Atmospheric Chemical Reaction Networks as Tools for Understanding Planetary Processes and the Influence of Biospheres on Their Host Worlds

by

Theresa Mason Hirst Fisher

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved May 2023 by the
Graduate Supervisory Committee:

Sara Walker, Chair
Hilairy Hartnett
Michael Line
Evgenya Shkolnik
Jordan Okie

ARIZONA STATE UNIVERSITY

December 2023

ABSTRACT

With the ability to observe the atmospheres of terrestrial exoplanets via transit spectroscopy on the near-term horizon, the possibility of atmospheric biosignatures has received considerable attention in astrobiology. While traditionally exoplanet scientists looking for life focused on biologically relevant trace gases such as O_2 and CH_4, this approach has raised the spectre of false positives. Therefore, to address these shortcomings, a new set of methods is required to provide higher confidence in life detection.

One possible approach is measuring the topology of atmospheric chemical reaction networks (CRNs). To investigate and assess this approach, the ability of network-theoretic metrics to distinguish the distance from thermochemical equilibrium in the atmosphere of hot jupiters was tested. After modeling the atmospheres of hot jupiters over a range of initial conditions using the VULCAN modeling package, atmospheric CRNs were constructed from the converged models and their topology measured using the Python package NetworkX. I found that network metrics were able to predict the distance from thermochemical equilibrium better than atmospheric species abundances alone.

Building on this success, I modeled 30,000 terrestrial worlds. These models divided into two categories: Anoxic Archean Earth-like planets that varied in terms of CH_4 surface flux (modeled as either biotic or abiotic in origin), and modern Earth-like planets with and without a surface flux of CCl_2F_2 (to represent the presence of industrial civilizations). I constructed atmospheric CRNs from the converged models, and analyzed their topology. I found that network metrics could distinguish between atmospheres with and without the presence of life or technology. In particular, mean degree and average shortest path length consistently performed better at distinguishing between abiotic and biotic Archean-like atmospheres than CH_4 abundance.

DEDICATION

Dedicated with love to my wife, Alex, whose support and encouragement has been invaluable. Together, the stars will be ours

ACKNOWLEDGMENTS

As a student of ASU, I acknowledge that the Tempe campus sits on the ancestral homelands of those American Indian tribes that have inhabited this place for centuries, including the Akimel O'odham (Pima) and Pee Posh (Maricopa) peoples.

I'm grateful to Dr. Shang-Min Tsai for his assistance with VULCAN; Dr. Chester "Sonny" Harmon, Dr. Giada Arney, and Dr. Shawn Domagal-Goldman for their advice on running atmos; Dr. Sarah Hörst for her suggestions on perturbing the Archean Earth models; Dr. Hyunju Kim for her guidance and work on machine learning; and the entirety of the Emergence@ASU lab for their insights and advice.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Figure Page

Chapter 1

INTRODUCTION: A NEED FOR A NEW APPROACH TO EXOPLANET BIOSIGNATURES

Now, my own suspicion is that the universe is not only queerer than we suppose, but queerer than we can *suppose...I suspect that there are more things in heaven and earth than are dreamed of, or can be dreamed of, in any philosophy.*

J.B.S. Haldane (1927)

1.1 Spectral Analysis of Exoplanet Atmospheres and Life Detection

The discovery of the first exoplanet around a sun-like star twenty five years ago (Mayor and Queloz, 1995) was followed by the spectral analysis of exoplanets and their atmospheres less than five years later (Coustenis *et al.*, 1998; Moutou *et al.*, 2001; Charbonneau *et al.*, 2002). While such analysis has been possible with Jovian-mass planets that orbit close to their stars, studying the atmosphere of Earth-like exoplanets outside a few preliminary observations (Swain *et al.*, 2021) has not yet been technically feasible (Madhusudhan, 2019). Fortunately, analyses of terrestrial exoplanets are likely to become easier in the near-term with the development of high-resolution observatories. For example, both the National Air and Space Administration and the European Space Agency have proposed or planned space observatories specifically oriented towards measuring the atmospheric spectra of exoplanets (*e.g.*,PLATO and the planned JWST successor) (National Academy of Sciences, 2021). Additionally, ground-based telescopes like the Extremely Large Telescope may have the capability to image exoplanets directly (Fujii *et al.*, 2018). Even more encouraging, the James

Webb Space Telescope has already demonstrated the ability to observe atmospheric spectra of hot jupiters, and terrestrial exoplanets are within its current capabilities, albeit marginally (Alderson *et al.*, 2023; Ahrer *et al.*, 2023).

Investigations of planetary atmospheres have been a topic of interest to the astrobiology community since its earliest days as a field in the 1960's (Simoncini *et al.*, 2013; Grenfell, 2017). Historically, this was conceptualized as a search for a "smoking gun"–single species that were thought to be unambiguous indicators of life. Oxygen in particular–both in the forms of O_2 and O_3–was the focus of considerable speculation, in part due to its relatively detectable spectral signature in the infrared and near-infrared bands (Fujii *et al.*, 2018; Tremblay *et al.*, 2019). O_2 in the presence of CH_4 was considered especially of interest, as the latter will be oxidized to CO_2, and thus the CH_4 must be actively being replenished in order to maintain such a conspicuous disequilibrium (Sagan *et al.*, 1993; Kaltenegger *et al.*, 2020).

However, as knowledge of exoplanets increased and the true diversity of worlds became apparent, these "smoking guns" came under scrutiny. O_2, it appears, is not without the potential risks for false positives (Domagal-Goldman *et al.*, 2014; Meadows *et al.*, 2018; Harman and Domagal-Goldman, 2018) and false negatives (Wordsworth and Pierrehumbert, 2014). For example, detectable levels of O_2 could be purely generated by photolysis on planets with high water vapor content (Kleinböhl *et al.*, 2018), which could otherwise be misinterpreted as evidence of photosynthesis. However, neither does the absence of O_2 indicate an abiotic world–after all, Earth's biological O_2 was virtually undetectable for much of its history, despite a thriving anaerobic biosphere (Reinhard *et al.*, 2017). Additionally, weathering of surface crust and aqueous reactions in the ocean, if aggressive enough, could scrub the atmosphere of biotically produced O_2, providing another potential false negative (Meadows, 2017).

CH_4 has also been subject to similar concerns (Guzmán-Marmolejo *et al.*, 2013).

Even CH_4 in the presence of oxygen may not be a reliable biosignature, due to the possibility of the atmosphere of an exoplanet's unseen moon contaminating the exoplanet's observed atmospheric spectra, leading to the appearance of a single body with a strong atmospheric disequilibrium (Rein *et al.*, 2014; Saxena, 2022). As such, the "smoking gun" approach now appears potentially fraught with ambiguity (Smith and Mathis, 2022).

In addition, exoplanet biospheres may be based on alternative biochemistries, where their biologically relevant molecules may be difficult to identify (Schwieterman *et al.*, 2018). For example, PH_3 has been proposed as a potential biosignature for anoxic worlds (Sousa-Silva *et al.*, 2020) as has CH_3Cl (Gebauer *et al.*, 2021), NH_3 and CH_3Cl for hydrogen-dominated atmospheres (Seager *et al.*, 2013b; Bains *et al.*, 2014; Wunderlich *et al.*, 2021), and halocarbons in biospheres using chlorinic photosynthesis (Haas, 2010). Even Earth's own biosignature gases would have been considerably different over the bulk of its history, due to the absence of oxygen (Arney *et al.*, 2016). The most noticeable sign of life would've instead likely been the presence of high levels of CO_2 and CH_4, with the absence of CO ruling out volcanism (Krissansen-Totton *et al.*, 2019; Wogan *et al.*, 2020). This too, however, is vulnerable to false positives from abiotic processes on warm, rocky planets experiencing out-gassing (Woitke *et al.*, 2020).

The difficulty posed by high-confidence observational evidence of exoplanet biospheres is further amplified by the uncertainty inherent in the spectroscopic measurements. The resolution required for atmospheric transmission spectra of terrestrial planets will be close to the noise floor of even large telescopes such as JWST and PLATO (Lustig-Yaeger *et al.*, 2019; Suissa *et al.*, 2020). This lack of resolution is especially true in the case of traditionally proposed biosignature gases, such as O_3, CH_4, and N_2O (Tremblay *et al.*, 2019). Arguably, these concerns have already been

revealed by the challenges regarding the detection of PH_3 in the atmosphere of Venus: an initial claim of detection (Greaves *et al.*, 2021) led to intense debate about whether the presence of the trace gas could be explained via abiotic processes (Truong and Lunine, 2021) or not (Bains *et al.*, 2021). Further observations have since cast doubt on the presence of PH_3 in Venus' atmosphere entirely (Akins *et al.*, 2021; Villanueva *et al.*, 2021). It can only be imagined how much greater the ambiguity will be in detecting biologically-relevant molecules in the atmosphere of exoplanets, given that they are orders of magnitude further away than Venus (our proverbial next door neighbor)!

Atmospheric technosignatures, such as the emission of chlorofluorocarbons (CFCs) (Haqq-Misra *et al.*, 2022b) or changes in the abundance of nitrogen-bearing species in the atmosphere related to industrial nitrate production (Kopparapu *et al.*, 2021; Haqq-Misra *et al.*, 2022a), while distinctive, are also not immune to the problem of false positives. Halocarbons like CFCs, for example, can be produced by volcanic activity rather than technology (Visscher *et al.*, 2004; Broadley *et al.*, 2018; Klobas and Wilmouth, 2019).

Therefore, a more holistic approach towards atmospheric exoplanet biosignatures and technosignatures may be necessary–ideally, one that is:

1. *robust against false positives and false negatives.* As ground-truthing exoplanets remains impossible for the foreseeable future, any exoplanet biosignature will have to provide a high degree of confidence in detection.

2. *unambiguous in the face of high uncertainty.* Due to observational constraints and the limited amount of data available to exoplanet researchers, an exoplanet biosignature should be as unambiguous as possible.

3. *agnostic across a wide range of possible biochemistries.* We do not yet know

what kinds of life might have evolved on exoplanets, and what we find may be very unfamiliar; a good exoplanet biosignature should account for this.

Such an approach will almost certainly require a systems-level perspective, examining the atmosphere not for specific species, but rather the complex patterns of interactions between species (Walker *et al.*, 2018, 2020; Sandora and Silk, 2020). However, current approaches still remain largely reductive in nature.

1.2 Current Approaches to Exoplanet Biosignatures

A broad variety of approaches have been considered for determining whether or not an exoplanet is inhabited. The majority of these have focused on information derived from spectral data, since higher resolution observations (such as direct imaging of the surface) are unlikely in the near-term (Schneider *et al.*, 2010).

1.2.1 Atmospheric Trace Gas Approaches

In addition to O_2 and CH_4, a number of other atmospheric species have been investigated as potential biosignature candidates.

Seager *et al.* (2013a) proposed a three-type classification system for atmospheric trace gases. The types in this system are Type I, which are the direct result of the thermodynamics of metabolic processes (usually redox reactions), *e.g.* the carbon dioxide and water produced by the metabolism of carbohydrates in heterotrophs; Type II, which are the byproducts of biomass production, *e.g.*, the oxygen produced during the photosynthesis of sugars and starches by phototrophs; and Type III, molecules produced by biology for specialized purposes, *e.g.*, volatile organic compounds produced by plants as a defense mechanism (Seager *et al.*, 2013a).

Based on this classification scheme, Seager *et al.* (2016) created a preliminary list of small volatile molecules of potential biological relevance, with a particular focus

on Type II biosignatures due to their high abundance on Earth. As the list has over fourteen thousand molecules, however, it still occupies an extremely large space of possible biosignatures, and makes a targeted search more difficult.

1.2.2 Other Spectral Approaches

Spectral non-atmospheric biosignatures have also been proposed. The "red edge" is a spectral signature caused by the reflectance of chlorophyll-using vegetation between 700 and 750 nm. It has been suggested as a possible biosignature (Tinetti *et al.*, 2006; Kiang *et al.*, 2018), and, judging from Earth's example, is viable over a large fraction of a biosphere's history (O'Malley-James and Kaltenegger, 2019). The circular polarization of reflected light has also been suggested as a possible biosignature, taking advantage of the chiral nature of most phototrophic pigments (Sparks *et al.*, 2020). However, reflectance of certain minerals can mimic the red edge, raising the risk of a false positive of life detection (Seager *et al.*, 2005). Additionally, the specific pigmentation used by extraterrestial biospheres must be assumed carefully (Kiang *et al.*, 2007), as, for example, the red edge was likely more of a "purple edge" during Earth's anoxic period due to the then-wider prevalence of retinal photopigments (Sanromá *et al.*, 2013). Biofluorescent pigments on planets that receive high UV fluxes have also been suggested as a potential spectral signature of life, though this also suffers many of the same drawbacks as the red edge with regards to false positives from mineral reflectance and fluorescence (O'Malley-James and Kaltenegger, 2018).

Due to the possibilities of false positive and negatives associated with the "smoking gun" approach demonstrated above, there is movement in astrobiology away from "smoking guns" and towards the use of more generalizable and statistically-grounded methods (Walker *et al.*, 2018, 2020).

1.2.3 Thermodynamic Approaches

A more systems-level approach under consideration is, rather than searching for specific trace gases, to instead analyze the overall thermodynamic state of the atmosphere. As life as a physical process can often drive its environment away from thermodynamic equilibrium, looking for atmospheres that are considerably far from equilibrium may be an indicator of life (Lovelock and Margulis, 1974). Furthermore, this degree of thermodynamic disequilibria can be quantified in a number of ways.

Simoncini *et al.* (2013) focused on the power, P, required to maintain the atmospheric disequilibrium as their primary metric, using the equation

$$P = -J_{\mathrm{A}} \Delta_{\mathrm{r}} G = J_{\mathrm{A}} RT \ln \frac{K_{\mathrm{eq}} p_{\mathrm{A}}}{p - p_{\mathrm{A}}} \tag{1.1}$$

where J_{A} is the flux of the species A being oxidized or reduced, $\Delta_{\mathrm{r}} G$ is the Gibbs energy gained by the system during the reverse reaction, R is the univeral gas constant (8.314 joules per Kelvin), T is the temperature in Kelvin, and p and p_{A} represent the total partial pressure and the partial pressure of species A, respectively.

Using this technique, they found that, in the case of the Earth, 0.67 terrawatts are required to preserve the disequilibrium between O_2 and CH_4. However, Simoncini *et al.* (2013) point out this is on the same order for magnitude as the power requirements of many abiotic geochemical processes on Earth.

Krissansen-Totton *et al.* (2016) developed a more complex metric, ϕ, based on the total Gibbs energy (in the form of ΔG) available in the atmosphere. The core of the thermodynamic equilibrium calculation for ϕ is the equation

$$G_{(T,\mathrm{P})} = \sum_i n_i \left(G^{\circ}_{i(T,\mathrm{P}_r)} + RT \ln \left(a_i \right) \right) = \sum_i n_i \left(G^{\circ}_{i(T,\mathrm{P}_r)} + RT \ln \left(P n_i \gamma_{fi} / n_T \right) \right) \tag{1.2}$$

where $G_{(T,\mathrm{P})}$ is the sum total Gibbs free energy of all species in the system, $G^{\circ}_{i(T,\mathrm{P}_r)}$ is the standard energy of formation for the ith species at temperature T and pressure P_r, P is the pressure, and n_i and γ_{fi} are the number of moles and the activity coefficient of species i, respectively.

In practice, this was calculated in terms of $\Delta G_{(T,\mathrm{P})}$, which Krissansen-Totton *et al.* (2016) demonstrated yields the same output as using $G_{(T,\mathrm{P})}$. To determine ϕ, the total Gibbs energy is calculated based on the mole fraction of of each species present at equilibrium, and then compared to the *actual* mole fraction n_i observed in the planet's atmosphere, producing the value ϕ, which represents the available Gibbs energy in the atmosphere:

$$\phi = \Delta G_{(T,P)}\left(n_i\right) - \Delta G_{(T,P)}\left(\bar{n}_i\right) \tag{1.3}$$

where $\bar{n}_i$ is the mole fraction that minimizes $\Delta G_{(T,P)}$.

The thermodynamics approach offers the benefit of being potentially agnostic to the biology driving the disequilibrium, and in principle working for both oxic and anoxic biospheres (Krissansen-Totton *et al.*, 2018), which alleviates some of the concerns associated with the "smoking gun" approach. However, the thermodynamic approach is also not without its drawbacks. In the case of ϕ, the metric requires calculations for species both in their gaseous and aqueous states for full effectiveness, as the ϕ for the Earth is unremarkable compared to that of other planetary atmospheres in the Solar System *unless* the presence of Earth's oceans is accounted for (as under equilibrium conditions, the N_2 and O_2 in the atmosphere should react with the liquid H_2O in the oceans to form aqueous HNO_3). Thus, its use to investigate exoplanets would require knowledge of whether a surface ocean existed on the exoplanet in question.

