected to \(i\):

\[
    k_i = \sum_{j \in V} a_{ij}w_{ij},
\]

where \(a_{ij}\) is an element of an adjacency matrix of a given network and equal to 1 when \(i\) and \(j\) are connected by an edge and 0 otherwise. This is the total sum of reaction rates for all reactions where a given chemical compound participates as a reactant or a product, which is equivalent to the total flux defined as the sum of total in-flux from and out-flux to all its neighbors in the atmospheric network. Note that the total flux is different from the net flux defined as the in-flux subtracted by the out-flux. The average of the total flux over the set of all chemical compounds in the atmosphere can be quantified by the mean degree, \(\langle k \rangle\), defined as the average weight over all edges in a network:

\[
    \langle k \rangle = \frac{1}{N} \sum_{i \in V} k_i, \tag{4}
\]

where \(N\) is the total number of nodes (chemical compounds) in the network.

The shortest path length from \(i\) to \(j\) for a weighted network is defined as

\[
    \ell(i, j) = \min(\gamma(i, j) \in \Gamma(i, j) \left[ \sum_{u, v \in \gamma(i, j)} w_{uv} \right]), \tag{5}
\]

where \(\gamma(i, j)\) is a path from \(i\) to \(j\) and \(\Gamma(i, j)\) is the set of all possible paths from \(i\) to \(j\) (Antoniou & Tsompa 2008). Note that in a bipartite network the path would correspond to the more familiar definition of pathway as a linear sequence of reactions. However, in the unipartite representation we use here, the path should instead be thought of as dependence among chemical species in a series of reactions where species \(i\) and \(j\) are a reagent in the starting reaction and a product in the ending reaction, respectively. Hence, the shortest path length from compound \(i\) to \(j\) in the chemical reaction network represents the smallest amount of the sum of fluxes of chemical compounds along a dependency pathway connecting the two compounds. The average shortest path length over all pairs of nodes in the network can be written as

\[
    \langle \ell \rangle = \frac{1}{N(N - 1)} \sum_{i,j \in V} \ell(i, j), \tag{6}
\]

and this can be considered as the average minimal amount of flux along a dependency pathway between every pair of compounds in the network.

Clustering coefficient indicates the weighted density of edges between neighbors of a given node:

\[
    C_u = \frac{1}{d_u(d_u - 1)} \sum_{uv}(\hat{w}_{uv}\hat{w}_{uw}\hat{w}_{wv})^{1/3}, \tag{7}
\]
Figure 3. (a) The abundance of CH$_4$, CO, NH$_3$, and N$_2$ as a function of the temperature of the atmosphere at constant pressure of 100 mbar, demonstrating the transition from abundant reduced species (CH$_4$ and NH$_3$) to abundant oxidized species (CO and N$_2$) as temperatures increase. Increasing the $K_{zz}$ from equilibrium (left) to $K_{zz} = 10^{10}$ (right) leads to the transition occurring at higher temperatures than the equilibrium case. The dashed lines mark 600, 1200, and 1800 K, which are the temperatures of the corresponding networks shown in panel (b). (b) Network diagrams of the atmospheric chemical reaction network of a hot Jupiter over a range of temperatures, at solar metallicity and $K_{zz} = 0$ (top) and $K_{zz} = 10^{10}$ (bottom), demonstrating how network topological properties measurably change as a function of physical parameters—in this case, irradiation temperature. The networks at irradiation temperatures 600, 1200, and 1800 K are represented by yellow, orange, and red, respectively. The size of the nodes represents the flux-weighted degree of each node, and the width of edges indicate their reaction rate. Both the size of nodes and edges are rescaled consistently and normalized at each temperature and $K_{zz}$. 
where \( d_e \) is the number of edges connected to a node \( u \), and \( w_{uv} = \frac{w_{uv}}{\max(v)} \) is normalized by the maximum weight in the network (Onnela et al. 2005). \( C_u \) is assigned to 0 for \( d_e < 2 \). A group of nodes with high clustering coefficient will be tightly knit. Average clustering coefficient (Barabasi 2016)

\[
(C) = \frac{1}{N} \sum_{i \in V} C_i
\]

is the average of clustering coefficients over all nodes and quantifies the tendency of interdependence between two chemical compounds sharing a neighbor compound.

Average neighbor degree is the average weighted degree of the nodes connected to a given node (Barrat et al. 2004) and can be written as

\[
k_{nn,i} = \frac{1}{k_i} \sum_{j \in N(i)} w_{ij} d_j,
\]

where \( k_i \) is the degree of node \( i \), \( N(i) \) is the set of nodes connected to \( i \), \( w_{ij} \) is the weight of the edge that links nodes \( i \) and \( j \), and \( d_j \) is the number of edges connected to node \( j \) ( \( d_j \) is equivalent to \( k_i \) in unweighted networks). The correlation between average neighbor degree and degree is useful for determining whether a given network is assortative (such that high-degree nodes tend to be connected to high-degree ones and avoiding low-degree ones) or the opposite is true and the network is disassortative (such that high-degree nodes tend to be connected to low-degree nodes and not other high-degree nodes). In atmospheric networks, it can identify statistically whether chemical compounds with a high total flux tend to participate in the same reactions. In this paper, we computed the mean value of average neighbor degree \( \langle k_{nn} \rangle \) over all nodes for each network. Note that, in general, the mean value of the average neighbor degree alone cannot determine assortativity. However, in our atmospheric network models where a few nodes have a degree higher by several orders of magnitude than other nodes and dominate the total flux within the network, the high mean value can indicate that the high-degree nodes tend to be connected to each other, which indicates that the network is assortative.

Node betweenness centrality is defined as

\[
g(v) = \sum_{s,t \in V} \frac{s(s, t|v)}{s(s, t)},
\]

where \( V \) is the set of nodes, \( \sigma(s, t) \) is the number of shortest \( (s, t) \)-paths, and \( \sigma(s, t|v) \) is the number of those paths passing through some node \( v \) other than \( s, t \), if \( s = t \), \( \sigma(s, t) = 1 \), and if \( v \notin \{s, t\}, \sigma(s, t|v) = 0 \). This measure can quantify the influence of a given compound on total flux along shortest paths between any two compounds occurring within the network. Nodes with high betweenness centrality can sometimes be low degree but essential to dynamics and function since they play a key structural role by connecting many otherwise disconnected or distant nodes. Hence, it can be considered as global-scale connectivity in contrast to the degree associated with local connectivity. We calculated the average node betweenness, \( g(v) \), over each network. In atmospheric chemical reaction networks, the average node betweenness measures how often any given species is included in the shortest path between two other species in the network.

Edge betweenness centrality is similarly defined to node betweenness centrality, except that it represents the number of shortest paths that transverse through edge \( e \), instead of a node:

\[
g(e) = \sum_{s,t \in V} \frac{s(s, t|e)}{s(s, t)},
\]

where \( V \) is the set of nodes, \( \sigma(s, t) \) is the number of shortest \( (s, t) \)-paths, and \( \sigma(s, t|e) \) is the number of those paths passing through edge \( e \). We calculated the average edge betweenness, \( \langle g(e) \rangle \), which captures the impact of edges within the network on large-scale connectivity. This indicates how frequently any given reaction is part of the shortest path between any two species in the network.