More generally speaking, measures of thermodynamic disequilbrium are not im-

mune to false positives in the form of large disequilibria driven by geochemical or photochemical activity (Seager and Bains, 2015; Seager, 2018). Indeed, Seager (2018) points out that all terrestrial atmospheres in our solar system are out of equilibrium. Nor are they immune to false negatives, as a large disequilibrium may end up being consumed by the biosphere as a source of energy, particularly during global shifts in metabolism like the one that occurred during the Great Oxygenation Event on Earth (Wogan and Catling, 2020).

1.3 Astrobiology and Complex System Science

It is clear that an even more holistic approach than thermodynamics is necessary if we are to develop robust and reliable exoplanet biosignatures. To reach this goal, astrobiology will have to embrace complex systems science.

Complex systems science, as the name suggests, considers how the components of a system are related to one another, and how the interactions of the components may be greater than just the sum of the system's parts (Shalizi, 2006; Siegenfeld and Bar-Yam, 2020). Unsurprisingly, complex systems science has played a significant role in ecology (Ford, 1999), particularly in ecosystems ecology, where it came into prominence in response to the inherent irreducibility of ecosystem dynamics (Jørgensen *et al.*, 1992). This in turn led to a further development of complex systems science itself, to accommodate for ecological features (such as cross-scale interactions and environmental variability) that were not typically found in "classical" complex systems (Anand *et al.*, 2010).

Despite the significance of complex systems science in ecology, the field's methodology has seen little use in astrobiology. Researchers investigating the origin of life were the first to seize upon complex systems techniques (Elser, 2003; Chela-Flores, 2013), and have been rewarded with increasingly sophisticated theoretical tools, such

as assembly theory (Cronin, 2020); this has since given rise to potential biosignatures for use within our Solar System (Marshall *et al.*, 2021). However, astrobiology as pertaining to exoplanets has been slow to adopt complex systems thinking (Walker *et al.*, 2018; Smith and Mathis, 2022). This is to the field's detriment, as biosignatures will be fundamentally enmeshed with the context of the ecosystem that generates them (Green *et al.*, 2021). Take, for example, CH_4–it is not just the product of a biological activity that's potentially detectable in exoplanet atmospheres (Thompson *et al.*, 2022), but is itself the foundation of whole microbial communities, some of which don't even use CH_4 directly in their metabolism, instead syntrophically relying on methanogens to supply reduced compounds to enable the reduction of sulfate and other inorganic species (Valentine, 2002)! Ecological modeling has been a part of biosignature research regarding our own solar system for quite some time (Grasby and Londry, 2007; Wright *et al.*, 2010; Fisher and Schulze-Makuch, 2013; Sumner, 2014; Cabrol, 2018; Ceja and Kane, 2019), but exoplanet scientists interested in, for example, CH_4 as a biosignature, have only recently begun to incorporate ecological modeling into their models (Affholder *et al.*, 2021; Nicholson *et al.*, 2022; Seeburger *et al.*, 2023), much less complex systems science.

1.3.1 Network Biosignatures

One approach to investigating complex systems is through the use of network science (Mitchell, 2006; Amaral and Ottino, 2004), "the new flag of complex systems theory" (Bascompte, 2007). Network, or graph, theory, represents complex systems as a network, with the subcomponents of the network being assigned points, or *nodes*, and then linked together via *edges* to represent the relationship and interactions between subcomponents. Network representations both allow a more intuitive understanding of the systems in question, and provide an avenue of quantifying their

structure, or *topology*, and behavior (Barabási and Pósfai, 2016). Network science has already seen use in the context of ecology, from the level of individual organisms all the way up to metabolic networks in ecosystems (Proulx *et al.*, 2005; May, 2006; Bascompte, 2007). The study of chemical reaction networks (CRNs), in particular, is a well-established topic within the field of complex systems science (Clarke, 1974; Temkin *et al.*, 1996; Angeli, 2009) and ecology and biology (Conradi *et al.*, 2006; Lente, 2015).

Within astrobiology, network-theory-informed approaches have been applied to estimating the likelihood of SETI communication (Davis, 2019), detecting mineralogical biosignatures (Chan *et al.*, 2019), and modeling the increase of chemical complexity that may have led to the origin of life (Adam *et al.*, 2021).

With regard to exoplanets, understanding the chemical reaction networks of atmospheres is of particular interest to the astrobiology community. In the context of atmospheres, Solé and Munteanu (2004) was one of the first efforts to analyze the topology of atmospheric CRNs. Their analysis constructed networks representing the atmospheric chemistry of Earth, Mars, Venus, Titan, a generic hydrocarbon-chemistry network for gas giants, and the interstellar medium. For comparison to biological systems, they also included the biochemical network of *E. coli*. This work was re-examined more recently by Wong *et al.* (2023), which conducted a more rigorous analysis, additionally incorporating atmospheres of early Earth and Pluto.

Solé and Munteanu (2004) found that the CRN of Earth's atmosphere uniquely exhibited a scale-free degree distribution, though Wong *et al.* (2023) was unable to replicate this result, finding that the degree distribution was closer to log-normal. With that said, the Earth's degree distribution and network topology was still unique among the solar systems atmosphere, exhibiting hierarchical and modular features. Modules, in this case, are defined as reactions with high topological overlap; for

example, Earth's atmospheric CRN seems to be dominated by two modules, made up of the respective reactants and products of reactions involving Cl and OH. These modules are thought to occur due to the extremely reactive nature of the Cl and OH radicals, capable of reacting with almost every other element present in the Earth's atmospheres. Wong *et al.* (2023) detected similar communities within the network, dominated by NO_2, the Chapman O_2-O_3-O chemistry cycle, CH_2-bearing chlorine compounds, and chlorofluorocarbons in addition to the OH and Cl modules.

This topological distinction was reflected in the network metrics, as well, with the mean degree of the Earth's network being lower than any other network (6.27, versus the next lowest mean degree of 8.33 from Venus) and a higher average shortest path length than any other network (2.75, versus the next highest of 2.07, also from Venus) (see Section 2.3.1 and Table 2.1 for definitions of these metrics). Solé and Munteanu (2004) suggested that these unique topological characteristics could be due to the coupling between the Earth's atmosphere and biosphere, particularly since the *E. coli* network yielded similar network metrics (mean degree of 6.24, average shortest path length of 3.02, and modular features). Wong *et al.* (2023) found a similarly high average shortest path length value for the modern Earth atmosphere, and noted that it was comparable to that of the metabolic network, neural network, and marine food web that they examined; they also found that both Earth's atmospheric network and these biological networks exhibited lower transitivity values than the other atmospheric networks.

Similar patterns of organization are seen at all scales of the Earth's biosphere (Wuchty *et al.*, 2006; Kim *et al.*, 2019), from cellular biochemical networks (Barabasi and Oltvai, 2004; Daniels *et al.*, 2018) to those in individual organisms (Jeong *et al.*, 2000) all the way up to whole ecosystems (Solé *et al.*, 2002; Marquet *et al.*, 2005). These patterns are distinguishable from non-biological chemical reaction networks

(Jolley and Douglas, 2011), and thus it is possible that the topology of a planet's atmospheric CRN may be indicative of the presence of a biosphere interfacing with the atmosphere. This may be a reflection of 'universal laws of life' (Solé *et al.*, 2002; Mariscal and Fleming, 2018), patterns or motifs that emerge due to life's constraints as a physical system, regardless of the type of biochemistry involved; existing suggested laws include allometric scaling of cardiovascular blood vessels (West *et al.*, 1997), the vascular system of plants (West *et al.*, 1999), body size (Kuikka, 2003), and metabolic rate (Sousa *et al.*, 2008; Burger *et al.*, 2020). Similar network topology is also seen in technological networks (Solé *et al.*, 2011), especially in telecommunications networks such as the internet (Vázquez *et al.*, 2002; Wang *et al.*, 2005), suggesting that network science could be applied to the search for technosignatures, as well. However, it's worth noting that the exact nature of the topology associated with biological and technological systems is not universally agreed upon (Broido and Clauset, 2019; Smith *et al.*, 2021).

While the work of Solé and Munteanu (2004) and Wong *et al.* (2023) provide an intriguing glimpse of the promise of network theory-based biosignatures for planetary atmospheres, their approach is not without its limits. In the hunt for signs of life, astrobiologists will not be comparing Jovian worlds to terrestrial ones–rather, they shall be examining *populations* of Earth-like worlds that could be largely identical in terms of atmospheric composition. Furthermore, for CRN topology to be a viable biosignature, metrics to quantify it also need to be developed. Mathematical properties, such as algebraic closure and self-maintenance have been proposed (Centler and Dittrich, 2007), as has the ability for the network to return to thermodynamic equilibrium (Estrada, 2012). Network topological properties, such as the degree distribution and average shorter path length (explored in detail in Section 2.3.1), may also prove useful, either individually or in aggregate. In both cases, the central question is, can

these network measurements provide more information about the atmosphere and its behavior than from the abundance of gases or their thermodynamic state alone?

1.4 Conclusion

The search for life requires creative tools grounded in complex systems science, and several candidates for aiding in biosignature detection have been put forward. In particular, approaches grounded in network theory may have considerable promise in yielding a robust and universal biosignature for exoplanet atmospheres, but require further investigation. To remedy this need, I investigate atmospheric chemical reaction network topology as a method of understanding planetary atmospheres in the following chapters.

First, in Chapter 2 I provide a detailed, step-by-step description of the methods used in analyzing chemical reaction networks, and in large populations of modeled biotic and abiotic atmospheres, including how distinguishability between the distributions of network metrics of these populations can be quantified, and how the results might be validated.

In Chapter 3 I then evaluate the responsiveness of network topology using thermochemical equilibrium on hot Jupiters as a proof of concept that network topology can be reflective of the chemical and physical behavior of modeled atmospheres.

Having demonstrated that CRN topology can be useful for analyzing atmospheric chemistry, in Chapter 4 I take the methods developed and test them by applying them to terrestrial planets with biology (in the form of methanogenesis on populations of modeled Archean Earth analogues) and technological civilization (in the form of CFC-12 emissions on a population of modern Earth analogue), comparing the ability of network metrics to distinguish between biotic/abiotic and technological/non-technological influence atmospheres to distinguishability based on the abundance of

atmospheric species alone. My results indicate that network metrics perform better in distinguishing between the modeled atmospheres than abundance of atmospheric species alones.

The final chapter summarizes and concludes my findings, explores future applications for network theory biosignatures for exoplanets, and discusses the observational constraints required to make use of them.

REFERENCES

Adam, Z. R., A. C. Fahrenbach, S. M. Jacobson, B. Kacar and D. Y. Zubarev, "Radiolysis generates a complex organosynthetic chemical network", Scientific Reports 11, 1, 1–10 (2021).

Affholder, A., A. Bixel, B. Sauterey, R. Ferrière, D. Apai and S. Mazevet, "Predicting quantitative biosignature patterns from populations of exoplanets", in "AGU Fall Meeting Abstracts", vol. 2021, pp. P55D–1969 (2021).

Ahrer, E.-M., L. Alderson, N. M. Batalha, N. E. Batalha, J. L. Bean, T. G. Beatty, T. J. Bell, B. Benneke, Z. K. Berta-Thompson, A. L. Carter, I. J. M. Crossfield, N. Espinoza, A. D. Feinstein, J. J. Fortney, N. P. Gibson, J. M. Goyal, E. M.-R. Kempton, J. Kirk, L. Kreidberg, M. López-Morales, M. R. Line, J. D. Lothringer, S. E. Moran, S. Mukherjee, K. Ohno, V. Parmentier, C. Piaulet, Z. Rustamkulov, E. Schlawin, D. K. Sing, K. B. Stevenson, H. R. Wakeford, N. H. Allen, S. M. Birkmann, J. Brande, N. Crouzet, P. E. Cubillos, M. Damiano, J.-M. Désert, P. Gao, J. Harrington, R. Hu, S. Kendrew, H. A. Knutson, P.-O. Lagage, J. Leconte, M. Lendl, R. J. MacDonald, E. M. May, Y. Miguel, K. Molaverdikhani, J. I. Moses, C. A. Murray, M. Nehring, N. K. Nikolov, D. J. M. Petit dit de la Roche, M. Radica, P.-A. Roy, K. G. Stassun, J. Taylor, W. C. Waalkes, P. Wachiraphan, L. Welbanks,

P. J. Wheatley, K. Aggarwal, M. K. Alam, A. Banerjee, J. K. Barstow, J. Blecic, S. L. Casewell, Q. Changeat, K. L. Chubb, K. D. Colón, L.-P. Coulombe, T. Daylan, M. de Val-Borro, L. Decin, L. A. Dos Santos, L. Flagg, K. France, G. Fu, A. García Muñoz, J. E. Gizis, A. Glidden, D. Grant, K. Heng, T. Henning, Y.-C. Hong, J. Inglis, N. Iro, T. Kataria, T. D. Komacek, J. E. Krick, E. K. H. Lee, N. K. Lewis, J. Lillo-Box, J. Lustig-Yaeger, L. Mancini, A. M. Mandell, M. Mansfield, M. S. Marley, T. Mikal-Evans, G. Morello, M. C. Nixon, K. Ortiz Ceballos, A. A. A. Piette, D. Powell, B. V. Rackham, L. Ramos-Rosado, E. Rauscher, S. Redfield, L. K. Rogers, M. T. Roman, G. M. Roudier, N. Scarsdale, E. L. Shkolnik, J. Southworth, J. J. Spake, M. E. Steinrueck, X. Tan, J. K. Teske, P. Tremblin, S.-M. Tsai, G. S. Tucker, J. D. Turner, J. A. Valenti, O. Venot, I. P. Waldmann, N. L. Wallack, X. Zhang, S. Zieba and JWST Transiting Exoplanet Community Early Release Science Team, "Identification of carbon dioxide in an exoplanet atmosphere", Nature 614, 7949 (2023).

Akins, A. B., A. P. Lincowski, V. S. Meadows and P. G. Steffes, "Complications in the alma detection of phosphine at venus", The Astrophysical Journal Letters 907, 2, L27 (2021).

Alderson, L., H. R. Wakeford, M. K. Alam, N. E. Batalha, J. D. Lothringer, J. Adams Redai, S. Barat, J. Brande, M. Damiano, T. Daylan *et al.*, "Early release science of the exoplanet wasp-39b with jwst nirspec g395h", Nature 614, 7949, 664–669 (2023).

Amaral, L. A. and J. M. Ottino, "Complex networks: Augmenting the framework for the study of complex systems", The European Physical Journal B 38, 147–162 (2004).

Anand, M., A. Gonzalez, F. Guichard, J. Kolasa and L. Parrott, "Ecological systems as complex systems: challenges for an emerging science", Diversity 2, 3, 395–410 (2010).

Angeli, D., "A tutorial on chemical reaction networks dynamics", in "2009 European Control Conference (ECC)", pp. 649–657 (IEEE, 2009).

Arney, G., S. D. Domagal-Goldman, V. S. Meadows, E. T. Wolf, E. Schwieterman, B. Charnay, M. Claire, E. Hébrard and M. G. Trainer, "The pale orange dot: the spectrum and habitability of hazy Archean Earth", Astrobiology 16, 11, 873–899 (2016).

Bains, W., J. J. Petkowski, S. Seager, S. Ranjan, C. Sousa-Silva, P. B. Rimmer, Z. Zhan, J. S. Greaves and A. M. Richards, "Phosphine on venus cannot be explained by conventional processes", Astrobiology 21, 10, 1277–1304 (2021).

Bains, W., S. Seager and A. Zsom, "Photosynthesis in hydrogen-dominated atmospheres", Life 4, 4, 716–744 (2014).

Barabasi, A.-L. and Z. N. Oltvai, "Network biology: understanding the cell's functional organization", Nature Reviews Genetics 5, 2, 101–113 (2004).

Barabási, A.-L. and M. Pósfai, *Network Science* (Cambridge University Press, Cambridge, United Kingdom, 2016).

Bascompte, J., "Networks in ecology", Basic and Applied Ecology 8, 6, 485–490 (2007).

Broadley, M. W., P. H. Barry, C. J. Ballentine, L. A. Taylor and R. Burgess, "End-permian extinction amplified by plume-induced release of recycled lithospheric volatiles", Nature Geoscience 11, 9, 682–687 (2018).

Broido, A. D. and A. Clauset, "Scale-free networks are rare", Nature Communications 10, 1, 1017 (2019).

Burger, J. R., C. Hou, C. Hall and J. R. Brown, "Universal rules of life: metabolic rates, biological times and the equal fitness paradigm", bioRxiv (2020).

Cabrol, N. A., "The coevolution of life and environment on Mars: An ecosystem perspective on the robotic exploration of biosignatures", Astrobiology 18, 1, 1–27 (2018).

Ceja, A. Y. and S. Kane, "The search for extraterrestrial life: An astro-ecological modeling approach for characterizing exoplanet habitability", in "American Astronomical Society Meeting Abstracts# 233", vol. 233, pp. 346–01 (2019).

Centler, F. and P. Dittrich, "Chemical organizations in atmospheric photochemistries—A new method to analyze chemical reaction networks", Planetary and Space Science 55, 4, 413–428 (2007).

Chan, M. A., N. W. Hinman, S. L. Potter-McIntyre, K. E. Schubert, R. J. Gillams, S. M. Awramik, P. J. Boston, D. M. Bower, D. J. Des Marais, J. D. Farmer, T. Z. Jia, P. L. King, R. M. Hazen, R. J. Léveillé, D. Papineau, K. R. Rempfert, M. Sánchez-Román, J. R. Spear, G. Southam, J. C. Stern and H. J. Cleaves, "Deciphering biosignatures in planetary contexts", Astrobiology 19, 9, 1075–1102 (2019).

Charbonneau, D., T. M. Brown, R. W. Noyes and R. L. Gilliland, "Detection of an extrasolar planet atmosphere", The Astrophysical Journal 568, 1, 377 (2002).

Chela-Flores, J., "From systems chemistry to systems astrobiology: life in the universe

as an emergent phenomenon", International Journal of Astrobiology 12, 1, 8–16 (2013).

Clarke, B. L., "Graph theoretic approach to the stability analysis of steady state chemical reaction networks", The Journal of Chemical Physics 60, 4, 1481–1492 (1974).

Conradi, C., J. Saez-Rodriguez, E.-D. Gilles and J. Raisch, "Chemical reaction network theory... a tool for systems biology", Proceedings of the 5th MATHMOD (2006).

Coustenis, A., J. Schneider, R. Wittemberg, E. Chazsefiere, T. Guillot, A. Penny, T. Greene, H. Rauer and D. Bockelee-Morvan, "High resolution ground-based spectroscopy of 51 Peg b: Search for atmospheric signatures", in "Brown Dwarfs and Extrasolar Planets", vol. 134, p. 296 (1998).

Cronin, L., "Exploring the transition from prebiotic chemistry to biology using molecular assembly theory", in "AGU Fall Meeting Abstracts", vol. 2020, pp. P001–01 (2020).

Daniels, B. C., H. Kim, D. Moore, S. Zhou, H. Smith, B. Karas, S. A. Kauffman and S. I. Walker, "Logic and connectivity jointly determine criticality in biological gene regulatory networks", Physical Review Letters 121, 13 (2018).

Davis, R., "Searching for extraterrestrial intelligence by locating potential ET communication networks in space", arXiv:1907.05259 [physics] (2019).

Domagal-Goldman, S. D., A. Segura, M. W. Claire, T. D. Robinson and V. S. Meadows, "Abiotic ozone and oxygen in atmospheres similar to prebiotic Earth", The Astrophysical Journal 792, 2, 90 (2014).

Elser, J. J., "Biological stoichiometry: a theoretical framework connecting ecosystem ecology, evolution, and biochemistry for application in astrobiology", International Journal of Astrobiology 2, 3, 185–193 (2003).

Estrada, E., "Returnability as a criterion of disequilibrium in atmospheric reactions networks", Journal of Mathematical Chemistry 50, 6, 1363–1372 (2012).

Fisher, T. M. and D. Schulze-Makuch, "Nutrient and population dynamics in a subglacial reservoir: a simulation case study of the Blood Falls ecosystem with implications for astrobiology", International Journal of Astrobiology 12, 4, 304–311 (2013).

Ford, F. A., *Modeling the Environment: An Introduction to System Dynamics Models of Environmental Systems* (Island press, 1999).

Fujii, Y., D. Angerhausen, R. Deitrick, S. Domagal-Goldman, J. L. Grenfell, Y. Hori, S. R. Kane, E. Pallé, H. Rauer, N. Siegler, K. Stapelfeldt and K. B. Stevenson, "Exoplanet biosignatures: observational prospects", Astrobiology 18, 6, 739–778 (2018).

Gebauer, S., I. Vilović, J. L. Grenfell, F. Wunderlich, F. Schreier and H. Rauer, "Influence of biomass emissions upon habitability, biosignatures and detectability in Earth-like atmospheres", arXiv p. arXiv:2102.00220 (2021).

Grasby, S. E. and K. L. Londry, "Biogeochemistry of hypersaline springs supporting a mid-continent marine ecosystem: an analogue for martian springs?", Astrobiology 7, 4, 662–683 (2007).

Greaves, J. S., A. Richards, W. Bains, P. B. Rimmer, H. Sagawa, D. L. Clements,

S. Seager, J. J. Petkowski, C. Sousa-Silva, S. Ranjan *et al.*, "Phosphine gas in the cloud decks of venus", Nature Astronomy 5, 7, 655–664 (2021).

Green, J., T. Hoehler, M. Neveu, S. Domagal-Goldman, D. Scalice and M. Voytek, "Call for a framework for reporting evidence for life beyond earth", Nature 598, 7882, 575–579 (2021).

Grenfell, J. L., "A review of exoplanetary biosignatures", Physics Reports 713, 1–17 (2017).

Guzmán-Marmolejo, A., A. Segura and E. Escobar-Briones, "Abiotic production of methane in terrestrial planets", Astrobiology 13, 6, 550–559 (2013).

Haas, J. R., "The potential feasibility of chlorinic photosynthesis on exoplanets", Astrobiology 10, 9, 953–963 (2010).

Haldane, J. B. S., *Possible Worlds and Other Essays* (Chatto and Windus, London, 1927), 1st edn.

Haqq-Misra, J., T. J. Fauchez, E. W. Schwieterman and R. Kopparapu, "Disruption of a planetary nitrogen cycle as evidence of extraterrestrial agriculture", The Astrophysical Journal Letters 929, 2, L28 (2022a).

Haqq-Misra, J., E. W. Schwieterman, H. Socas-Navarro, R. Kopparapu, D. Angerhausen, T. G. Beatty, S. Berdyugina, R. Felton, S. Sharma, G. G. De la Torre and D. Apai, "Searching for technosignatures in exoplanetary systems with current and future missions", Acta Astronautica 198, 194–207 (2022b).

Harman, C. E. and S. Domagal-Goldman, "Biosignature false positives", in "Handbook of Exoplanets", edited by H. J. Deeg and J. A. Belmonte, pp. 1–22 (Springer International Publishing, Cham, 2018).

Jeong, H., B. Tombor, R. Albert, Z. N. Oltvai and A.-L. Barabási, "The large-scale organization of metabolic networks", Nature 407, 6804, 651–654 (2000).

Jolley, C. and T. Douglas, "Topological biosignatures: large-scale structure of chemical networks from biology and astrochemistry", Astrobiology 12, 1, 29–39 (2011).

Jørgensen, S. E., B. C. Patten and M. Straškraba, "Ecosystems emerging: toward an ecology of complex systems in a complex future", Ecological Modelling 62, 1-3, 1–27 (1992).

Kaltenegger, L., Z. Lin and S. Rugheimer, "Finding signs of life on transiting Earthlike planets: high-resolution transmission spectra of Earth through time around FGKM host stars", The Astrophysical Journal 904, 1, 10 (2020).

Kiang, N. Y., S. Domagal-Goldman, M. N. Parenteau, D. C. Catling, Y. Fujii, V. S. Meadows, E. W. Schwieterman and S. I. Walker, "Exoplanet biosignatures: at the dawn of a new era of planetary observations", Astrobiology 18, 6, 619–629 (2018).

Kiang, N. Y., A. Segura, G. Tinetti, R. E. Blankenship, M. Cohen, J. Siefert, D. Crisp, V. S. Meadows and others, "Spectral signatures of photosynthesis: Coevolution with other stars and the atmosphere on extrasolar worlds", Astrobiology 7, 3, 486–7 (2007).

Kim, H., H. B. Smith, C. Mathis, J. Raymond and S. I. Walker, "Universal scaling across biochemical networks on Earth", Science Advances 5, 1, eaau0149 (2019).

Kleinböhl, A., K. Willacy, A. J. Friedson, P. Chen and M. R. Swain, "Buildup of abiotic oxygen and ozone in moist atmospheres of temperate terrestrial exoplanets and its impact on the spectral fingerprint in transit observations", The Astrophysical Journal 862, 2, 92 (2018).

Klobas, J. E. and D. Wilmouth, "Volcanogenic chlorofluorocarbons and the recent cfc anomalies", (2019).

Kopparapu, R., G. Arney, J. Haqq-Misra, J. Lustig-Yaeger and G. Villanueva, "Nitrogen dioxide pollution as a signature of extraterrestrial technology", The Astrophysical Journal 908, 2, 164 (2021).

Krissansen-Totton, J., G. N. Arney, D. C. Catling, R. Felton, J. Fortney, R. Garland, P. Irwin, R. Kopparapu, O. Lehmer, J. Lustig-Yaeger and others, "Atmospheric disequilibrium as an exoplanet biosignature: Opportunities for next generation telescopes", BAAS 51, 3, 158 (2019).

Krissansen-Totton, J., D. S. Bergsman and D. C. Catling, "On detecting biospheres from chemical thermodynamic disequilibrium in planetary atmospheres", Astrobiology 16, 1, 39–67 (2016).

Krissansen-Totton, J., S. Olson and D. C. Catling, "Disequilibrium biosignatures over Earth history and implications for detecting exoplanet life", Science Advances 4, 1, eaao5747 (2018).

Kuikka, J. T., "Scaling laws in physiology: relationships between size, function, metabolism and life expectancy", International Journal of Nonlinear Sciences and Numerical Simulation 4, 4, 317–328 (2003).

Lente, G., *Deterministic Kinetics in Chemistry and Systems Biology: the Dynamics of Complex Reaction Networks* (Springer, 2015).

Lovelock, J. E. and L. Margulis, "Atmospheric homeostasis by and for the biosphere: the Gaia hypothesis", Tellus 26, 1-2, 2–10 (1974).

Lustig-Yaeger, J., V. S. Meadows and A. P. Lincowski, "The detectability and characterization of the TRAPPIST-1 exoplanet atmospheres with JWST", The Astronomical Journal 158, 1, 27 (2019).

Madhusudhan, N., "Exoplanetary atmospheres: key insights, challenges and prospects", Annual Review of Astronomy and Astrophysics 57, 617–663 (2019).

Mariscal, C. and L. Fleming, "Why we should care about universal biology", Biological Theory 13, 2, 121–130 (2018).

Marquet, P. A., R. A. Quiñones, S. Abades, F. Labra, M. Tognelli, M. Arim and M. Rivadeneira, "Scaling and power-laws in ecological systems", Journal of Experimental Biology 208, 9, 1749–1769 (2005).

Marshall, S. M., C. Mathis, E. Carrick, G. Keenan, G. J. Cooper, H. Graham, M. Craven, P. S. Gromski, D. G. Moore, S. I. Walker *et al.*, "Identifying molecules as biosignatures with assembly theory and mass spectrometry", Nature Communications 12, 1, 3033 (2021).

May, R. M., "Network structure and the biology of populations", Trends in Ecology & Evolution 21, 7, 394–399 (2006).

Mayor, M. and D. Queloz, "A Jupiter-mass companion to a solar-type star", Nature 378, 6555, 355–359 (1995).

Meadows, V. S., "Reflections on O_2 as a biosignature in exoplanetary atmospheres", Astrobiology 17, 10, 1022–1052 (2017).

Meadows, V. S., C. T. Reinhard, G. N. Arney, M. N. Parenteau, E. W. Schwieterman, S. D. Domagal-Goldman, A. P. Lincowski, K. R. Stapelfeldt, H. Rauer, S. DasSarma, S. Hegde, N. Narita, R. Deitrick, J. Lustig-Yaeger, T. W. Lyons, N. Siegler

and J. L. Grenfell, "Exoplanet biosignatures: understanding oxygen as a biosignature in the context of its Environment", Astrobiology 18, 6, 630–662 (2018).

Mitchell, M., "Complex systems: Network thinking", Artificial Intelligence 170, 18, 1194–1212 (2006).

Moutou, C., A. Coustenis, J. Schneider, R. St Gilles, M. Mayor, D. Queloz and A. Kaufer, "Search for spectroscopical signatures of transiting HD 209458b's exosphere", Astronomy & Astrophysics 371, 1, 260–266 (2001).

National Academy of Sciences, "Decadal survey on astronomy and astrophysics 2020", (2021).

Nicholson, A., S. Daines, N. Mayne, J. Eager-Nash, T. Lenton and K. Kohary, "Predicting biosignatures for nutrient-limited biospheres", Monthly Notices of the Royal Astronomical Society 517, 1, 222–239 (2022).

O'Malley-James, J. T. and L. Kaltenegger, "Biofluorescent worlds: global biological fluorescence as a biosignature", Monthly Notices of the Royal Astronomical Society 481, 2, 2487–2496 (2018).

O'Malley-James, J. T. and L. Kaltenegger, "Expanding the timeline for Earth's photosynthetic red edge biosignature", The Astrophysical Journal Letters 879, 2, L20 (2019).

Proulx, S. R., D. E. Promislow and P. C. Phillips, "Network thinking in ecology and evolution", Trends in Ecology & Evolution 20, 6, 345–353 (2005).

Rein, H., Y. Fujii and D. S. Spiegel, "Some inconvenient truths about biosignatures involving two chemical species on Earth-like exoplanets", Proceedings of the National Academy of Sciences 111, 19, 6871–6875 (2014).

Reinhard, C. T., S. L. Olson, E. W. Schwieterman and T. W. Lyons, "False negatives for remote life detection on ocean-bearing planets: lessons from the early Earth", Astrobiology 17, 4, 287–297 (2017).

Sagan, C., W. R. Thompson, R. Carlson, D. Gurnett and C. Hord, "A search for life on Earth from the Galileo spacecraft", Nature 365, 6448, 715–721 (1993).

Sandora, M. and J. Silk, "Biosignature surveys to exoplanet yields and beyond", Monthly Notices of the Royal Astronomical Society 495, 1, 1000–1015 (2020).

Sanromá, E., E. Pallé, M. Parenteau, N. Kiang, A. Gutiérrez-Navarro, R. López and P. Montañés-Rodríguez, "Characterizing the purple Earth: modeling the globally integrated spectral variability of the Archean Earth", The Astrophysical Journal 780, 1, 52 (2013).

Saxena, P., "Photobombing earth 2.0: diffraction-limit-related contamination and uncertainty in habitable planet spectra", The Astrophysical Journal Letters 934, 2, L32 (2022).

Schneider, J., A. Léger, M. Fridlund, G. J. White, C. Eiroa, T. Henning, T. Herbst, H. Lammer, R. Liseau, F. Paresce and others, "The far future of exoplanet direct characterization", Astrobiology 10, 1, 121–126 (2010).

Schwieterman, E. W., N. Y. Kiang, M. N. Parenteau, C. E. Harman, S. DasSarma, T. M. Fisher, G. N. Arney, H. E. Hartnett, C. T. Reinhard, S. L. Olson, V. S. Meadows, C. S. Cockell, S. I. Walker, J. L. Grenfell, S. Hegde, S. Rugheimer, R. Hu and T. W. Lyons, "Exoplanet biosignatures: a review of remotely detectable signs of life", Astrobiology 18, 6, 663–708 (2018).

Seager, S., "The search for habitable planets with biosignature gases framed by a

'biosignature Drake Equation"', International Journal of Astrobiology 17, 4, 294–302 (2018).

Seager, S. and W. Bains, "The search for signs of life on exoplanets at the interface of chemistry and planetary science", Science Advances 1, 2, e1500047 (2015).

Seager, S., W. Bains and R. Hu, "A biomass-based model to estimate the plausibility of exoplanet biosignature gases", The Astrophysical Journal 775, 2, 104 (2013a).

Seager, S., W. Bains and R. Hu, "Biosignature gases in H2-dominated atmospheres on rocky exoplanets", The Astrophysical Journal 777, 2, 95 (2013b).

Seager, S., W. Bains and J. Petkowski, "Toward a list of molecules as potential biosignature gases for the search for life on exoplanets and applications to terrestrial Biochemistry", Astrobiology 16, 6, 465–485 (2016).

Seager, S., E. L. Turner, J. Schafer and E. B. Ford, "Vegetation's red edge: a possible spectroscopic biosignature of extraterrestrial plants", Astrobiology 5, 3, 372–390 (2005).

Seeburger, R., P. M. Higgins, N. P. Whiteford and C. S. Cockell, "Linking methanogenesis in low-temperature hydrothermal vent systems to planetary spectra: Methane biosignatures on an archean-earth-like exoplanet", Astrobiology 23, 4, 415–430 (2023).

Shalizi, C. R., "Methods and techniques of complex systems science: An overview", Complex Systems Science in Biomedicine pp. 33–114 (2006).

Siegenfeld, A. F. and Y. Bar-Yam, "An introduction to complex systems science and its applications", Complexity 2020, 1–16 (2020).

Simoncini, E., N. Virgo and A. Kleidon, "Quantifying drivers of chemical disequilibrium: theory and application to methane in the Earth's atmosphere", Earth System Dynamics 4, 317–331 (2013).