2.3. Thermodynamic Analysis

A number of different thermodynamically motivated measures of atmospheric disequilibria have been proposed (e.g., Line & Yung 2013, Krissansen-Totton et al. 2016, 2018). Among these, the measure developed by Krissansen-Totton et al. (2016, 2018) has been widely discussed owing to its potential applicability to life detection on terrestrial worlds. Here we are interested in including thermodynamic statistics as another component of information in atmospheric models that might be exploited to provide more accurate inferences of relevant planetary properties, such as the degree of disequilibria (e.g., inferred \( K_{eq} \)). Initially, we measured the thermochemical disequilibrium by calculating the total available Gibbs free energy, or \( \phi \), for each atmospheric model following Krissansen-Totton et al. (2016). To calculate \( \phi \) requires summing the absolute Gibbs free energy of all species at a given temperature \( T \) and pressure \( P \):

\[
G_{(T,P)} = \sum n_i (G^0_{(T,P)} + RT \ln(a_i)) = \sum n_i (G^0_{(T,P)} + RT \ln \left( \frac{P n_i \gamma_i}{n_T} \right)),
\]

where \( G^0_{(T,P)} \) is the standard energy of formation for the ith species at temperature \( T \) and pressure \( P \), \( P \) is standard pressure (1 bar), and \( n_i \) and \( \gamma_i \) are the number of moles and the activity coefficient of species \( i \), respectively.

In practice, Gibbs free energy is almost always calculated in terms of reference to the standard Gibbs free energy of formation at standard temperature (273.15 K) and pressure (1 bar), denoted as \( \Delta G_{(T,P)} \), \( \Delta G_{(T,P)} \) as Krissansen-Totton et al. (2016) demonstrated, yields the same output as using \( G_{(T,P)} \) for the purposes of determining the minimal value of available Gibbs free energy required for calculating \( \phi \), but it is simpler to calculate.

To determine \( \phi \), the total Gibbs free energy in Equation (12) was calculated based on the mole fraction \( n_i \) of each species present at equilibrium (\( K_{eq} = 0 \)) and then compared to the mole fraction calculated from simulated atmospheric data for each model in our grid, producing the calculated value of \( \phi \) for that model atmosphere, which represents the available Gibbs free energy in the atmosphere. A higher value of \( \phi \) indicates that the atmosphere is further away from equilibrium. \( \phi \) is given by the equation

\[
\phi = \Delta G_{(T,P)}(n_i) - \Delta G_{(T,P)}(n_i).
\]
By construction, $\phi$, as defined by Krissansen-Totton et al., requires an equilibrium value $\phi = 0$. However, equilibrium values cannot readily be established for real-world observations, since many, if not all, planetary atmospheres are never observed to exist at thermochemical equilibrium. Thus, it is a theoretical construct not easily relatable to empirical observations. In our case with a model at equilibrium ($K_{zz} = 0$), observational uncertainty introduced by our interpolating function (see next section), meant to mimic observational uncertainty, introduces variability in the calculation of $\phi$ for the equilibrium case. One solution is to impose a somewhat arbitrary designation of the $\phi = 0$ case as a delta function exactly at $K_{zz} = 0$ and the specified temperature and pressure, from which the difference to interpolated values could be calculated. Since the atmospheric model at $K_{zz} = 0$ may or may not reflect the real-world conditions of the atmospheric chemistry in question (or its equilibrium), this assumption introduces the potential for additional model error. This is also unlike how we treated the network parameters or chemical species abundances, which have well-defined values independent of normalizing to an exact equilibrium scenario, and those values are calculated with observational uncertainty. This meant that we could not make direct comparisons between the utility of $\phi$ and other atmospheric variables because the former includes an arbitrary normalization, whereas the other measures do not. We therefore elected to use $G_{T,P,N}$, hereafter simply referred to as “$G$,” which describes the summed Gibbs free energy of all species in the entire atmosphere, as the primary thermodynamic statistic for each atmosphere (instead of using $\phi$ directly), because this did not require us to define a nonobservable equilibrium case for comparison.

2.4. Generating Statistical Distributions via Interpolation of Chemical Species Abundances, Network Topology, and Gibbs Free Energy

To model observational uncertainty, the calculated outputs of the VULCAN simulations (species abundances, network topology, and Gibbs free energy) were used as the basis of multidimensional interpolation functions created using SciPy’s griddata package. These functions attempt to “fit” the estimated position of a data point, in this case, estimated atmospheric model outputs such as species abundances, network measures, and total Gibbs free energy, based on the values of the inputs. Here the inputs are the initial conditions of irradiation temperature and metallicity input into the VULCAN code. That is, when given a set of planetary parameters, for example, an irradiation temperature distribution with a $T_{\text{mean}} = 1200$ K and metallicity distribution of $0.5_{\text{solar}}$ metallicity, the interpolation function will estimate what the most likely values of physical parameters (such as mixing ratios of specific species) would be if we were to simulate an atmosphere with specified conditions. The same process is used for estimating network topology: the output represents what the likely network topology would be if we had simulated an atmosphere with the specified conditions, built a chemical reaction network based on the resulting abundances of atmospheric gases, and then measured the topological properties of that network. The result is a series of distributions of atmospheric topological measures and physical parameters (Figures 4 and A1–A10).

We fed the interpolation functions several 10,000-point normal distributions of initial irradiation temperature, metallicity, and pressure values. The centers of these distributions started at 400 K and then increased at 100 K increments until 2000 K, leading to 17 distributions in total. The resulting propagated probability distributions for atmospheric variables were then statistically analyzed; generally, they were compared on the basis of $K_{zz}$ and temperature (see Figures 5 and 6, and A11–A14). To accommodate the limits of certainty that would be present in actual observational data, uncertainties (in the form of upper and lower bounds on error bars) ranging from temperatures of ±50 K to ±1000 K were incorporated to generate the distributions. The griddata package cannot interpolate values outside the original range given (i.e., here between 400 and 3000 K for the mean temperature of the normal distributions); therefore, many of the conditions drawn from the ±1000 K Gaussian distribution that was fed into the interpolation function had to be discarded. Consequently, the resulting parameter distributions interpolated from the high-uncertainty conditions exhibit a bimodal skew. This is an artifact. However, given that this only occurs in situations that already have high uncertainty, and only toward the boundary of the data, it is not expected that this will significantly impact the conclusions of our analysis.