Smith, H. B., H. Kim and S. I. Walker, "Scarcity of scale-free topology is universal across biochemical networks", Sci. Rep. 11, 1, 6542 (2021).

Smith, H. B. and C. Mathis, "The futility of exoplanet biosignatures", arXiv:2205.07921 (2022).

Solé, R. V., S. A. Levin, J. H. Brown, V. K. Gupta, B.-L. Li, B. T. Milne, C. Restrepo and G. B. West, "The fractal nature of nature: power laws, ecological complexity and biodiversity", Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 357, 1421, 619–626 (2002).

Solé, R. V. and A. Munteanu, "The large-scale organization of chemical reaction networks in astrophysics", Europhysics Letters (EPL) 68, 2, 170–176 (2004).

Solé, R. V., S. Valverde and C. Rodriguez-Caso, "Convergent evolutionary paths in biological and technological networks", Evolution: Education and Outreach 4, 3, 415–426 (2011).

Sousa, T., T. Domingos and S. Kooijman, "From empirical patterns to theory: a formal metabolic theory of life", Philosophical Transactions of the Royal Society B: Biological Sciences 363, 1502, 2453–2464 (2008).

Sousa-Silva, C., S. Seager, S. Ranjan, J. J. Petkowski, Z. Zhan, R. Hu and W. Bains, "Phosphine as a biosignature gas in exoplanet atmospheres", Astrobiology 20, 2, 235–268 (2020).

Sparks, W. B., M. N. Parenteau, R. E. Blankenship, T. A. Germer, C. Patty, K. M. Bott, C. M. Telesco and V. S. Meadows, "Spectropolarimetry of primitive phototrophs as global surface biosignatures", arXiv:2010.02133 (2020).

Suissa, G., A. M. Mandell, E. T. Wolf, G. L. Villanueva, T. Fauchez and R. kumar Kopparapu, "Dim prospects for transmission spectra of ocean Earths around M stars", The Astrophysical Journal 891, 1, 58 (2020).

Sumner, D. Y., "Microbial communities: Tracing growth processes from Antarctic lakes to early earth to other planets", in "AGU Fall Meeting Abstracts", vol. 2014, pp. B41P–01 (2014).

Swain, M. R., R. Estrela, G. M. Roudier, C. Sotin, P. Rimmer, A. Valio, R. West, K. Pearson, N. Huber-Feely and R. T. Zellem, "Detection of an atmosphere on a rocky exoplanet", arXiv:2103.05657 (2021).

Temkin, O. N., A. V. Zeigarnik and D. Bonchev, *Chemical Reaction Networks: A Graph-Theoretical Approach* (CRC Press, 1996).

Thompson, M. A., J. Krissansen-Totton, N. Wogan, M. Telus and J. J. Fortney, "The case and context for atmospheric methane as an exoplanet biosignature", Proceedings of the National Academy of Sciences 119, 14, e2117933119 (2022).

Tinetti, G., S. Rashby and Y. L. Yung, "Detectability of red-edge-shifted vegetation on terrestrial planets orbiting M stars", The Astrophysical Journal Letters 644, 2, L129 (2006).

Tremblay, L., M. R. Line, K. B. Stevenson, T. Kataria, R. T. Zellem, J. J. Fortney and C. V. Morley, "The detectability and constraints of biosignature gases in the near &

mid-infrared from transit transmission spectroscopy", arXiv:1912.10939 [astro-ph] (2019).

Truong, N. and J. Lunine, "Volcanically extruded phosphides as an abiotic source of venusian phosphine", Proceedings of the National Academy of Sciences 118, 29, e2021689118 (2021).

Valentine, D. L., "Biogeochemistry and microbial ecology of methane oxidation in anoxic environments: a review", Antonie Van Leeuwenhoek 81, 271–282 (2002).

Villanueva, G., M. Cordiner, P. Irwin, I. de Pater, B. Butler, M. Gurwell, S. Milam, C. Nixon, S. Luszcz-Cook, C. Wilson *et al.*, "No evidence of phosphine in the atmosphere of venus from independent analyses", Nature Astronomy 5, 7, 631–635 (2021).

Visscher, H., C. V. Looy, M. E. Collinson, H. Brinkhuis, J. H. Van Konijnenburg-Van Cittert, W. M. Kürschner and M. A. Sephton, "Environmental mutagenesis during the end-permian ecological crisis", Proceedings of the National Academy of Sciences 101, 35, 12952–12956 (2004).

Vázquez, A., R. Pastor-Satorras and A. Vespignani, "Large-scale topological and dynamical properties of the Internet", Physical Review E 65, 6, 066130 (2002).

Walker, S. I., W. Bains, L. Cronin, S. DasSarma, S. Danielache, S. Domagal-Goldman, B. Kacar, N. Y. Kiang, A. Lenardic, C. T. Reinhard, W. Moore, E. W. Schwieterman, E. L. Shkolnik and H. B. Smith, "Exoplanet biosignatures: future directions", Astrobiology 18, 6, 779–824 (2018).

Walker, S. I., L. Cronin, A. Drew, S. Domagal-Goldman, T. Fisher, M. Line and

C. Millsaps, "Probabilistic biosignature frameworks", in "Planetary Astrobiology", p. 477 (University of Arizona Press, 2020).

Wang, W.-X., B.-H. Wang, B. Hu, G. Yan and Q. Ou, "General dynamics of topology and traffic on weighted technological networks", Physical Review Letters 94, 18, 188702 (2005).

West, G. B., J. H. Brown and B. J. Enquist, "A general model for the origin of allometric scaling laws in biology", Science 276, 5309, 122–126 (1997).

West, G. B., J. H. Brown and B. J. Enquist, "A general model for the structure and allometry of plant vascular systems", Nature 400, 6745, 664–667 (1999).

Wogan, N., J. Krissansen-Totton and D. C. Catling, "Abundant atmospheric methane from volcanism on terrestrial planets is unlikely and strengthens the case for methane as a biosignature", arXiv:2009.07761 (2020).

Wogan, N. F. and D. C. Catling, "When is chemical disequilibrium in Earth-like planetary atmospheres a biosignature versus an anti-biosignature? Disequilibria from dead to living worlds", The Astrophysical Journal 892, 2, 127 (2020).

Woitke, P., O. Herbort, C. Helling, E. Stüeken, M. Dominik, P. Barth and D. Samra, "Coexistence of CH4, CO2 and H2O in exoplanet atmospheres", arXiv:2010.12241 (2020).

Wong, M. L., A. Prabhu, J. Williams, S. M. Morrison and R. M. Hazen, "Toward network-based planetary biosignatures: atmospheric chemistry as unipartite, unweighted, undirected networks", Journal of Geophysical Research: Planets 128, 6, e2022JE007658 (2023).

Wordsworth, R. and R. Pierrehumbert, "Abiotic oxygen-dominated atmospheres on terrestrial habitable zone planets", The Astrophysical Journal Letters 785, 2, L20 (2014).

Wright, K., D. Gleeson, C. Williamson, S. Grasby, J. Spear, R. Pappalardo and A. Templeton, "A sulfur-based glacial ecosystem as a model for the habitability of Europa and Mars", in "Astrobiology Science Conference 2010: Evolution and Life: Surviving Catastrophes and Extremes on Earth and Beyond", vol. 1538, p. 5526 (2010).

Wuchty, S., E. Ravasz and A.-L. Barabási, "The architecture of biological networks", in "Complex Systems Science in Biomedicine", pp. 165–181 (Springer, 2006).

Wunderlich, F., M. Scheucher, J. L. Grenfell, F. Schreier, C. Sousa-Silva, M. Godolt and H. Rauer, "Detectability of biosignatures on lhs 1140 b", Astronomy & Astrophysics 647, A48 (2021).

Chapter 2

A METHOD OF ANALYZING THE ATMOSPHERIC CHEMICAL REACTION NETWORKS OF LARGE POPULATIONS OF PLANETARY ATMOSPHERES

2.1 Abstract

Historically, the search for life on exoplanets has focused on the search for "smoking guns", single molecular species such as O_2 or CH_4 that were believed to be unambigous signs of life. However, this approach invites the very real risk of false positives and false negatives, and it is clear that a more reliable, systems-level approach is required. One such method is analyzing the topology of atmospheric chemical reaction networks (CRNs). Here, I demonstrate this technique step-by-step, first using a single, simple chemical reaction network to explain CRN topological metrics, and then by measuring and comparing the distributions of topological metrics from two populations of modeled Archean Earth-like atmospheres. One population is wholly abiotic, while the other incorporates methanogenesis reactions to represent a simple biosphere. I then quantify how distinguishable these distributions are using percentage of distribution overlap and the Kolmogirov-Smirnoff test. Finally, I use perturbation testing to validate our results.

Atmospheric CRN topology provides a powerful tool for better understanding the chemical and physical behavior of exoplanet atmospheres. With further development, it may provide a solution to the dilemma of false positives and negatives in biosignature detection.

2.2 Introduction

With the prospect of spectral analysis of terrestrial exoplanet atmospheres on the near-term horizon, growing attention has been paid towards the implications for astrobiology (Catling *et al.*, 2018; Walker *et al.*, 2018). As discussed in Section 1.1, astrobiologists were primarily concerned historically with searching for atmospheric "smoking guns", single species that were believed to be unambiguously associated with life, such as O_2 and CH_4 (Grenfell, 2017; Fujii *et al.*, 2018). However, it has become apparent that a single species cannot necessarily be trusted as a reliable indicator of life–O_2 has come under scrutiny has the potential risks for false positives (Domagal-Goldman *et al.*, 2014; Meadows *et al.*, 2018; Harman and Domagal-Goldman, 2018) and false negatives (Wordsworth and Pierrehumbert, 2014). For example, detectable levels of O_2 could be purely generated by photolysis on planets with high water vapor content(Kleinböhl *et al.*, 2018), which could otherwise be misinterpreted as evidence of photosynthesis. On the other hand, neither does the absence of O_2 indicate an abiotic world–after all, Earth's biological O_2 was miniscule for much of its history, despite a thriving anaerobic biosphere (Reinhard *et al.*, 2017).

CH_4 has also been subject to similar concerns (Guzmán-Marmolejo *et al.*, 2013); even CH_4 in the presence of oxygen may not be a reliable biosignature, due to the possibility of the atmosphere of an exoplanet's unseen moon contaminating the spectra of its host exoplanet, and leading to the appearance of a single body with a strong atmospheric disequilibrium (Rein *et al.*, 2014; Saxena, 2022).

Thus, an approach that focuses not on single species, but rather one that examines atmospheres at a systems level, is necessary (Walker *et al.*, 2018, 2020; Sandora and Silk, 2020). One natural approach to examining the behavior of complex systems is by representing them as networks (Mitchell, 2006; Amaral and Ottino, 2004), which

both allows a more intuitive understanding of the systems in question, and provides an avenue of quantifying their structure and behavior (Barabási and Pósfai, 2016) (see Section 1.3 for more detail).

Chemical reaction networks (CRNs) are a well-established topic within the field of network science (Clarke, 1974; Temkin *et al.*, 1996; Angeli, 2009). In the context of atmospheres, Solé and Munteanu (2004) was one of the first efforts to analyze the topology of atmospheric CRNs. Their analysis constructed networks representing the atmospheric chemistry of Earth, Mars, Venus, Titan, a generic hydrocarbon-chemistry network for gas giants, and the interstellar medium. For comparison to biological systems, they also included the biochemical network of *E. coli*. This work was re-examined more recently by Wong *et al.* (2023), which conducted a more rigorous analysis, additionally incorporating atmospheres of early Earth and Pluto, as well as several biological networks (see Section 1.3.1 for more detail).

However, unlike Solé and Munteanu (2004) and Wong *et al.* (2023), rather than comparing a single terrestrial planet to say, a Jovian, I will be investigating populations terrestrial planets that could be largely identical in terms of atmospheric composition (Affholder *et al.*, 2021). Thus, in order to differentiate between worlds in this population, I must include additional information in the network about the atmospheric chemistry as a system beyond its composition.

One option is to weight the reaction edges by the rate (defined as the temperature-dependent rate constant in the form of an Arrhenius equation, multiplied by the concentration of the reactants) of the reaction that the edge represents. These edge weights would represent the net flux between any two connected species–that is, the rate at which a specific reactant is being converted to a given product. Understanding how fluxes changes in response to different initial conditions would allow us to better understand the patterns of behavior in the chemistry within planetary atmospheres,

and compare worlds with otherwise similar atmospheres, particularly in terms of disequilibrium chemistry (Fisher *et al.*, 2022). For example, this would allow us to more easily determine if the behavior of the atmospheric chemistry is dominated by a few species or reactions, or if it is determined in a more emergent or diffuse manner. Given the dramatic role life can play in affecting a planet's atmosphere (Lovelock and Margulis, 1974; Kasting and Siefert, 2002; Krissansen-Totton *et al.*, 2016), being able to gauge what factors are controlling the atmospheric chemistry is critical.

2.3 Methods

2.3.1 Analysis of a Simple Network

As an example for how similar atmospheric CRNs might be analyzed and compared, let us start with a simple set of chemical reactions, each with a reaction rate of R_n, calculated as the reaction rate constant in terms of molecules of product s^{-1} cm^{-3}, multiplied by the molecules of reactant:

$$4H_2 + CO_2 \rightarrow CH_4 + 2H_2O, R_1 = 4.168 \tag{2.1}$$

$$CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O, R_2 = 283.1 \tag{2.2}$$

$$2H_2O + h\nu \rightarrow O_2 + 2H_2, R_3 = 0.00789 \tag{2.3}$$

After assigning each species present a node and connecting them reactant to product to construct a network, I can then increase the usefulness of the network by weighting the edges with the net reaction rate, here defined as the reaction rate constant of the reaction multiplied by the concentration of the reactants. For example, the edge between H_2 and H_2O would incorporate the net rate calculated from the rate of conversion of H_2 to H_2O from $R_1 = 4.168$ in Equation 2.1, and the conversion of

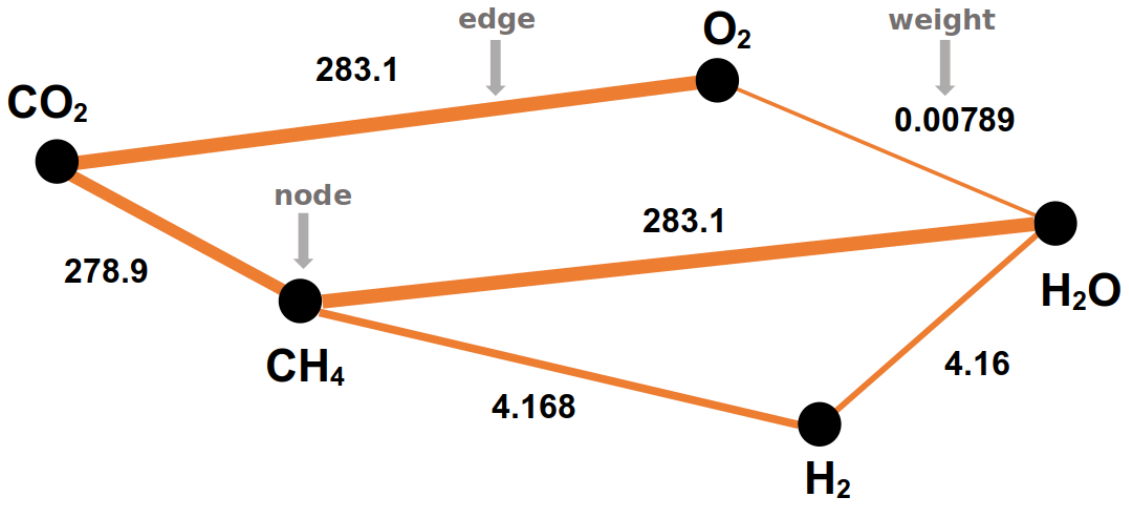


Figure 2.1: A simple reaction network. The weight of the edges is determined by calculating the net rate of the reactions the edge represents.

H_2O to H_2 from $R_3 = 0.00789$ in Equation 2.3, to give an edge weight of 4.168-0.00789 = 4.16 (see Figure 2.1)

Networks can be represented as *unipartite*, in which there is only one type of node present (in this case, those representing the reactants and products in the chemical system), or *bipartite*, where two types of nodes are present, with the second representing, for example, the reactions themselves in addition to the nodes representing the reactants and products. In the interest of simplicity, I will only be using unipartite network representations. This by no means is to discount bipartite networks; for more complex biospheres, bipartite networks could be useful, as biological reactions could be represented using the second set of nodes.

Having constructed this weighted unipartite network, I can then analyze its topology using a variety of network metrics (see Table 2.1). These metrics can be measured using network-theoretic software packages, such as the NetworkX package for Python (Hagberg *et al.*, 2008) or Cytoscape (Shannon *et al.*, 2003).