3. Results

3.1. Distinguishability of Chemical Species Abundances, Network Topology, and Gibbs Free Energy for the Different States of Disequilibria

A key goal of our analyses was to determine whether the likely value for a given variable can be statistically distinguished for different $K_{zz}$. This is a necessary precondition for a variable to be a good predictor of $K_{zz}$, and therefore a strong indicator of disequilibria: the candidate variable must first take on statistically distinguishable values to correlate with different degrees of disequilibrium as driven by the strength of vertical mixing quantified by $K_{zz}$. This implies that the likelihood distributions must themselves be distinguishable for different $K_{zz}$ across varying temperatures and uncertainty. We therefore measured the distance between distributions using the Kolmogorov–Smirnov (K–S) test (Massey 1951). The two-sample K–S test is a nonparametric test useful for distinguishing between two different samples, both with respect to the location and shape of the samples’ cumulative distribution functions. The output of the K–S test is a value between 0 and 1 representing the maximum distance between two cumulative distribution functions, along with its associated $p$ value. The distance is given as

$$D_{n,m} = \sup_{x}|F_{1,n}(x) - F_{2,m}(x)|,$$

(14)

where $F_{1,n}(x)$ and $F_{2,m}(x)$ are the empirical distribution functions of the samples and $\sup_{x}$ is the supremum function. Comparatively large K–S values imply distinguishable distributions, which provide targets for follow-up as possible candidates for accurate inference of $K_{zz}$, since it is these cases that most readily distinguish the $K_{zz}$ values. We considered $p < 0.005$ to be our threshold for significance. For comparisons between two 10,000-point distributions, this significance threshold corresponds to a K–S value of ~0.00257. Our results show that likelihood distributions of network measures, corresponding to different $K_{zz}$, are distinguishable with respect
to physical parameters such as temperature; see Figure 7 and Tables A1–A6). In general, the distributions become more similar to one another at higher temperatures; this result is expected given the previous observations that the effect of vertical mixing on the distance from equilibrium tends to weaken as temperatures increase (Mollière et al. 2015; Thorngren et al. 2019).

The distinguishability of network measurement distributions by the use of the K-S tests (See Figure 7) yields the average neighbor degree with the highest value across all $K_{zz}$ values, followed by average shortest path length and node betweenness centrality. This holds true even as the uncertainty of the values increases up to ±250 K (Figures 5 and 6). The higher K-S scores of these measurements are likely a result of their sensitivity to the reaction rates incorporated into the edge weighting of the network. Other network measurements, such as node betweenness centrality and average neighbor degree, do not track $K_{zz}$ as well (see Tables A1–A6 for full comparison of KS values).

Among the measured molecular abundances, NH$_3$ appeared to be the most distinguishable, having a slightly lower K-S value than the most distinguishable network measurements, followed by CO. The other molecular abundances analyzed are only weakly distinguishable from each other as a function of...
$K_{zz}$ likely due to the fact that the abundance of these species is not significantly altered at different vertical mixing strengths.

The thermodynamic measurements, $\phi$ and $G$, are less promising when evaluated using the K-S metric. The former yields K-S $= 1.0$ (maximally distinguishable) across all $K_{zz}$ values. However, this is an artifact due to the use of a delta function as a reference equilibrium value (one reason we did not in the end use $\phi$ as the primary thermodynamic measure in this work, though see Figures A11 and A12 in the Appendix). The latter actually decreases in distinguishability from the equilibrium case as $K_{zz}$ increases.

Thus, across both $K_{zz}$ values and levels of uncertainty, certain measurements (e.g., average neighbor degree, NH$_3$ abundance) perform much better than the other metrics used. This effect grows stronger as distance from equilibrium is increased.

Figure 5. $T = 1200$ K. Distributions of network measures and thermodynamic parameters, interpolated from a 10,000-point normal distribution of initial conditions centered on a mean temperature of 1200 K, at $K_{zz}$ values of 0 and $10^{10}$, and uncertainties in temperature of 50, 250, and 1000 K. The distributions with different $K_{zz}$ values can be distinguished from each other for the network measures (especially average clustering coefficient and mean degree), even with uncertainties of ±250 K. Total available Gibbs free energy, however, does not appear to be useful for distinguishing between atmospheres at equilibrium and disequilibrium. The bimodal distributions seen in the ±1000 K case are an artifact of adding noise to the data. Appendix For the description of the network measures in the figure, see Table 1 and Section 2.2.
3.2. Evaluating Model Error

We analyzed the effect of perturbations on our analysis through several interventions to the reaction network. Like the uncertainty inserted during interpolation, these tests were conducted to determine the analysis pipeline’s tolerance for errors that might occur either during observation or during inverse modeling. As such, we did not alter the VULCAN models themselves, but rather perturbed the network constructed from the output model.

First, we perturbed the data by removing the CH$_4$–CO quenching pathway (selected owing to the critical role it plays in the atmospheric chemistry of hot Jupiters; e.g., Moses et al. 2011) from the chemical reaction networks constructed. We also removed all species containing oxygen, carbon, or nitrogen from the network (see Figures A15 and A16 in the
Section 2.2.

The K-S statistic of an observable molecule abundance, and network topological measures. Distributions were calculated from a normal distribution of initial conditions, centered at 900 K, 50 comet, and 100 mb. Values were calculated with respect to the distribution at equilibrium ($K_{zz} = 0$). Uncertainty for the distributions was ±50 K. A K-S value of 0 indicates that the two distributions sampled are indistinguishable from each other; the closer the value is to 1, the more distinguishable the distributions are. For comparative values across $K_{zz}$ values, see Table A1. For the description of the network measures in the figure, see Table 1 and Section 2.2.

Figure 7. K-S metric values for the distributions of observable molecule abundance, and network topological measures. Distributions were calculated from a normal distribution of initial conditions, centered at 900 K, 50 comet, and 100 mb. Values were calculated with respect to the distribution at equilibrium ($K_{zz} = 0$). Uncertainty for the distributions was ±50 K. A K-S value of 0 indicates that the two distributions sampled are indistinguishable from each other; the closer the value is to 1, the more distinguishable the distributions are. For comparative values across $K_{zz}$ values, see Table A1. For the description of the network measures in the figure, see Table 1 and Section 2.2.

Appendix), either singularly or in combination with another element being removed. We did this both to see how the global network structure would change and to try to understand the subtler underlying network features that might have been obscured by the extremely high weight edges. We also did perturbation testing by removing key species—CO$_2$, CH$_4$, H$_2$O, and NH$_3$—and their associated reactions.

We found that perturbing the networks changed the distribution of network measurements, even when only the CH$_4$–CO pathway was removed (see Figure 8). This change in distribution was likely due to the fact that the networks are edge-weighted by reaction rates and that this pathway has an extremely high edge weight, especially at high temperatures. Unsurprisingly, the changes were even more pronounced when large swaths of species were removed, such as all carbon-containing compounds (see Figure A15).