One measurement is the *degree*, which is one of the most frequently used network

Table 2.1: Network measures used in this study with their definitions, notation, physical interpretation, and values from the example network shown in Figure 2.1

Notation	Network measurement	Description in the context of CRNs	Physical interpretation	Value in example network
$\langle k \rangle$	Mean degree (Equation 2.5)	Average of the total flux driven by each chemical compound across the network	High weighted mean degree indicates that the chemical system is dominated by highly reactive species	459
$\langle \ell \rangle$	Average shortest path length (Equation 2.7)	Average of the minimum amounts of flux along dependency pathways between every pair of compounds	Fewest reactions required to convert one species to another in a given direction of flow, as weighted by flux	87.0
$\langle C \rangle$	Average clustering coefficient (Equation 2.9)	The average tendency of a chemical compound to have neighbor compounds that participate in the same chemical reaction.	How often species tend to co-participate in reactions with the same reactants	0.052
$\langle k_{nn} \rangle$	Average neighbor degree (Equation 2.10)	Average of the total flux driven by the neighbors of each chemical compound, averaged across the network	High average neighbor degree indicates that reactive species in the system co-participate in many reactions	1.60
$\langle g(v) \rangle$	Average node betweenness centrality (Equation 2.11)	The average tendency of a compound to be part of a shortest (flux-weighted reaction) path between every pair of compounds.	Which species play a key role in multiple reaction pathways	0.133
$\langle g(e) \rangle$	Average edge betweenness centrality (Equation 2.12)	The average tendency of a reaction to be part of a shortest (flux-weighted reaction) path between every pair of compounds	Which reactions play a key role in multiple reaction pathways	0.120

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 as a reactant. However, for a weighted network like the atmospheric network models I study here, the degree of node i, k_i is the *sum of weights* of the edges connected to i:

$$k_i = \sum_{j \in V} a_{ij} w_{ij} \tag{2.4}$$

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 its all 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:

$$\langle k \rangle = \frac{1}{N} \sum_{i \in V} k_i \tag{2.5}$$

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

In the case of the example network in Figure 2.1, the weighted mean degree is 459. This demonstrates that mean degree is very sensitive to the presence of reactions with high rate values–even if they're comparatively few in a number within a reaction network, such reactions can strongly influence the mean degree of the CRN.

The *shortest path length* from i to j for a weighted network is defined as

$$\ell(i,j) = \min_{\gamma(i,j) \in \Gamma(i,j)} \left[\sum_{u,v \in \gamma(i,j)} w_{uv} \right] \tag{2.6}$$

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 and 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 I use here, the path should instead be thought as the dependence amongst chemical species in a series of reactions where species i and j are participants in the end member reaction in the set each. Thus, 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

$$\langle \ell \rangle = \frac{1}{N(N-1)} \sum_{i,j \in V} \ell(i,j) \tag{2.7}$$

and this can be considered as the average minimal amount of flux along a dependency pathway between every pair of compounds in the network. In the example network, the weighted average shortest path length is 87.0. Like mean degree, the reaction rate-weighted average shortest pathway in atmospheric CRN tends to be sensitive to reactions with high rate values.

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

$$C_u = \frac{1}{d_u\left(d_u - 1\right)} \sum_{vw} \left(\hat{w}_{uv}\hat{w}_{uw}\hat{w}_{vw}\right)^{1/3} \tag{2.8}$$

where d_u is the number of edges connected to a node u, and $\hat{w}_{uv} = w_{uv}/max(w)$ is normalized by the maximum weight in the network (Onnela *et al.*, 2005). C_u is assigned to 0 for $d_u < 2$. A group of nodes with high clustering coefficient will be tightly knit. Average clustering coefficient (Barabási and Pósfai, 2016),

$$\langle C \rangle = \frac{1}{N} \sum_{i \in V} C_i \tag{2.9}$$

is the average of clustering coefficients over all nodes and quantifies the tendency of interdependence between two chemical compounds sharing a neighbor compound. In the case of the example network, the weighted average clustering coefficient is 0.052– unlike mean degree and average shortest path length, average clustering coefficient is better at capturing the fluxes across the entire network, and is less likely to be influenced by a few high-rate reactions.

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

$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in N(i)} w_{ij} d_j \tag{2.10}$$

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 as each other. In this paper, I 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.

In the case of the example network, the average neighbor degree for the network is 1.60. In this case, the network is too small to be able to infer anything about its assortativity from this value.

Node betweenness centrality is defined as:

$$g(v) = \sum_{s,t \in v} \frac{\sigma(s,t \mid v)}{\sigma(s,t)} \tag{2.11}$$

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 $v \sum s,t, \sigma(s,t|v) = 0$. This measure can quantify the influence of a given compound on the 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. I calculated the average node betweenness, $\langle g(v) \rangle$, 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.

The average node betweenness centrality for the example network is 0.133. Like average clustering coefficient, this measure reflects the flux values across the entire network.

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

$$g(e) = \sum_{s,t \in v} \frac{\sigma(s,t \mid e)}{\sigma(s,t)} \tag{2.12}$$

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. I 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. The average edge betweenness centrality of the example network is 0.120, again indicative of the flux values across the network, much like average node betweenness centrality.

2.3.2 CRN Analysis of Populations of Planetary Atmospheres

Having applied network measures to a simple CRN, I turn to using the same approach on large populations of planetary atmospheres. As a demonstration, consider a group of 500 terrestrial worlds, each having an atmosphere resembling the early Archean Earth, but without the presence of life. I model these worlds using the atmos package (Arney *et al.*, 2016) combined with the PyATMOS Python wrapper that allows running large batches of atmos in parallel (Bella *et al.*, 2018).

Atmos is a combined photochemical and climate Fortran model developed by Arney *et al.* (2016) based off of work by Kasting *et al.* (1979) and Zahnle *et al.* (2006),

though for the purposes of atmospheric CRN analysis, I will only use the photochemical portion. The 1-dimensional photochemistry model is divided into 200 plane parallel layers, stretching from the surface to an altitude of 100 km, with layers spaced at 0.5 km intervals. The mixing ratio of each species is found by solving flux and mass continuity equations in each layer (in the form of a Jacobian matrix) simultaneously using a reverse-Euler method Butcher (2016), providing exact solutions at steady state. The Jacobian is solved self-consistently at each time step–that is, the redox fluxes of gases entering and exiting the atmosphere sum to zero (Zahnle *et al.*, 2006). The model includes 76 species, though only 66 of them are considered long-lived and included in the Jacobian. Vertical transport by molecular and eddy diffusion is included, and boundary conditions that drive the model can be set for each species at the surface and the top of the atmosphere.

Conveniently, atmos comes pre-equipped with a model of the Archean Earth atmosphere (see Table 2.2 for initial conditions); I will use this template largely unchanged, varying only the surface flux of CH_4, to represent CH_4 produced from the reaction of CO_2 with H_2 generated from the transformation of ferromagnesian minerals into serpentinite via water-rock reactions (Thompson *et al.*, 2022). I will draw these surface flux values from a normal distribution ranging from $\sim 10^{-2}$ to 10^8 molecules/cm^2/second, with a mean of 10^6 molecules/cm^2/second (based off of Etiope and Sherwood Lollar, 2013's estimates for the abiotic production CH_4 on Earth) and a variance of 3.9. The model contains 492 reactions, and includes sophisticated methods to handle the formation of haze and aerosols; see Arney *et al.* (2016) for more details and a full list of reactions and their rate constants. Atmos also comes with several pre-set insolation profiles, to represent the planet's star; I will be using the one for Gliese 576, a red dwarf, due to the fact it's thought to be easier to observe terrestrial exoplanets around red dwarfs with current and near-future technology (Fu-

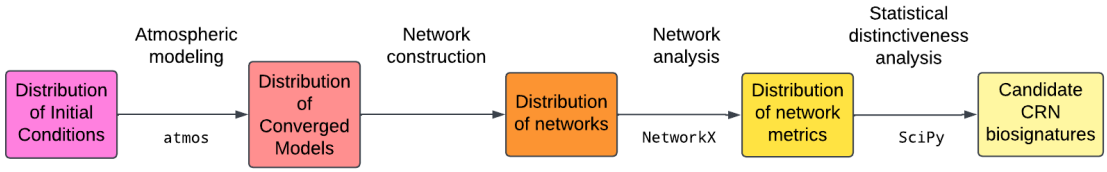


Figure 2.2: Planetary atmospheres and used to create chemical reaction networks. The metrics of these networks are then analyzed and investigated for potential biosignature candidates.

jii *et al.*, 2018; Lincowski *et al.*, 2019; Arney, 2019). However, this does mean that photolytic reactions are entirely absent.

Once these models have converged, I then run them through an analysis pipeline that extracts the reactions and their rates from the atmos output files; while which reactions are present is fixed at the start of the model, the reaction rates changes as function of the abundance of the reactants. These reactions and rates are used to build chemical reaction networks with rate-weighted edges (see Figure 2.2). These networks are then analyzed using NetworkX, yielding a population of network metric values, each network with its own mean degree, average shortest path length, etc (see Figure 2.3).

Table 2.2: Initial conditions for all long-lived species present in the atmos template for Archean Earth-like atmospheres used in this example

Species	Deposition velocity (molecules/cm^2/s)	Constant mixing ratio (molecules/cm^2/s)	Constant upward flux (molecules/cm^2/s)
O	1.0	1.00×10^{-8}	n/a
O_2	0	n/a	n/a

H_2O	0	n/a	n/a
H	1.0	n/a	n/a
OH	1.0	n/a	n/a
HO_2	1.0	n/a	n/a
H_2O_2	2.0×10^{-1}	n/a	n/a
H_2	2.4×10^{-4}	n/a	1.000×10^{10}
CO	1.2×10^{-4}	n/a	n/a
HCO	1.0	n/a	n/a
H_2CO	2.0×10^{-1}	n/a	n/a
CH_4	0	n/a	10^{-2}-10^{8}
CH_3	1.0	n/a	n/a
C_2H_6	0	n/a	n/a
NO	3.0×10^{-4}	n/a	n/a
NO_2	3.0×10^{-3}	n/a	n/a
HNO	1.0	n/a	n/a
O_3	7.0×10^{-2}	n/a	n/a
HNO_3	2.0×10^{-1}	n/a	n/a
N	0	n/a	n/a
C_3H_2	0	n/a	n/a
C_3H_3	0	n/a	n/a
CH_3C_2H	0	n/a	n/a
CH_2CCH_2	0	n/a	n/a
C_3H_5	0	n/a	n/a
C_2H_5CHO	0	n/a	n/a

C_3H_6	0	n/a	n/a
C_3H_7	0	n/a	n/a
C_3H_8	0	n/a	n/a
C_2H_4OH	0	n/a	n/a
C_2H_2OH	0	n/a	n/a
C_2H_5	0	n/a	n/a
C_2H_4	0	n/a	n/a
CH	0	n/a	n/a
CH_3O_2	0	n/a	n/a
CH_3O	0	n/a	n/a
CH_2CO	0	n/a	n/a
CH_3CO	0	n/a	n/a
CH_3CHO	0	n/a	n/a
CH_23	0	n/a	n/a
C_2H	0	n/a	n/a
C_2	0	n/a	n/a
C_2H_3	0	n/a	n/a
HCS	0	n/a	n/a
CS_2	0	n/a	n/a
CS	0	n/a	n/a
OCS	0	n/a	n/a
S	0	n/a	n/a
HS	0	n/a	n/a
H_2S	2.0×10^{-2}	n/a	3.5000×10^{8}

SO_3	0	n/a	n/a
S_2	0	n/a	n/a
HSO	1.0	n/a	n/a
H_2SO_4	1.0	n/a	n/a
SO_2	1.0	n/a	3.5000×10^9
SO	0	n/a	n/a
CO_2	0	2.00×10^{-2}	n/a
SO_4 aerosol	1.0×10^{-2}	n/a	n/a
S_8 aerosol	1.0×10^{-2}	n/a	n/a
hydrocarbon aerosol 1	1.0×10^{-2}	n/a	n/a
hydrocarbon aerosol 2	1.0×10^{-2}	n/a	n/a

To make statistical analysis easier, I can convert these populations into probability distributions. This can be done, for example, by converting the population into a kernel density estimate using SciPy's *gaussian_kde* function (Virtanen *et al.*, 2020) and evaluating the area under the resulting curve using the minimum and maximum values of the population as bounds. This produces a probability distribution for every value in the population, normalized to between 1 and 0 (see Figure 2.4).

To compare this distribution to one that might result from a population of life-bearing worlds, I model another 500 terrestrial Archean Earth-like planets using atmos, but this time with two significant changes. First, I add a reaction representing the biological production of CH_4 to the reaction list of the models. Because of how the reaction list is formatted in atmos, only two reactants are allowed per reaction,

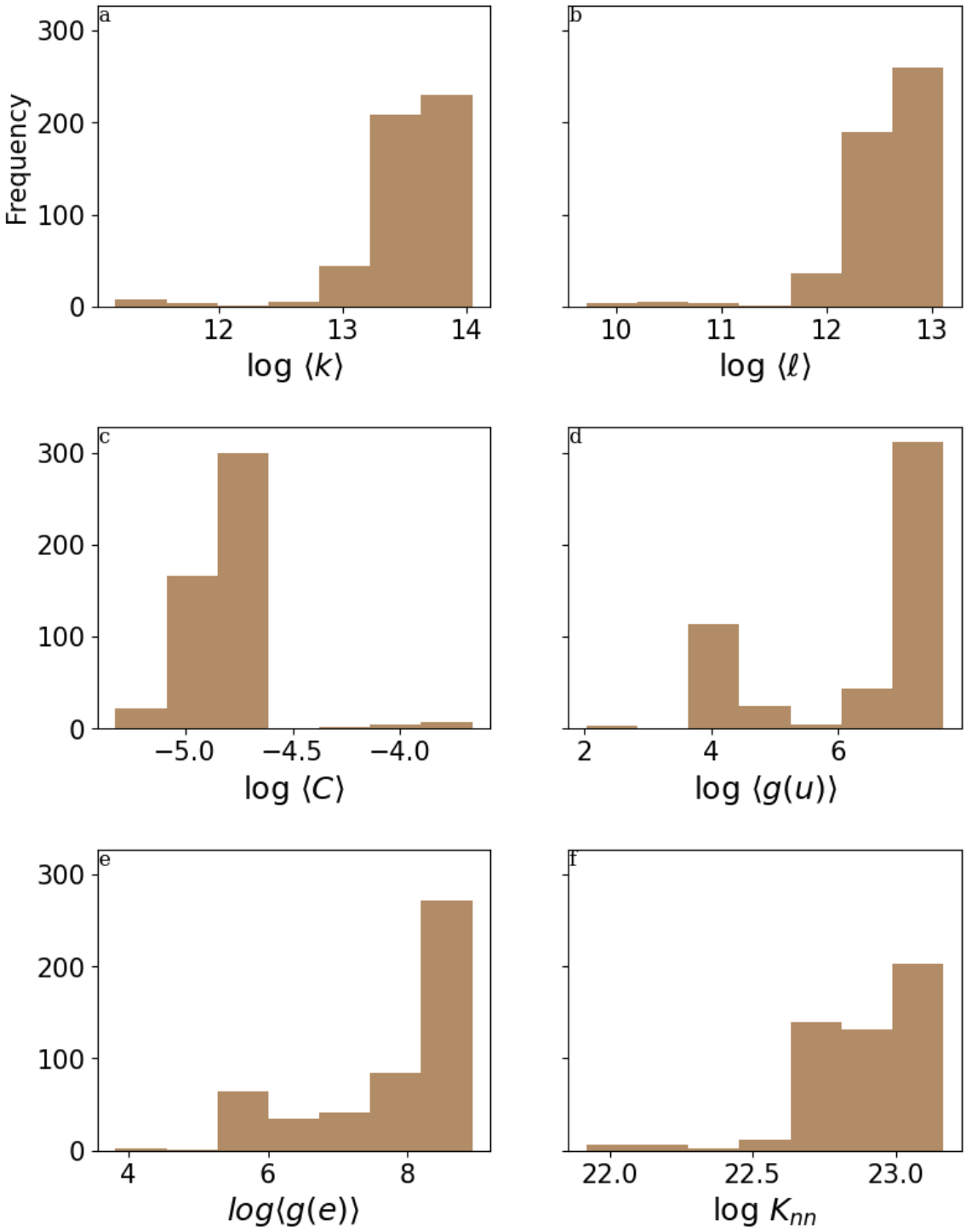


Figure 2.3: Frequency distributions of network metrics of a population of abiotic worlds: a) mean degree b) average shortest path length c) average clustering coefficient d) node betweenness centrality e) edge betweenness centrality f) average neighbor degree.