3.3. Predictive Efficacy of Multivariate Data for the Disequilibria of Atmospheric Networks

We next sought to evaluate the predictive power of individual variables or sets of the variables of atmospheric chemical reaction networks for accurately determining the degree of disequilibria. The K-S test for a variable can be used to identify those variables that can distinguish the atmospheric networks in equilibrium from those in disequilibrium, but it cannot be used to characterize the different states of disequilibria quantitatively. In other words, K-S values do not provide information about how accurately a given variable measured from an atmospheric network can predict the $K_{zz}$ value associated with the network. To this end, we used XGBoost (Chen & Guestrin 2016; Sharma 2018), a supervised machine-learning algorithm, to evaluate the accuracy of the different variables for predicting $K_{zz}$, by building a classifier that predicts the $K_{zz}$ value of an atmospheric network among $K_{zz} = 0$, $10^6$, $10^8$, and $10^{10}$ from a given variable or a set of them. This approach enables quantifying the capability of the different variables to distinguish atmospheric networks associated with different $K_{zz}$ values. Note that the prediction through machine learning relies solely on data measured from our atmosphere model without assuming a functional form for the classifier. For the actual implementation of XGBoost, we split the data set into training and testing sets with 80:20 ratio randomly using a Python package sklearn. model_selection.train_test_split and train XGBClassifier on the training sets with min_child_weight = 10, max_depth = 5, and n_estimators = 1000.

Figure 9 shows the predicting accuracy of variables as a function of the mean temperature. The Gibbs free energy in the figure, G, shows prediction accuracy around 70% throughout the mean temperature range from 400 to 2000 K regardless of uncertainty in the temperature. On the other hand, the predicting accuracy of the abundance of the four chemical species varies drastically depending on the mean temperature, especially for CO and NH$_3$. The predictive power of CO abundance is maximum, close to 1, when the mean temperature is low, but then it decreases monotonically as the mean temperature increases, until it becomes lower than 0.25 at 1800 K. The distinctive patterns of predictivity of network topological measures in Figure 9 divide the topological variables into two different groups. Mean degree $\langle k \rangle$ and average clustering coefficient $\langle C \rangle$ show uniformly low predicting accuracy ($<0.25$) over the whole temperature range, as might be expected based on their low distinguishability across different disequilibria with the K-S test. The predictive power of the other topological variables fluctuates as the mean temperature increases and their maximum predicting accuracy is larger than or close to Gibbs free energy. The topological variables with higher predictive power than mean degree and average clustering coefficient tend to show higher complexity compared to the topological measures with low predictive power; this might be expected given how computing more complex topology requires more information about the network as a whole than simpler topological measures do. Figures A17 and A18 in the Appendix show how as the uncertainty in the mean temperature increases, the predictive accuracy of any individual variables decreases compared to the results in Figure 9, with the exception of the Gibbs free energy and the abundance of CO. The Gibbs free energy and the abundance of CO show the highest predictive power over the whole range of the temperatures and at low temperature, respectively.

3.4. Boosting Up the Predictive Power and Distinguishability for the Disequilibria of Atmospheric Networks via the Multivariate Data Analysis

Our next question was how we can increase the predictive power of the variables for $K_{zz}$ and hence their distinguishability for the chemical disequilibria of atmospheric networks. To this end, first we identified variables that contribute most to the prediction at different mean temperatures with a given uncertainty. Figure 10 shows individual variables with the three highest values for their predictive power (in red) compared to others (in gray) when the uncertainty in the mean temperature is ±50 K. The members in the group of top predictors change depending on the mean temperature, as can
Figure 8. $T = 1200 \, \text{K}$ w/ perturbations w/ pathway removed, uncertainty $\pm 50 \, \text{K}$. Distributions of network measures when the CH$_4$–CO quenching pathway is removed from the network (dashed line) vs. complete networks (solid line), at 1200 K and 1 solar metallicity: clustering coefficient and average neighbor degree allow distinction between the two. For the description of the network measures in the figure, see Table 1 and Section 2.2.
be inferred from Figure 9. This is shown explicitly in Figure 10, where we utilized the multivariables composed of the top predictors to measure predicting accuracy and the result (yellow). The result demonstrated that the multivariate information boosts up the predicting accuracy and hence distinguishability drastically compared to individual variable analysis. The predicting accuracy of the set of top predictors (denoted by \( T^* \)) is higher than 99% at the mean temperature...
between 500 and 1100 K, more than 95% between 1200 and 1700 K, and around 80% between 1800 and 1900 K. Figure 10 also shows the predicting accuracy measured by using two other different multivariate groups: the set of the abundances of NH₃, CH₄, CO, and H₂O (denoted by A*); and the set of six network topological measures (denoted by N*). The two groups

Figure 10. Boosting up the predicting accuracy for the disequilibrium state of the atmospheric chemical reaction networks with multivariate information. Each panel of the plot shows the predicting accuracy of variables for atmospheric networks associated with κₐ = 0, 10⁶, 10⁸, and 10¹⁰ at a given mean temperature between 500 and 1900 K with uncertainty ±50 K. The three variables with the highest predictive power at each mean temperature compared to other individual variables (in gray) are colored red. (For detailed information about topological variables used in this plot, see Table 1 and Section 2.2). T*, A*, and N* represent multidimensional variable sets composed of the three top predictors, the abundance of the four chemical species shown in this plot, and the six network topological measures represented in Figure 9, respectively. The three different multivariable sets demonstrated higher predicting accuracy than individual variables even in the cases where the highest predictors are not included.
show similar predictive power to the set of top predictors identified at each different temperature even when individual variables in the two former groups do not show high predictive power compared to the top predictors. The same analysis on the data sets with uncertainty in the mean temperature $\pm 250$ K and $\pm 500$ K can be found in the Appendix (Figures A19–A21).

3.5. Robustness of Network Topology as Predictor for the Disequilibria of Atmospheric Networks

To test the robustness of network topological measures or the set of them as a predictive metric, we measured network topology after one of four key species—NH$_3$, CH$_4$, CO, and H$_2$O—was removed from networks and computed their predicting accuracy for $K_{zz}$ based on the perturbed network data sets. Figure 11 shows the predictive power of four individual topological variables—$\langle \ell \rangle$, $\langle k_{\text{max}} \rangle$, $\langle g(v) \rangle$, and $\langle g(e) \rangle$—and the set of topological measures, $N^*$, before (gray) and after (green) removal of each of four chemical species are robust predictors for $K_{zz}$ as both measure groups show high predictive powers after the perturbation. While the predicting accuracy of the four individual topological variables shows considerable changes (decreasing or increasing), the predictive power of $N^*$ as a combined set of topological measures is consistent before and after the perturbation, and this indicates its robustness as a predictor for the chemical disequilibrium against perturbation.

4. Discussion

Current approaches to infer the properties of exoplanets from their atmospheres aim to identify specific molecular species from spectral data because these allow identifying atmospheric composition. However, there is other information about planetary atmospheres, such as the degree of disequilibria, that is not directly observational accessible, does not strictly depend on knowledge of molecular abundance, and is nonetheless a prime target for remote inference to learn more about
exoplanet characteristics. This is particularly important in the context of inferring alien life, where the presence of a biosphere drives specific features of the planetary system that are not necessarily uniquely reflected in the production of atmospheric gases. In the current work, we have expanded the types of informative variables exoplanet researchers can target when building new methods for remote inference. Molecular species abundance data capture features of composition. The network topology measures introduced here, by contrast, capture global features of the patterns of interactions among molecules. Our analyses demonstrate that different combinations of these two levels of model description, combined with Gibbs free energy, tuned to the particular planetary parameters (metallicity, T) can lead to more accurate inference of the disequilibrium state of planetary atmospheres.

Our results indicate that topological measures can be used to distinguish between atmospheres with different $K_{zz}$ values and thus correspondingly different distances from thermochemical equilibrium. While metrics that focused primarily on individual nodes, such as mean degree, performed poorly at distinguishing between different distances from equilibrium, more complex measurements such as average neighbor degree performed very well. We suspect that this is due to the fact that changing $K_{zz}$ primarily affects the reaction rates of only a comparative few species; thus, the change in simpler network measurements was washed out by the rest of the network. Metrics such as average neighbor degree, however, were able to detect the impact of these few changes propagating throughout the network, hence their greater sensitivity. These results corroborate how network representations are particularly well suited to capturing global features of atmospheric reaction networks. Furthermore, the set of topological variables displays high predictive power that is comparable to the set of major chemical species abundances and robust against the perturbation in the data sets such as missing central species.

These findings appear resilient even in the face of considerable observational uncertainty, as the distributions of average network clustering coefficient, for example, remain distinguishable even with uncertainty in temperature of ±250 K. To a lesser extent, we can also distinguish between systems that have been perturbed via the removal of critical pathways.

A key result is how no one measure will single-handedly provide confidence in the inference of disequilibria over a wide range of planetary conditions given the observational and model uncertainty. Instead, our results demonstrate how a multivariate approach tuned to particular planetary properties (T, metallicity) is necessary. Average clustering coefficient, for example, demonstrates a strong ability to determine whether the atmosphere is in a state of equilibrium or not, but it lacks the sensitivity to indicate how far away the system is from equilibrium. Mean degree and average shortest path length, on the other hand, provide much greater sensitivity to the distance from equilibrium, even if they are less sensitive overall compared to average clustering coefficient. Abundances and the Gibbs free energy can provide grosser distinctions but at higher sensitivities.

A caveat of our approach is that the global topological measures of networks are calculated from prior knowledge of the abundances and rates in forward models. One can therefore reasonably raise the question of what advantage an approach based on network topology really brings if it requires information already captured to a lesser degree in other variables, e.g., individual species abundances (which are statistically combined to yield weighted network topology). There are two key reasons that the network-based approach provides utility beyond just knowing abundances. The first is that we have shown that the set of network topological measures is a robust predictor for atmospheric disequilibrium against chemical perturbation. We expect it to be less sensitive to model error and more robust to fluctuations in the data. This is because network projections represent a coarse-graining of the chemistry happening in the data and, like other macrovariables, are not as sensitive to microscopic variation. Thus, even though in a modeling pipeline the network construction requires the information of abundances, the topological measurements themselves, once calculated, provide only coarse-grained information of the chemical systems described and therefore can describe many atmospheres that may vary in their details but still exhibit the same bulk features. They therefore can potentially provide a more robust map between what we can model and what we can infer. Even if network representations of atmospheric chemistry are not ultimately adopted as a tool for remote inference, they do provide utility in understanding disequilibria properties of atmospheres in terms of the collective properties of atmospheric reactions.

The second reason is that our work motivates network topology as a target for new remote inference methods, in particular by motivating development of approaches to directly infer network topology from atmospheric data rather than inferring abundances. For example, rather than using Bayesian retrieval methods to retrieve abundances from spectral data, pipelines could be developed that aim to directly extract network topology. If this can be done, it allows a more direct means of getting at features (e.g., topology) that most accurately capture global features of the state of a planetary atmosphere, including its disequilibria conditions. The latter point is a major part of our motivation for studying atmospheric network properties alongside more traditional variables for inference, such as abundances. This is due to the potential of network-based approaches for future development of statistically motivated methods for the remote inference of life. Biosignatures based on collective interdependent interactions among molecules in the atmospheres of exoplanets have been proposed by a number of researchers as a promising candidate for future life detection efforts (Cenler & Dittrich 2007; Estrada 2012; Schwieterman et al. 2018; Walker et al. 2020). The reasoning is multifold but relies on the idea that life is a systems-level property and that when detecting life we need agnostic, quantitative approaches that aim to measure the presence of life at a systems level. To accomplish this goal, two near-term tasks are required: (1) network topology must be validated as a statistically significant distinguisher of a biosphere on a planet, and (2) it must be demonstrated that network properties are remotely inferable and provide reliable information about the physics and chemistry of planetary atmospheres. Efforts to address task 1 are underway (Kim et al. 2019; Solé and Munteanu 2004), and there is ample evidence that biologically driven networks display distinctive topological properties. Here we have attempted to take the first steps toward task 2 that are rigorously grounded in the realities of remote detection. Our goal was to demonstrate that useful information about planetary atmospheres is contained in their reaction network topology, and indeed we find that this is the
case. Taken in a multivariate analysis, network topology provides additional statistical power for inferring the state of atmospheric disequilibria. In doing so, we have demonstrated that network topology can be a useful tool for exoplanet characterization, even outside of life detection. However, much remains to be done to develop these methods further and to make them a powerful statistical tool for the exoplanet science community.

We gratefully acknowledge support from the National Aeronautics and Space Administration via grants NNX15AD53G and GR40991.

Appendix
Methods

A.1. Atmospheric Modeling with VULCAN

The network analysis was a multistep process, using several different techniques. The first step of the process was to use the VULCAN modeling package (Tsai et al. 2017) to simulate the atmospheric chemistry of a hot Jupiter over a range of temperatures (400–3000 K, in ~100 K increments) and, later, metallicities (0.1–300 times that of the solar system, evaluated on a log scale) and pressures (from 50 to 150 mb, reflecting the likely depth in the atmosphere that will be observable), and, finally, vertical mixing coefficients, ranging from zero to $10^{10}$ s$^{-1}$ cm$^{-1}$.

The vertical mixing coefficient, or $K_{zz}$, represents how much mixing is occurring between the lower layers of the atmosphere, which are at higher temperature and pressure, and the upper layers, which are at lower temperature and pressure. At a $K_{zz}$ value of zero, there is no mixing between layers, and each layer is assumed to very quickly reach chemical equilibrium; thus, we consider this to be our equilibrium case. As the atmosphere becomes more mixed, it moves away from equilibrium, since molecules can be produced at the lower levels of the atmosphere in quantities that are not thermodynamically favorable at the higher levels and then “dredged” up to the higher levels at a rate faster than they can be depleted by equilibrium chemistry. Thus, $K_{zz}$ is used as a proxy for how far away the atmospheric chemistry is from equilibrium.

A.2. Network Analysis

Once the models converged to a solution, the results were saved as a .vul file, a format specific to VULCAN that can be treated as a Python dictionary once unpacked with the pickle function. Next, the chemical reaction network was created, based on the hot Jupiter reaction list included with VULCAN. We constructed a pipeline for importing the abundance of each constituent component, the reactions they participated in as reactants, and the associated rate constant of each reactant from model outputs, and we calculated edge weights from these values.