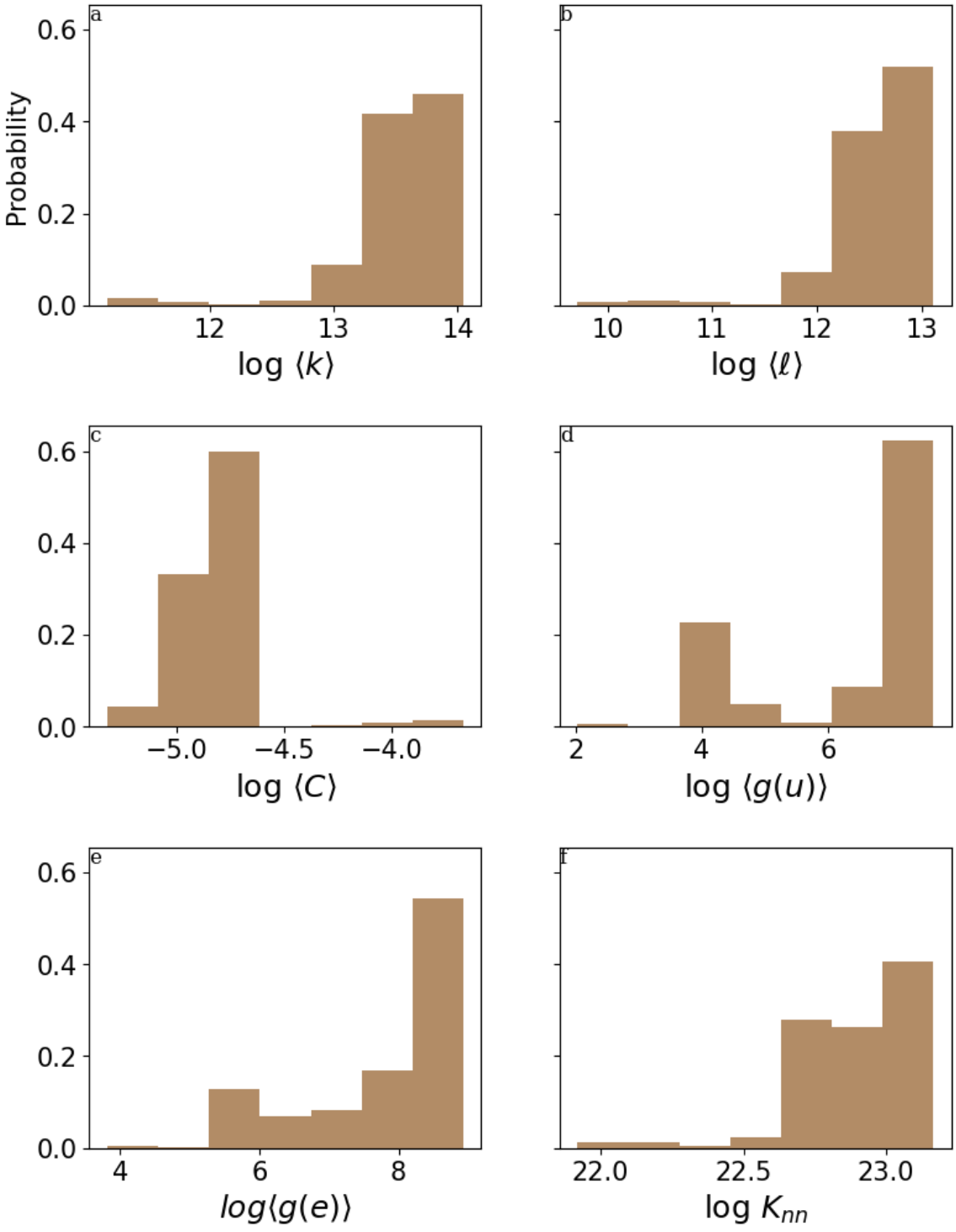


Figure 2.4: Probability distributions of a population of abiotic worlds: a) mean degree b) average shortest path length c) average clustering coefficient d) node betweenness centrality e) edge betweenness centrality f) average neighbor degree.

without stoichiometric coefficients. Consequently, the methanogenesis reaction

$$CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O \tag{2.13}$$

in the second set of models was included in the CRN as a three-step reaction, involving several fictitious short-lived intermediate species:

$$H_2 + H_2 \rightarrow \text{`}H_4\text{'} \tag{2.14}$$

$$\text{`}H_4\text{'} + CO_2 \rightarrow \text{`}H_4C_2\text{'} \tag{2.15}$$

$$\text{`}H_4\text{'} + \text{`}H_4CO_2\text{'} \rightarrow CH_4 + H_2O + H_2O \tag{2.16}$$

These fictitious intermediates are considered to be short-lived species, which are modeled separately in atmos than long-lived species, being excluded from the model's Jacobian. While they are included in the chemical reaction network, their effect on the network topology is expected to be minimal due to their short lifespan.

Each step in this series of reactions had the same rate constant modeled by an Arrhenius equation, using rate constants for anaerobic microbes in the form of

$$m_E = 3.3 * e^{\frac{-6.94\times10^4}{R} - \frac{1}{T} - \frac{1}{298}} \tag{2.17}$$

(Tijhuis *et al.*, 1993; Seager *et al.*, 2013)

Where R is the universal gas constant 8.314 J/K/mole, T is the temperature in Kelvin, -6.94×10^4 is the activation energy in J/mol and m_E refers to the maintenance energy required to sustain the population of methanogenic microbes in J/g biomass/s.

Secondly, I increase the surface flux of CH_4, to represent its now much-higher rate from biological production. This is done by drawing the flux values from a normal distribution ranging from $\sim10^3$ to 10^{13} molecules/cm^2/second, with a mean of 10^{11} molecules/cm^2/second (from Catling and Zahnle, 2020) and a variance of 3.9.

Once the biotic atmosphere models have converged, they can then be run through the same pipeline as the abiotic models, resulting in a population of network values (see Figure 2.5) and a probability distribution of network values (see Figure 2.6) which can be compared to that of the abiotic worlds.

2.4 Results and Discussion

The resulting distributions can be compared to each other to determine if network topology can help distinguish between populations of abiotic and biotic worlds (see Figure 2.7). However, the point of using CRN metrics is to be able to more robustly differentiate between these populations; as such, the distinguishability must be quantifiable.

2.4.1 Quantifying Distinguishability Between Distributions

The distinguishability between these distributions can be quantified using a variety of methods. One is the percentage of overlap shared between the distributions; less overlap indicates network metric values that easier to distinguish from each other, and, potentially, candidate biosignatures. This, in turn, lowers the risk of false positives and negatives. Examining the abiotic and biotic network metric distributions, I find that several topological measurements, such as mean degree and average shortest path length, have 25% or less overlap, suggesting they may be good biosignature candidates (see Figure 2.8) due to a lower risk of false positives and negatives.

Another method is the Kolmorogov-Smirnoff metric (Massey, 1951) or K-S test, which measures the distance between two distributions. 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 zero and one representing the maximum

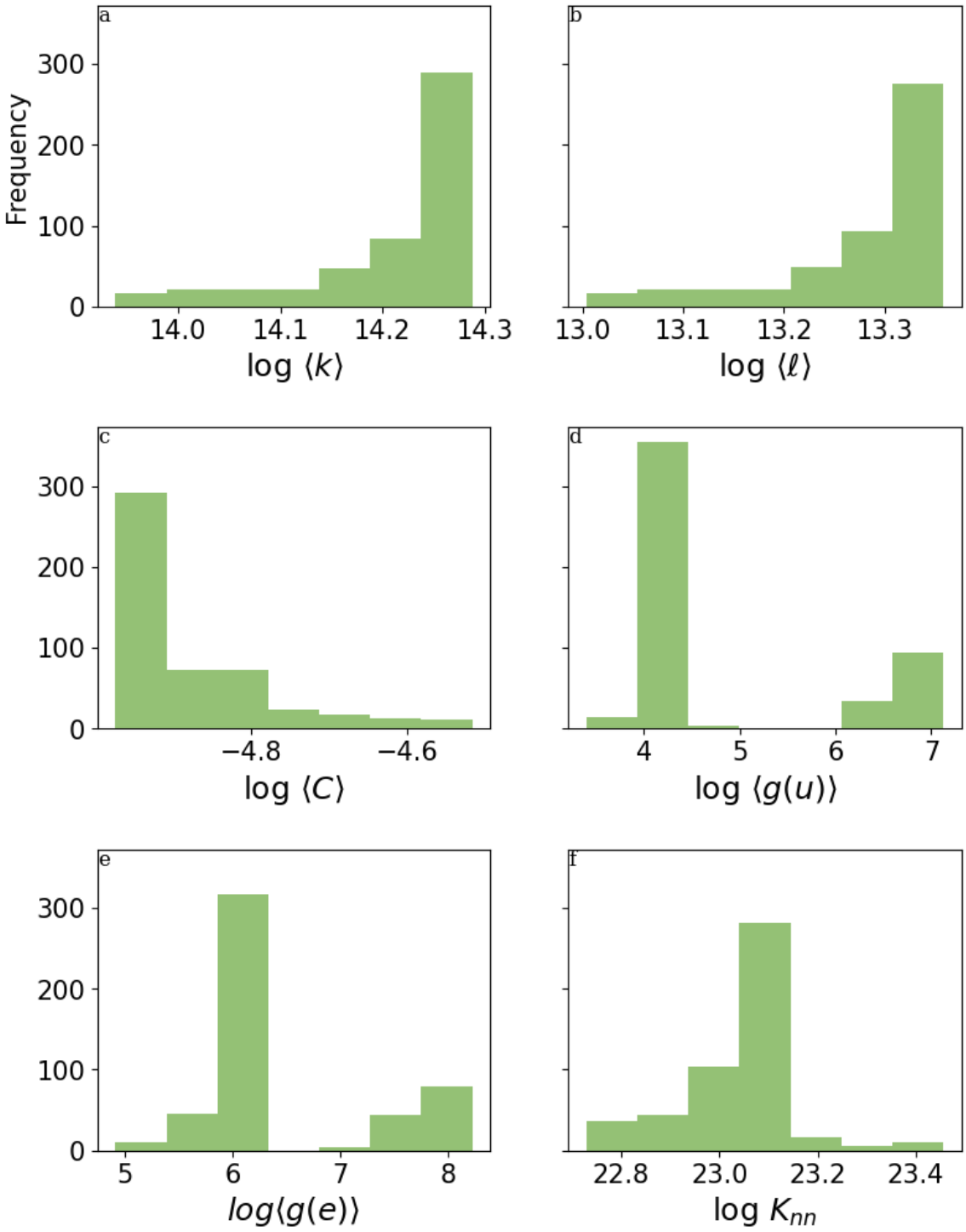


Figure 2.5: Frequency distributions of network metrics of a population of biotic worlds: a) mean degree b) average shortest path length c) average clustering coefficient d) node betweenness centrality e) edge betweenness centrality f) average neighbor degree.

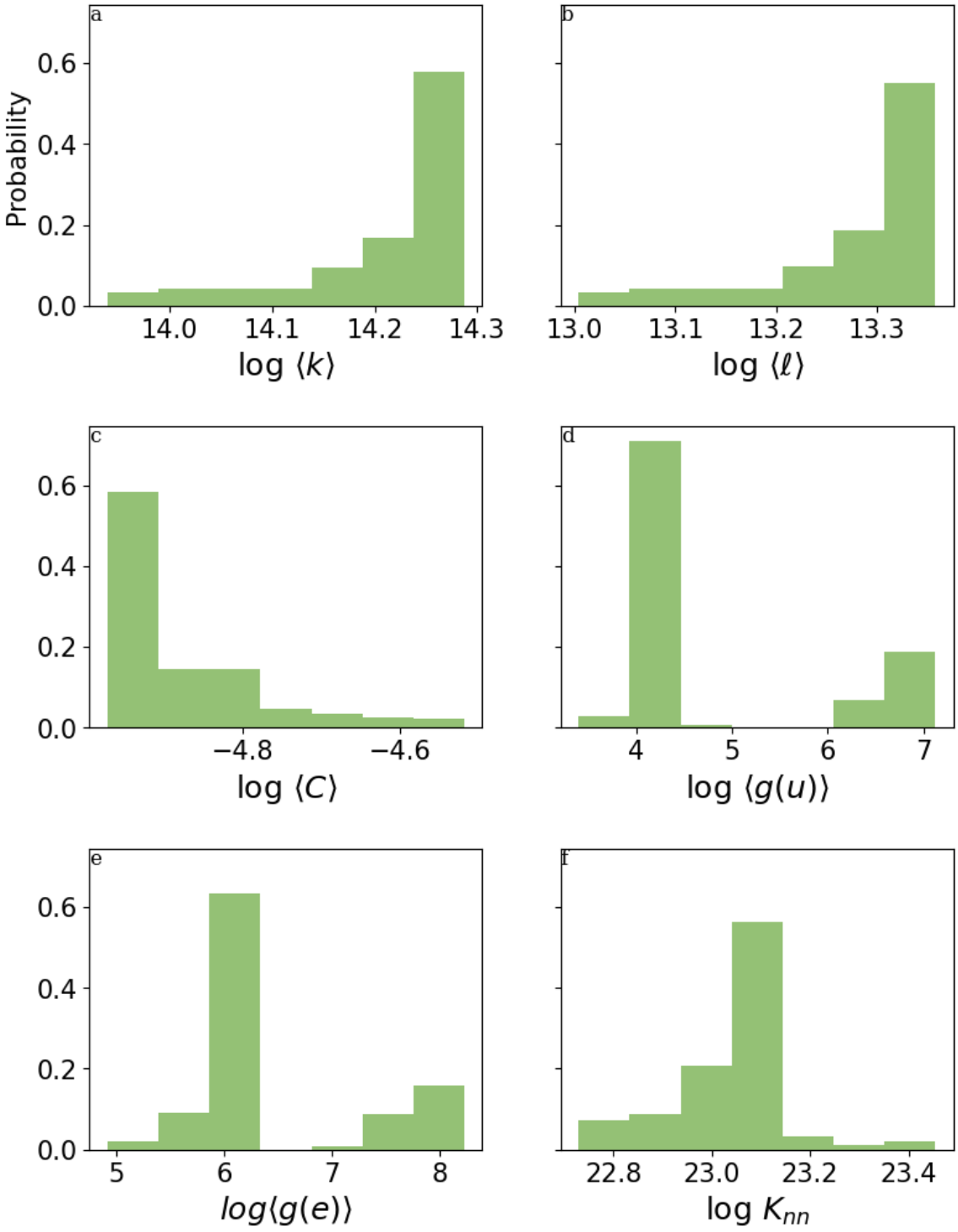


Figure 2.6: Probability distributions of a population of biotic worlds: a) mean degree b) average shortest path length c) average clustering coefficient d) node betweenness centrality e) edge betweenness centrality f) average neighbor degree.

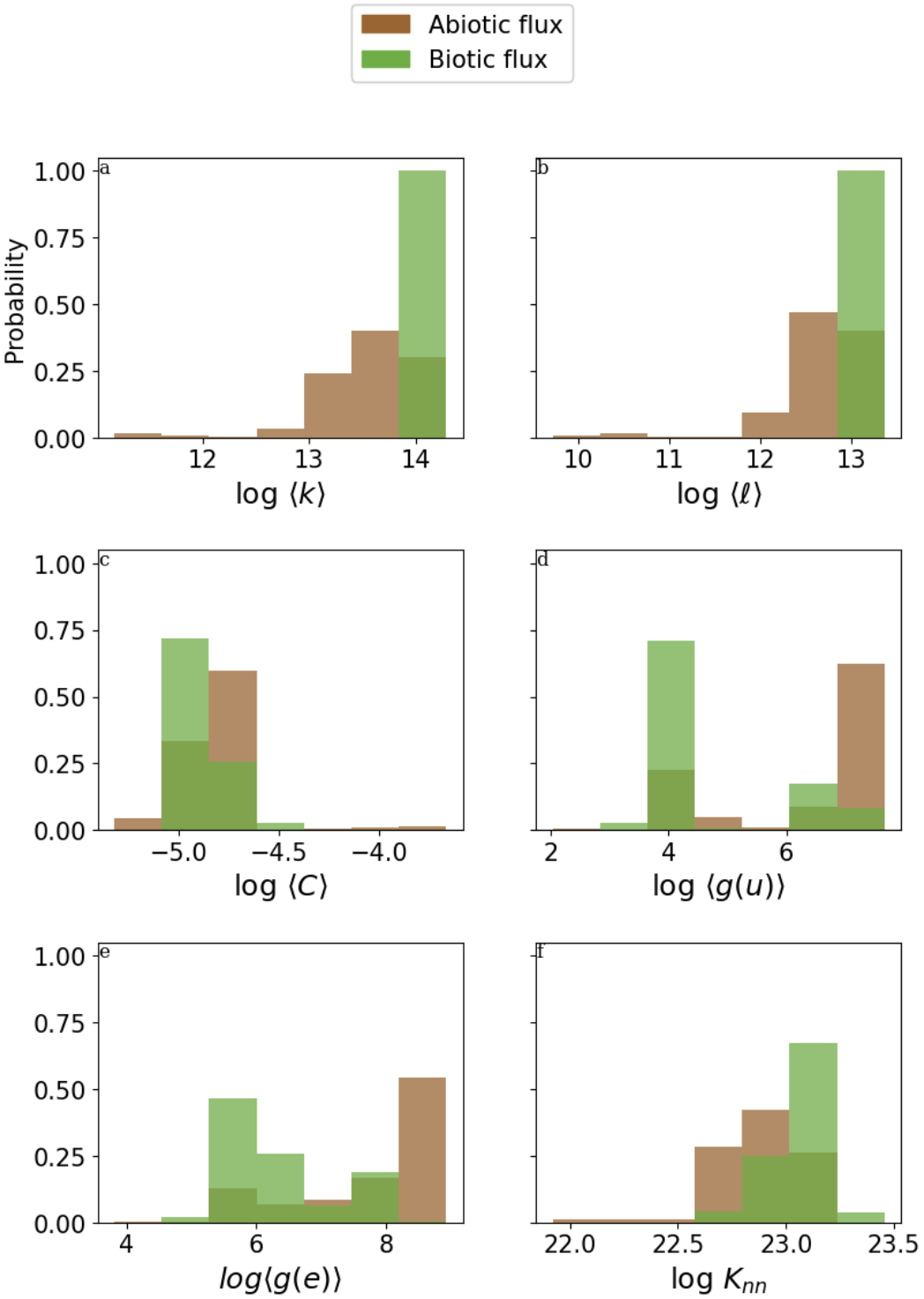


Figure 2.7: A comparison of the probability distributions of network metrics from populations of abiotic (brown) and biotic (green) worlds: a) mean degree b) average shortest path length c) average clustering coefficient d) node betweenness centrality e) edge betweenness centrality f) average neighbor degree.