Because the network is directed, forward and backward rates were required. The forward rate of a given reaction was calculated by multiplying the product of the abundance of the reactants by the forward reaction rate constant; conversely, the backward reaction rate was the multiplicative product of the products of the same reaction and the backward reaction rate constant. In the rare case where the reaction was second-order and not first-order with respect to a given chemical species, the corresponding abundance was squared.

The topological properties of these reaction-rate-weighted chemical reaction networks were then measured using the Python NetworkX package. Specifically, after ensuring that the network graphs were made up of the largest connected components, we used NetworkX’s degree, clustering, betweenness_centrality, average_shortest_path_length, and average_degree methods.

The network measurements and abundances, along with initial conditions (in terms of temperature, pressure, and metallicity), were used as the basis of multidimensional interpolation functions created using SciPy’s griddata package. These functions can then be used to estimate a value for the network parameters based on an initial temperature and metallicity.

A.3. Thermodynamic Analysis

To compare our findings to existing thermochemical disequilibrium metrics, we also calculated the total available Gibbs free energy, or $\phi$, for each atmosphere modeled (Krissansen-Totton et al. 2016). We used a version of the code provided by Krissansen-Totton et al. (2016) implemented in Python for consistency.

A.4. Comparison to Phi

The distribution of phi values also allowed distinguishing between different values of $K_{zz}$, though this was difficult to quantify owing to the inability to use the K-S test or Wasserstein distance metrics on the distributions, as, unlike network metrics, they lack an equilibrium value to be compared against.

A.5. Network Metrics

The values of mean degree, average shortest path length, average clustering coefficient, node betweenness centrality, edge betweenness centrality, and average neighbor degree were measured as a function of temperature, host start metallicity, and $K_{zz}$ (Figures A1–A10). These metrics were then fed into the interpolation function to generate distributions of values.
Figure A1. The abundance of CO as a function of temperature and $K_{zz}$ at 0.1, 1, 10, and 100 times solar metallicity. The x-axis is temperature, y-axis is $K_{zz}$, and color corresponds to the abundance of CO (see color bar for each individual plot for scale).
Figure A2. The abundance of CH$_4$ as a function of temperature and $K_{zz}$ at 0.1, 1, 10, and 100 times solar metallicity. The x-axis is temperature, y-axis is $K_{zz}$, and color corresponds to the abundance of CH$_4$ (see color bar for each individual plot for scale).
Figure A3. The abundance of NH$_3$ as a function of temperature and $K_{zz}$, at 0.1, 1, 10, and 100 times solar metallicity. The $x$-axis is temperature, $y$-axis is $K_{zz}$, and color corresponds to the abundance of NH$_3$ (see color bar for each individual plot for scale).
Figure A4. The abundance of H$_2$O as a function of temperature and $K_{zz}$, at 0.1, 1, 10, and 100 times solar metallicity. The x-axis is temperature, y-axis is $K_{zz}$, and color corresponds to the abundance of H$_2$O (see color bar for each individual plot for scale).
Figure A5. Mean degree as a function of temperature and $K_{zz}$ at 0.1, 1, 10, and 100 times solar metallicity. The x-axis is temperature, y-axis is $K_{zz}$, and color corresponds to the weighted mean degree (see color bar for each individual plot for scale).
Figure A6. Average shortest path length as a function of temperature and $K_{zz}$ at 0.1, 1, 10, and 100 times solar metallicity. The x-axis is temperature, y-axis is $K_{zz}$, and color corresponds to the weighted average shortest path length (see color bar for each individual plot for scale).
Figure A7. Average clustering coefficient as a function of temperature and $K_{zz}$, at 0.1, 1, 10, and 100 times solar metallicity. The $x$-axis is temperature, $y$-axis is $K_{zz}$, and color corresponds to the weighted average clustering coefficient (see color bar for each individual plot for scale).
Figure A8. Node betweenness centrality as a function of temperature and $K_{zz}$ at 0.1, 1, 10, and 100 times solar metallicity. The x-axis is temperature, y-axis is $K_{zz}$, and color corresponds to the weighted node betweenness centrality (see color bar for each individual plot for scale).
Figure A9. Edge betweenness centrality as a function of temperature and $K_{zz}$ at 0.1, 1, 10, and 100 times solar metallicity. The x-axis is temperature, y-axis is $K_{zz}$, and color corresponds to the weighted edge betweenness centrality (see color bar for each individual plot for scale).
Figure A10. Average neighbor degree as a function of temperature and $K_{zz}$ at 0.1, 1, 10, and 100 times solar metallicity. The $x$-axis is temperature, $y$-axis is $K_{zz}$, and color corresponds to the weighted average neighbor degree (see color bar for each individual plot for scale).
A.6. Metric Distributions

The distributions of metrics were compared as a function of $K_{zz}$ (Figures A11 and A12), as were the distributions of phi values (Figures A13 and A14). Distributions were generally more distinguishable from each other as thermochemical disequilibria (in the form of $K_{zz}$) increased; this finding was quantified by the use of the KS statistic (Tables A1–A6). Additionally, distributions where measured where carbon-bearing compounds had been removed from the network, as form of perturbation testing (Figures A15 and A16).

![Figure A11. Phi distributions over a range of uncertainties and $K_{zz}$ values, drawn from a 10,000-point normal distribution of initial conditions centered at 1200 K.](image)

![Figure A12. Phi distributions over a range of uncertainties and $K_{zz}$ values, drawn from a 10,000-point normal distribution of initial conditions centered at 2000 K.](image)
Figure A13. Distributions of network and thermodynamic measurements, interpolated from a 10,000-point normal distribution of initial conditions centered on a temperature of 1200 K, at $K_{zz}$ values of 0, $10^6$, and $10^{10}$, and uncertainties in temperature of 50, 250, and 1000 K. The bimodal distributions seen in the $+/-1000$ K distributions are an artifact of discarding members of the distribution that fell outside the bounds of the original model grid and thus could not be interpolated.
Figure A14. Distributions of molecular abundances, interpolated from a 10,000-point normal distribution of initial conditions centered on a temperature of 1200 K, at $K_{zz}$ values of 0, $10^6$, $10^8$, and $10^{10}$, and uncertainties in temperature of 50, 250, and 1000 K. The bimodal distributions seen in the $+/−1000$ K distributions are an artifact of inserting noise into the system.
Figure A15. $T = 1200$ K w/ perturbations w/ all C-bearing compounds removed. Distributions of network parameters when species containing carbon are removed from the network (dashed line) vs. complete networks (solid line), at 1200 K and 1 solar metallicity. Interestingly, while the absolute values of the parameters change, the overall shape of them remains the same, despite the significant reduction in the network.
Figure A16. Distributions of network parameters when species containing carbon are removed from the network, at 1200 K and 1 solar metallicity. While the absolute values of the parameters changed compared to the complete networks, the distributions are still distinguishable from each other as a function of $K_{zz}$.

Table A1

| W.r.t. $K_{zz}$ | $K_{zz} = 0$ | $K_{zz} = 10^6$ | $K_{zz} = 10^8$ | $K_{zz} = 10^{10}$ |
|----------------|--------------|-----------------|-----------------|-------------------|
| $K_{zz} = 0$   | 0            | 0               | 0               | 0                 |
| $K_{zz} = 10^6$| 0.0025, $p = 1.0$ | 0               | 0.0029, $p = 1.0$ | 0                 |
| $K_{zz} = 10^8$| 0.0011, $p = 1.0$ | 0.0029, $p = 1.0$ | 0               | 0.0074, $p = 0.7$ |
| $K_{zz} = 10^{10}$ | 0.0091, $p = 0.5$ | 0.0088, $p = 0.5$ | 0.0074, $p = 0.7$ | 0                 |

Note. Unsurprisingly, the equilibrium ($K_{zz} = 0$) case is the most readily distinguishable from the other cases, but the varying levels of disequilibrium ($K_{zz} > 0$) are still distinguishable from each other.
### Table A2
K-S Values for Mean Degree across All Interpolated Distributions, with Respect to Different $K_{zz}$ Values

| W.r.t. $K_{zz}$ = | $K_{zz} = 0$ | $K_{zz} = 10^6$ | $K_{zz} = 10^8$ | $K_{zz} = 10^{10}$ |
|-----------------|-------------|----------------|----------------|-----------------|
| $K_{zz} = 0$    | 0           |                |                |                 |
| $K_{zz} = 10^6$ | 0.0007, $p = 1.0$ | 0       |                |                 |
| $K_{zz} = 10^8$ | 0.0011, $p = 1.0$ | 0.0009, $p = 1.0$ | 0       |                 |
| $K_{zz} = 10^{10}$ | 0.0009, $p = 1.0$ | 0.0008, $p = 1.0$ | 0.0013, $p = 1.0$ | 0   |

**Note.** Distributions become more distinguishable from each other as the difference in $K_{zz}$ value between them increases.

### Table A3
K-S Values for Average Shortest Path Length across All Interpolated Distributions, with Respect to Different $K_{zz}$ Values

| W.r.t. $K_{zz}$ = | $K_{zz} = 0$ | $K_{zz} = 10^6$ | $K_{zz} = 10^8$ | $K_{zz} = 10^{10}$ |
|-----------------|-------------|----------------|----------------|-----------------|
| $K_{zz} = 0$    | 0           |                |                |                 |
| $K_{zz} = 10^6$ | 0.1458, $p = 0.0$ | 0       |                |                 |
| $K_{zz} = 10^8$ | 0.1740, $p = 0.0$ | 0.0712, $p = 0.0$ | 0       |                 |
| $K_{zz} = 10^{10}$ | 0.2085, $p = 0.0$ | 0.1559, $p = 0.0$ | 0.1459, $p = 0.0$ | 0   |

**Note.** Distributions become more distinguishable from each other as the difference in $K_{zz}$ value between them increases.

### Table A4
K-S Values for Node Betweenness Centrality across All Interpolated Distributions, with Respect to Different $K_{zz}$ Values

| W.r.t. $K_{zz}$ = | $K_{zz} = 0$ | $K_{zz} = 10^6$ | $K_{zz} = 10^8$ | $K_{zz} = 10^{10}$ |
|-----------------|-------------|----------------|----------------|-----------------|
| $K_{zz} = 0$    | 0           |                |                |                 |
| $K_{zz} = 10^6$ | 0.1255, $p = 0.0$ | 0       |                |                 |
| $K_{zz} = 10^8$ | 0.1601, $p = 0.0$ | 0.1721, $p = 0.0$ | 0       |                 |
| $K_{zz} = 10^{10}$ | 0.0876, $p = 0.0$ | 0.2001, $p = 0.0$ | 0.1423, $p = 0.0$ | 0   |

### Table A5
K-S Values for Average Edge Betweenness Centrality across All Interpolated Distributions, with Respect to Different $K_{zz}$ Values

| W.r.t. $K_{zz}$ = | $K_{zz} = 0$ | $K_{zz} = 10^6$ | $K_{zz} = 10^8$ | $K_{zz} = 10^{10}$ |
|-----------------|-------------|----------------|----------------|-----------------|
| $K_{zz} = 0$    | 0           |                |                |                 |
| $K_{zz} = 10^6$ | 0.0948, $p = 0.0$ | 0       |                |                 |
| $K_{zz} = 10^8$ | 0.0976, $p = 0.0$ | 0.1286, $p = 0.0$ | 0       |                 |
| $K_{zz} = 10^{10}$ | 0.1430, $p = 0.0$ | 0.1953, $p = 0.0$ | 0.1641, $p = 0.0$ | 0   |

### Table A6
K-S Values for Average Neighbor Degree across All Interpolated Distributions, with Respect to Different $K_{zz}$ Values

| W.r.t. $K_{zz}$ = | $K_{zz} = 0$ | $K_{zz} = 10^6$ | $K_{zz} = 10^8$ | $K_{zz} = 10^{10}$ |
|-----------------|-------------|----------------|----------------|-----------------|
| $K_{zz} = 0$    | 0           |                |                |                 |
| $K_{zz} = 10^6$ | 0.3059, $p = 0.0$ | 0       |                |                 |
| $K_{zz} = 10^8$ | 0.3949, $p = 0.0$ | 0.1981, $p = 0.0$ | 0       |                 |
| $K_{zz} = 10^{10}$ | 0.5138, $p = 0.0$ | 0.2890, $p = 0.0$ | 0.1256, $p = 0.0$ | 0   |
Finally, the effect of including network measurements in machine learning was tested in terms of predictive power over a range of temperatures (Figures A17–A21). Metrics were both tested in aggregate, and individually.