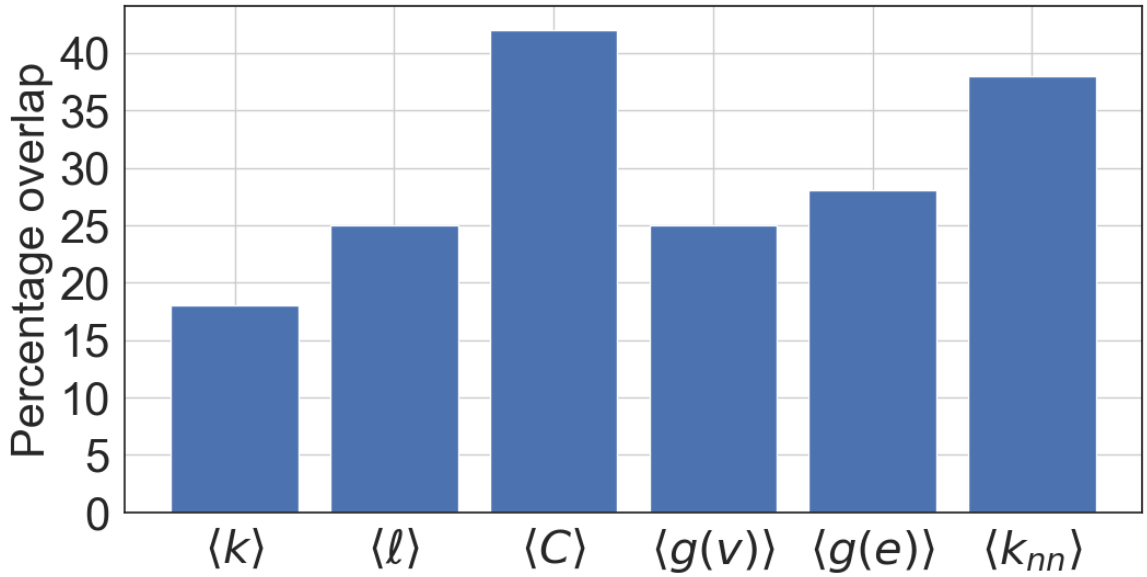


Figure 2.8: Mean degree, average clustering coefficient, and node betweenness centrality all have comparatively low percentages of overlap between abiotic and biotic distributions, suggesting they may be potentially useful as biosignatures.

distance between two cumulative distribution functions, along with its associated p-value indicating the statistical significance of the distance value. The distance is given as

$$D_{n,m} = \sup_x \mid F_{1,n}(x) - F_{2,m}(x) \tag{2.18}$$

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 biosignature candidates. The threshold for which a K-S value might be considered large is a function of the size of the distribution and the p-value yielded by the K-S test; for our distributions with an n=500 and p-values of approximately zero, that puts the threshold at around 0.011 (see Figure 2.9)

When applied to the two distributions using the *ks_2samp* function from SciPy.stats (Virtanen *et al.*, 2020), I find that mean degree and average shortest path length have

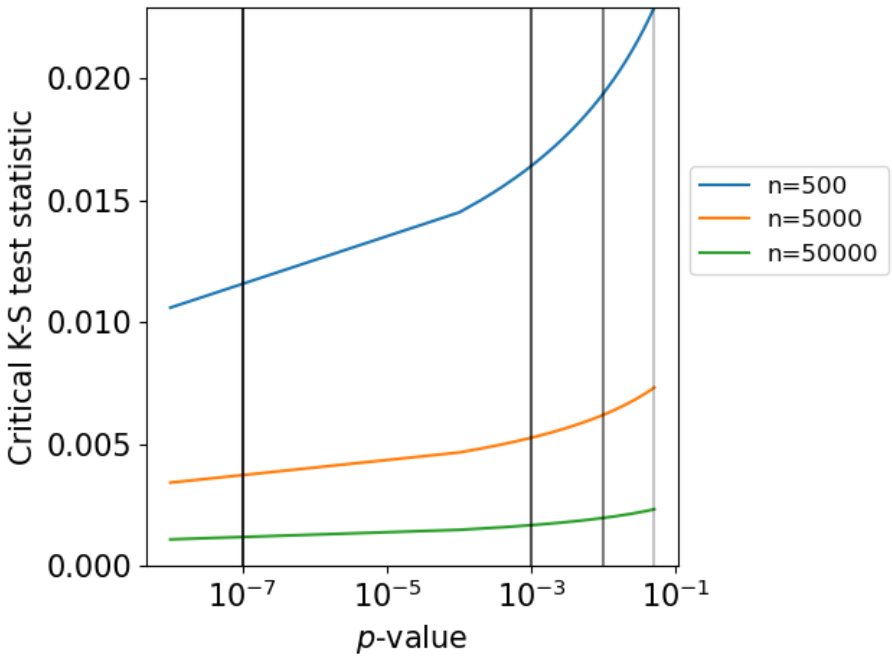


Figure 2.9: The threshold for a K-S test statistic to be considered significant for the distributions with n=500 and a p-value below 10^{-7} is 0.011.

the highest K-S value (see Figure 2.10), of 0.92; as average shortest path length and mean degree are both sensitive to reactions with high rates, it isn't surprising that they exhibit similar trends and values when measured over the same population of networks. All distributions were found to have a p-value of approximately zero, indicating high statistical significance.

2.4.2 Validating Results Using Perturbation Testing

While these results are exciting, I must confirm that the network metrics observed actually correspond in some way to the physical reality of the modeled atmospheres. As models of Archean-like atmospheres used in atmos have already been demonstrated to be self-consistent (Zahnle *et al.*, 2006) and ground-truthing is impossible, reliability tests such as split-half testing would not necessarily provide useful information (Welbanks *et al.*, 2023). However, validation can be accomplished via perturbation

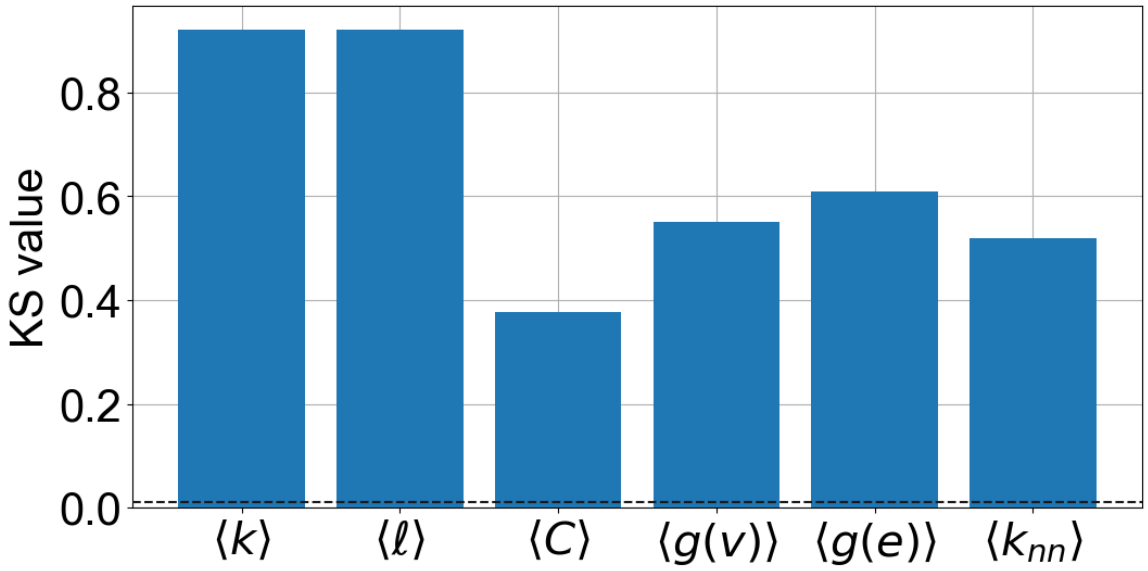


Figure 2.10: Mean degree and average shortest path length exhibit high K-S values when comparing probability distributions of network metrics from populations of abiotic and biotic worlds, well above the threshold of 0.011 (dotted line).

testing, where the atmosphere is significantly altered (by removing a key species or reaction pathway, for example), and then the resulting network metrics are compared to the unperturbed model.

For the population of Archean Earth-like worlds I'm investigating, I can perturb them easily by removing the species C_2H_2 and its associated reactions. C_2H_2 plays a central role in providing a sink for CH_4 via aerosol formation, *e.g.*,

$$C_2 + CH_4 \rightarrow C_2H + CH_3 \tag{2.19}$$

$$C_2H + C_2H_2 \rightarrow \text{Aerosol} + H \tag{2.20}$$

and thus removing it should lead to noticeable changes in the behavior of the atmospheric chemistry of both the abiotic and biotic populations of worlds.

Perturbing the atmospheres in this way does result in changes in the network metric probability distributions of the abiotic world population(see Figure 2.11), with

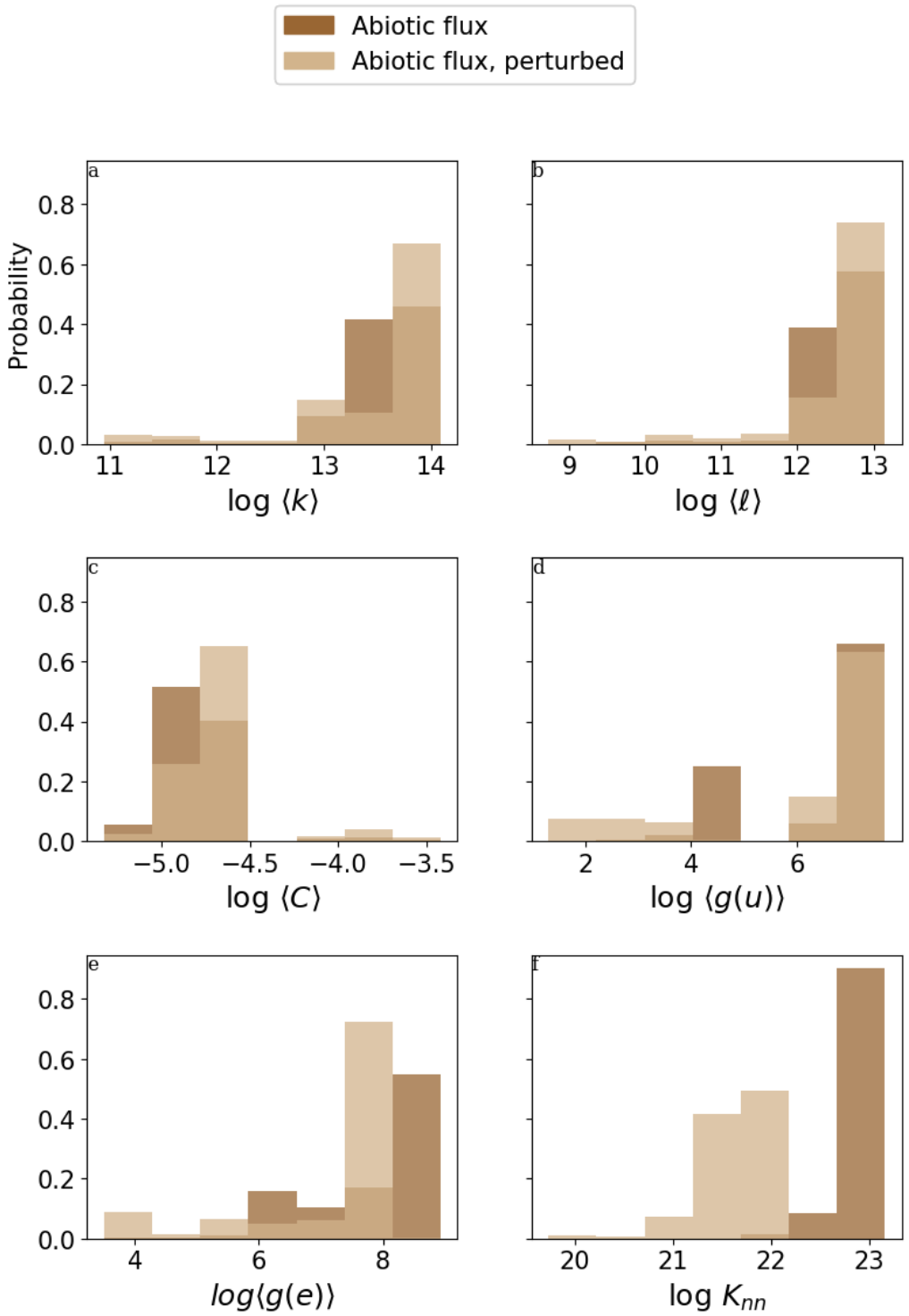


Figure 2.11: The probability distributions of network metrics from perturbed and non-perturbed abiotic worlds are noticeably distinguishable, providing confidence that network metrics are indicative of physical and chemical behavior within the modeled atmospheres.