Figure A17. Predictive power of variables for the disequilibrium state of the atmospheric chemical reaction networks. Each panel of the plot shows the predicting accuracy of an individual variable for atmospheric networks associated with $K_a = 0, 10^6, 10^8$, and $10^{10}$ as a function of the mean temperature with uncertainty $\pm 250$ K (where gray lines show the data on all plots for ease of comparison). The variables include $G$ (Gibbs free energy), the abundances of chemical species NH$_3$, CH$_4$, CO, and H$_2$O; and the average topological properties (see Table 1 and Section 2.2).
Figure A18. Predictive power of variables for the disequilibrium state of the atmospheric chemical reaction networks. Each panel of the plot shows the predicting accuracy of an individual variable for atmospheric networks associated with $K_0 = 0, 10^6, 10^8, \text{ and } 10^{10}$ as a function of the mean temperature with uncertainty ±500 K (where gray lines show the data on all plots for ease of comparison). The variables include $G$ (Gibbs free energy), the abundances of chemical species NH$_3$, CH$_4$, CO, and H$_2$O; and the average topological properties (see Table 1 and Section 2.2).
Figure A19. Boosting up the predicting accuracy for the disequilibrium state of the atmospheric chemical reaction networks with multivariate information. Each panel of the plot shows the predicting accuracy of variables for atmospheric networks associated with $K_z = 0, 10^6, 10^8,$ and $10^{10}$ at a given mean temperature between 500 and 1900 K with uncertainty $\pm 50$ K. The three variables with the highest predictive power at each mean temperature compared to other individual variables (in gray) are colored red (for detailed information about topological variables used in this plot, see Table 1 and Section 2.2). $T^*, A^*,$ and $N^*$ represent multidimensional variable sets composed of the three top predictors, the abundance of the four chemical species shown in this plot, and the six network topological measures represented in Figure 9, respectively. The three different multivariable sets demonstrated higher predicting accuracy than individual variables even in the cases where the highest predictors are not included.
Figure A20. Boosting up the predicting accuracy for the disequilibrium state of the atmospheric chemical reaction networks with multivariate information. Each panel of the plot shows the predicting accuracy of variables for atmospheric networks associated with $K_{zz} = 0, 10^6, 10^8,$ and $10^{10}$ at a given mean temperature between 500 and 1900 K with uncertainty $\pm 250$ K. The three variables with the highest predictive power at each mean temperature compared to other individual variables (in gray) are colored red (for detailed information about topological variables used in this plot, see Table 1 and Section 2.2). $T^*$, $A^*$, and $N^*$ represent multidimensional variable sets composed of the three top predictors, the abundance of the four chemical species shown in this plot, and the six network topological measures represented in Figure 9, respectively. The three different multivariable sets demonstrated higher predicting accuracy than individual variables even in the cases where the highest predictors are not included.
Figure A21. Boosting up the predicting accuracy for the disequilibrium state of the atmospheric chemical reaction networks with multivariate information. Each panel of the plot shows the predicting accuracy of variables for atmospheric networks associated with $K_{zz} = 0, 10^6, 10^8,$ and $10^{10}$ at a given mean temperature between 500 and 1900 K with uncertainty ±500 K. The three variables with the highest predictive power at each mean temperature compared to other individual variables (in gray) are colored red (for detailed information about topological variables used in this plot, see Table 1 and Section 2.2). $T^*$, $A^*$, and $N^*$ represent multidimensional variable sets composed of the three top predictors, the abundance of the four chemical species shown in this plot, and the six network topological measures represented in Figure 9, respectively. The three different multivariable sets demonstrated higher predicting accuracy than individual variables even in the cases where the highest predictors are not included.
References

Albert, R., & Barabási, A.-L. 2002, RvMP, 74, 47
Antoniou, I. E., & Tsompa, E. T. 2008, Discrete Dyn. Nat. Soc., 2008, ID 375452
Barabási, Albert-László 2016, Network Science (Cambridge: Cambridge Univ. Press)
Barrat, A., Barthelemy, M., Pastor-Satorras, R., & Vespignani, A. 2004, PNAS, 101, 3747
Benneke, B., & Seager, S. 2012, ApJ, 753, 100
Catling, D. C., Krissansen-Totton, J., Kiang, N. Y., et al. 2018, AsBio, 18, 709
Centler, F., & Dittrich, P. 2007, P&SS, 55, 413
Chen, T., & Guestrin, C. 2016, in Proc. 22nd ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (New York: ACM), 785
Cobb, A. D., Himes, M. D., Soboczenski, F., et al. 2019, AJ, 158, 33
Crossfield, I. J. M. 2015, PASP, 127, 941
Estrada, E. 2012, JMaCh, 50, 1363
Fortney, J. J., Dawson, R. I., & Komacek, T. D. 2021, JGRE, 126, e06629
Fuji, Y., Angerhausen, D., Deitrick, R., et al. 2018, AsBio, 18, 739
Gleiss, P. M., Stadler, P. F., Wagner, A., & Fell, D. A. 2001, Adv Complex Syst, 04, 207
Hayes, J. J. C., Kerins, E., Awpian, S., et al. 2020, MNRAS, 494, 4492
Heng, K., & Showman, A. 2015, AREPS, 43, 509
Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, A.-L. 2000, Natur, 407, 651
Jolley, C., & Douglas, T. 2012, AsBio, 12, 29
Jolley, C. C., & Douglas, T. 2010, ApJ, 722, 1921
Kiang, N. Y., Domagal-Goldman, S., Parenteau, M. N., et al. 2018, AsBio, 18, 619
Kim, H., Smith, H. B., Mathis, C., Raymond, J., & Walker, S. I. 2019, SciA, 5, eaau0149
Krippasen-Totton, J., Bergsman, D. S., & Catling, D. C. 2016, AsBio, 16, 39
Krippasen-Totton, J., Olson, S., & Catling, D. C. 2018, SciA, 4, eaao5747
Line, M. R., Teske, J., Burningham, B., Fortney, J. J., & Marley, M. S. 2015, ApJ, 807, 183
Line, M. R., & Yung, Y. L. 2013, ApJ, 779, 3
Madhusudhan, N. 2019, ARA&A, 57, 617
Madhusudhan, N., Harrington, J., Stevenson, K. B., et al. 2011a, Natur, 469, 64
Massey, F. J. 1951, JASA, 46, 68
Meadows, V. S. 2017, AsBio, 17, 1022
Molaverdikhani, K., Henning, T., & Mollière, P. 2019, ApJ, 883, 194
Mollière, P., van Boekel, R., Dullemond, C., Henning, T., & Mordasini, C. 2015, ApJ, 813, 47
Montañez, R., Medina, M. A., Solé, R. V., & Rodríguez-Casio, C. 2010, Bioessays, 32, 246
Moses, J. I., Visscher, C., Fortney, J. J., et al. 2011, ApJ, 737, 15
Newman, M. 2010, Networks: An Introduction (Oxford: Oxford Univ. Press)
Newman, M., Barabási, A.-L., & Watts, D. J. 2011, The Structure and Dynamics of Networks (Princeton, NJ: Princeton Univ. Press)
Newman, M. E. J. 2003, SIAM Rev, 167, 45
Onnela, J.-P., Saramäki, J., Kertész, J., & Kaski, K. 2005, PhRvE, 71, 065103
Schwieterman, E. W., Kiang, N. Y., Parenteau, M. N., et al. 2018, AsBio, 18, 663
Seager, S. 2013, Sci, 340, 577
Seager, S., & Deming, D. 2010, ARA&A, 48, 631
Sharma, N. 2018, XGBoost. The Extreme Gradient Boosting for Mining Applications (Munich: GRIN Verlag)
Simoncini, E., Virgo, N., & Kleidon, A. 2013, ESD, 4, 317
Smith, H. B., Kim, H., & Walker, S. I. 2021, NatSR, 11, 6542
Solé, R. V., & Munteanu, A. 2004, EL, 68, 170
Thomgren, D., Gao, P., & Fortney, J. J. 2019, ApJL, 884, L6
Tsai, S.-M., Lyons, J. R., Grosheintz, L., et al. 2017, ApJS, 228, 20
Venot, O., Hebrard, E., Agúndez, M., et al. 2012, A&AS, 546, A43
Walker, S. I., Bains, W., Cronin, L., et al. 2018, AsBio, 18, 779
Walker, S. I., Cronin, L., Drew, A., et al. 2020, Planetary Astrobiology (Tuscon, AZ: Univ. of Arizona Press), 477
Zahnle, K., Marley, M. S., Freedman, R. S., Lodders, K., & Fortney, J. J. 2009, ApJ, 701, L20