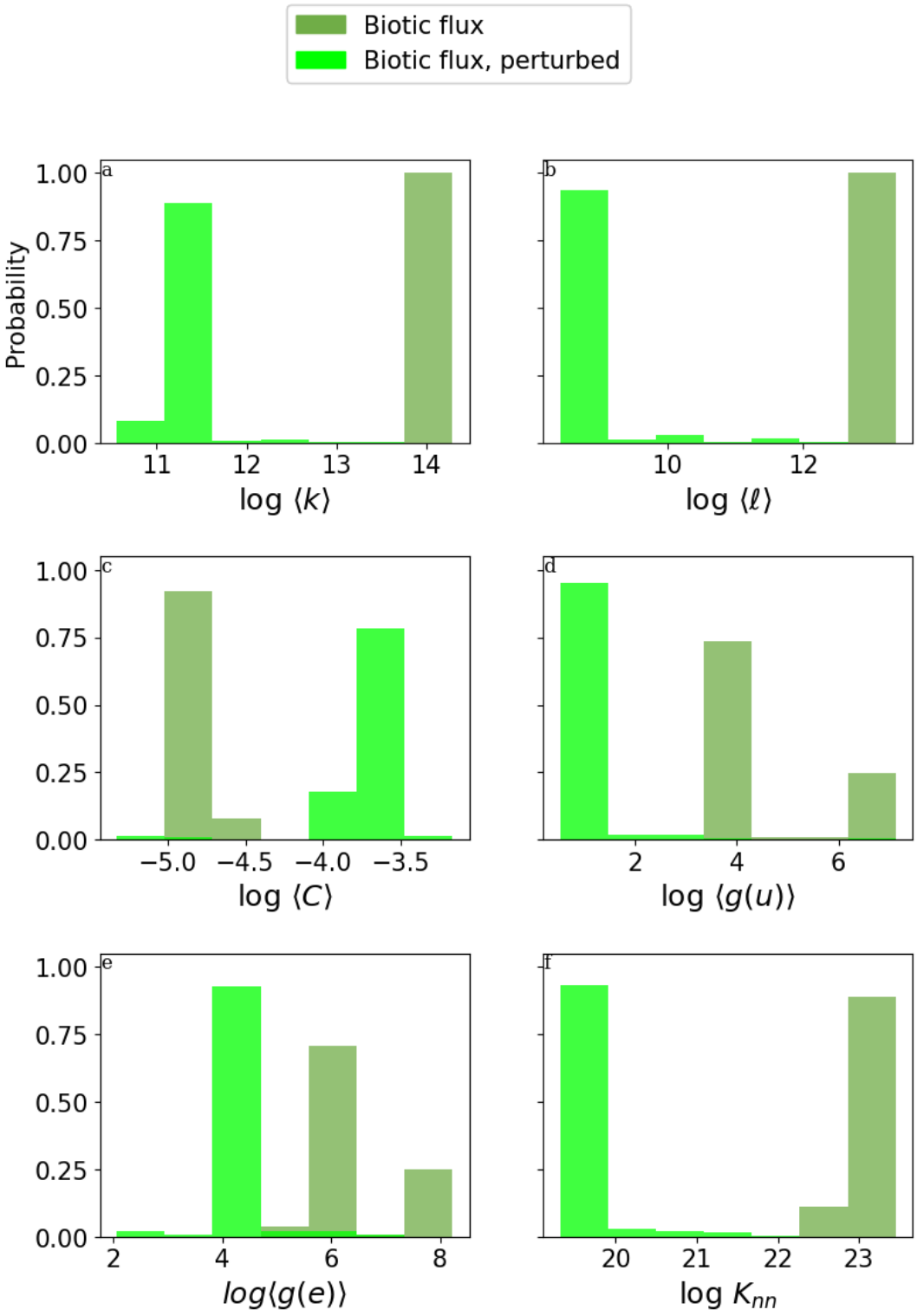


Figure 2.12: The network metric probability distributions from perturbed and non-perturbed biotic atmospheres are even more noticeably distinct than those of perturbed and non-perturbed abiotic atmospheres

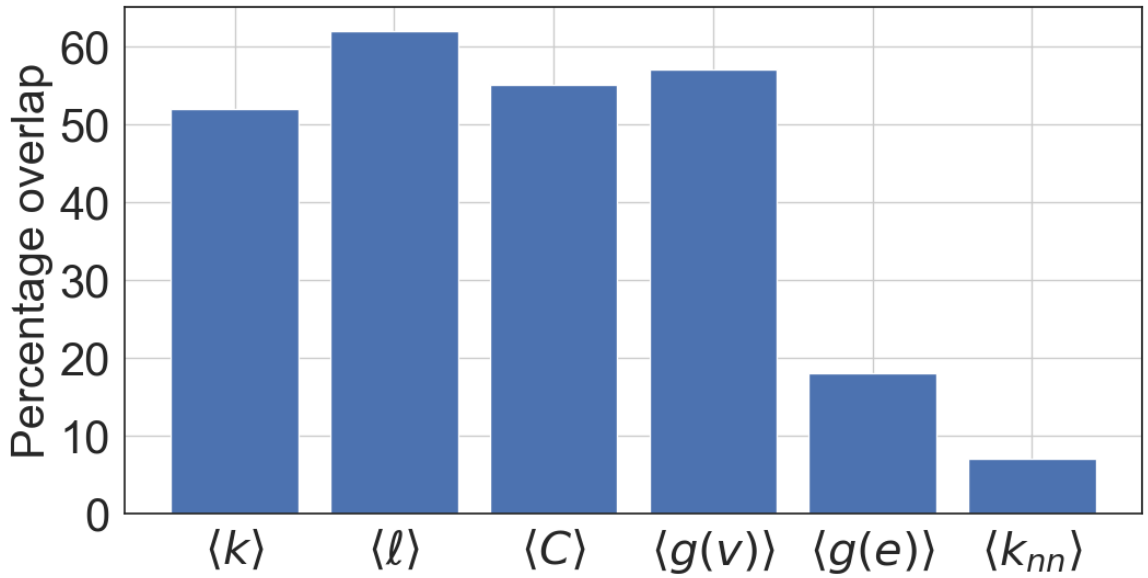


Figure 2.13: Perturbed and non-perturbed abiotic network metric distributions do not wholly overlap–in some cases, the percentage overlap is as low as 7%. This increases confidence that the network metrics are reflective of the physical and chemical behavior of the atmospheres.

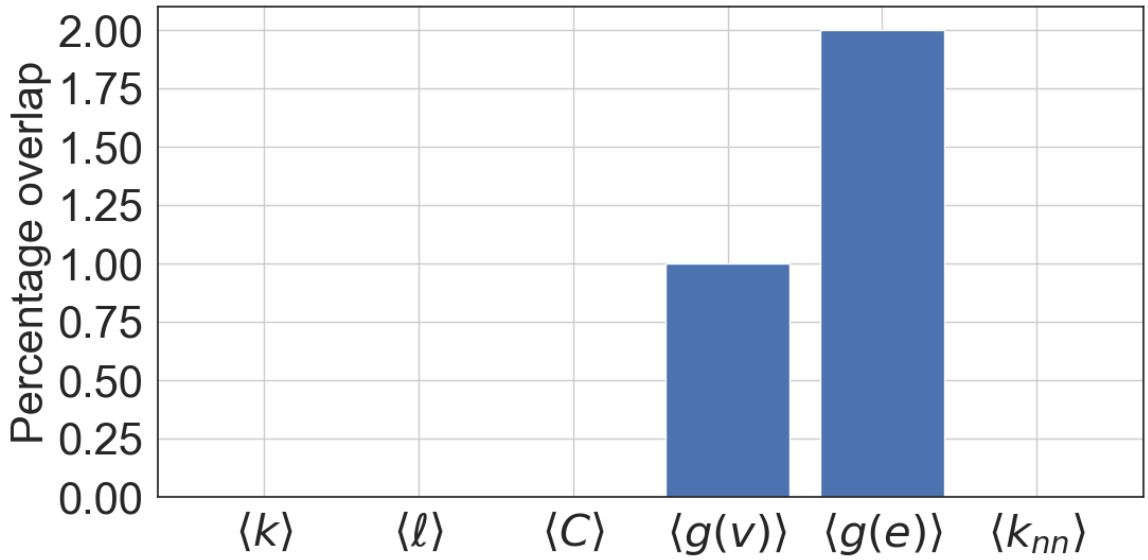


Figure 2.14: Perturbed and non-perturbed biotic network metric distributions are either completely non-overlapping, or show very low overlap. This further increase confidence that the network metrics are reflective of the physical and chemical behavior of the atmospheres.

distribution overlap dropping as low as 7% in the case of average neighbor degree (see Figure 2.13). The effect is even more striking when comparing the distributions from perturbed and non-perturbed biotic worlds (see Figure 2.12), likely due to the increase in the concentration of CH_4 in the atmospheres; several distributions are completely non-overlapping, and the rest only exhibit a very small overlap percentage (see Figure 2.14). This provides confidence that the network metrics are indeed reflective of the physical and chemical behavior within the modeled atmospheres, including the potential effects of a biosphere interfacing with an atmosphere; otherwise, the network metrics would have remained unchanging in the face of perturbation.

2.5 Conclusion

As I have shown, I can distinguish between large populations of abiotic and biotic plantary atmospheres using the topology of their chemical reaction networks. The use of network topology thus provides a systems-level alternative to looking for single "smoking gun", one that may provide more unambiguously differentiated between biotic and abiotic signals.

Furthermore, atmospheric CRN topology not only provides potential biosignatures, it can also help us to better understand the chemical behavior of the atmospheres. For example, the higher range of mean degree and average shortest path length in response to adding biological methane production to the Archean Earth-like models suggests that the rates of the reactions in the network have increased. This is likely due in turn to the massive uptick in CH_4 surface flux, and the addition of another set of reactions that produce CH_4.

Thus, network topology offers a powerful tool for those seeking to better understand the chemical and physical dynamics of exoplanet atmospheres.

REFERENCES

Affholder, A., A. Bixel, B. Sauterey, R. Ferrière, D. Apai and S. Mazevet, "Predicting quantitative biosignature patterns from populations of exoplanets", in "AGU Fall Meeting Abstracts", vol. 2021, pp. P55D–1969 (2021).

Amaral, L. A. and J. M. Ottino, "Complex networks: Augmenting the framework for the study of complex systems", The European Physical Journal B 38, 147–162 (2004).

Angeli, D., "A tutorial on chemical reaction networks dynamics", in "2009 European Control Conference (ECC)", pp. 649–657 (IEEE, 2009).

Antoniou, I. E. and E. T. Tsompa, "Statistical analysis of weighted networks", Discrete Dynamics in Nature and Society 2008, e375452 (2008).

Arney, G., S. D. Domagal-Goldman, V. S. Meadows, E. T. Wolf, E. Schwieterman, B. Charnay, M. Claire, E. Hébrard and M. G. Trainer, "The pale orange dot: the spectrum and habitability of hazy Archean Earth", Astrobiology 16, 11, 873–899 (2016).

Arney, G. N., "The k dwarf advantage for biosignatures on directly imaged exoplanets", The Astrophysical Journal Letters 873, 1, L7 (2019).

Barabási, A.-L. and M. Pósfai, *Network Science* (Cambridge University Press, Cambridge, United Kingdom, 2016).

Barrat, A., M. Barthélemy, R. Pastor-Satorras and A. Vespignani, "The architecture of complex weighted networks", Proceedings of the National Academy of Sciences 101, 11, 3747–3752 (2004).

Bella, A., A. Choprab, W. Fawcettc and R. Talebid, "NASA Frontier Development Lab technical memorandum", (2018).

Butcher, J. C., *Numerical methods for ordinary differential equations* (John Wiley & Sons, 2016).

Catling, D. C., J. Krissansen-Totton, N. Y. Kiang, D. Crisp, T. D. Robinson, S. DasSarma, A. J. Rushby, A. Del Genio, W. Bains and S. Domagal-Goldman, "Exoplanet biosignatures: a framework for their assessment", Astrobiology 18, 6, 709–738 (2018).

Catling, D. C. and K. J. Zahnle, "The archean atmosphere", Science Advances 6, 9, eaax1420 (2020).

Clarke, B. L., "Graph theoretic approach to the stability analysis of steady state chemical reaction networks", The Journal of Chemical Physics 60, 4, 1481–1492 (1974).

Domagal-Goldman, S. D., A. Segura, M. W. Claire, T. D. Robinson and V. S. Meadows, "Abiotic ozone and oxygen in atmospheres similar to prebiotic Earth", The Astrophysical Journal 792, 2, 90 (2014).

Etiope, G. and B. Sherwood Lollar, "Abiotic methane on earth", Reviews of Geophysics 51, 2, 276–299 (2013).

Fisher, T., H. Kim, C. Millsaps, M. Line and S. I. Walker, "Inferring exoplanet disequilibria with multivariate information in atmospheric reaction networks", The Astronomical Journal 164, 2, 53 (2022).

Fujii, Y., D. Angerhausen, R. Deitrick, S. Domagal-Goldman, J. L. Grenfell, Y. Hori, S. R. Kane, E. Pallé, H. Rauer, N. Siegler, K. Stapelfeldt and K. B. Stevenson, "Exoplanet biosignatures: observational prospects", Astrobiology 18, 6, 739–778 (2018).

Grenfell, J. L., "A review of exoplanetary biosignatures", Physics Reports 713, 1–17 (2017).

Guzmán-Marmolejo, A., A. Segura and E. Escobar-Briones, "Abiotic production of methane in terrestrial planets", Astrobiology 13, 6, 550–559 (2013).

Hagberg, A., P. Swart and D. S Chult, "Exploring network structure, dynamics, and function using NetworkX", Tech. rep., Los Alamos National Lab.(LANL), Los Alamos, NM (United States) (2008).

Harman, C. E. and S. Domagal-Goldman, "Biosignature false positives", in "Handbook of Exoplanets", edited by H. J. Deeg and J. A. Belmonte, pp. 1–22 (Springer International Publishing, Cham, 2018).

Jeong, H., B. Tombor, R. Albert, Z. N. Oltvai and A.-L. Barabási, "The large-scale organization of metabolic networks", Nature 407, 6804, 651–654 (2000).

Kasting, J. F., S. Liu and T. Donahue, "Oxygen levels in the prebiological atmosphere", Journal of Geophysical Research: Oceans 84, C6, 3097–3107 (1979).

Kasting, J. F. and J. L. Siefert, "Life and the evolution of earth's atmosphere", Science 296, 5570, 1066–1068 (2002).

Kleinböhl, A., K. Willacy, A. J. Friedson, P. Chen and M. R. Swain, "Buildup of abiotic oxygen and ozone in moist atmospheres of temperate terrestrial exoplanets and its impact on the spectral fingerprint in transit observations", The Astrophysical Journal 862, 2, 92 (2018).

Krissansen-Totton, J., D. S. Bergsman and D. C. Catling, "On detecting biospheres from chemical thermodynamic disequilibrium in planetary atmospheres", Astrobiology 16, 1, 39–67 (2016).

Lincowski, A. P., J. Lustig-Yaeger and V. S. Meadows, "Observing isotopologue bands in terrestrial exoplanet atmospheres with the James Webb Space Telescope: implications for identifying past atmospheric and ocean loss", The Astronomical Journal 158, 1, 26 (2019).

Lovelock, J. E. and L. Margulis, "Atmospheric homeostasis by and for the biosphere: the Gaia hypothesis", Tellus 26, 1-2, 2–10 (1974).

Massey, F. J., "The Kolmogorov-Smirnov test for goodness of fit", Journal of the American Statistical Association 46, 68–78 (1951).

Meadows, V. S., C. T. Reinhard, G. N. Arney, M. N. Parenteau, E. W. Schwieterman, S. D. Domagal-Goldman, A. P. Lincowski, K. R. Stapelfeldt, H. Rauer, S. DasSarma, S. Hegde, N. Narita, R. Deitrick, J. Lustig-Yaeger, T. W. Lyons, N. Siegler and J. L. Grenfell, "Exoplanet biosignatures: understanding oxygen as a biosignature in the context of its Environment", Astrobiology 18, 6, 630–662 (2018).

Mitchell, M., "Complex systems: Network thinking", Artificial Intelligence 170, 18, 1194–1212 (2006).

Onnela, J.-P., J. Saramäki, J. Kertész and K. Kaski, "Intensity and coherence of motifs in weighted complex networks", Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 71, 6 Pt 2, 065103 (2005).

Rein, H., Y. Fujii and D. S. Spiegel, "Some inconvenient truths about biosignatures involving two chemical species on Earth-like exoplanets", Proceedings of the National Academy of Sciences 111, 19, 6871–6875 (2014).

Reinhard, C. T., S. L. Olson, E. W. Schwieterman and T. W. Lyons, "False negatives for remote life detection on ocean-bearing planets: lessons from the early Earth", Astrobiology 17, 4, 287–297 (2017).

Sandora, M. and J. Silk, "Biosignature surveys to exoplanet yields and beyond", Monthly Notices of the Royal Astronomical Society 495, 1, 1000–1015 (2020).

Saxena, P., "Photobombing earth 2.0: diffraction-limit-related contamination and uncertainty in habitable planet spectra", The Astrophysical Journal Letters 934, 2, L32 (2022).

Seager, S., W. Bains and R. Hu, "A biomass-based model to estimate the plausibility of exoplanet biosignature gases", The Astrophysical Journal 775, 2, 104 (2013).

Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, "Cytoscape: a software environment for integrated models of biomolecular interaction networks", Genome Research 13, 11, 2498–2504 (2003).

Smith, H. B., H. Kim and S. I. Walker, "Scarcity of scale-free topology is universal across biochemical networks", Sci. Rep. 11, 1, 6542 (2021).

Solé, R. V. and A. Munteanu, "The large-scale organization of chemical reaction networks in astrophysics", Europhysics Letters (EPL) 68, 2, 170–176 (2004).

Temkin, O. N., A. V. Zeigarnik and D. Bonchev, *Chemical Reaction Networks: A Graph-Theoretical Approach* (CRC Press, 1996).

Thompson, M. A., J. Krissansen-Totton, N. Wogan, M. Telus and J. J. Fortney, "The case and context for atmospheric methane as an exoplanet biosignature", Proceedings of the National Academy of Sciences 119, 14, e2117933119 (2022).

Tijhuis, L., M. C. Van Loosdrecht and J. Heijnen, "A thermodynamically based correlation for maintenance gibbs energy requirements in aerobic and anaerobic chemotrophic growth", Biotechnology and Bioengineering 42, 4, 509–519 (1993).

Virtanen, P., R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt and SciPy 1.0 Contributors, "SciPy 1.0: fundamental algorithms for scientific computing in Python", Nature Methods 17, 261–272 (2020).

Walker, S. I., W. Bains, L. Cronin, S. DasSarma, S. Danielache, S. Domagal-Goldman, B. Kacar, N. Y. Kiang, A. Lenardic, C. T. Reinhard, W. Moore, E. W. Schwieterman, E. L. Shkolnik and H. B. Smith, "Exoplanet biosignatures: future directions", Astrobiology 18, 6, 779–824 (2018).

Walker, S. I., L. Cronin, A. Drew, S. Domagal-Goldman, T. Fisher, M. Line and C. Millsaps, "Probabilistic biosignature frameworks", in "Planetary Astrobiology", p. 477 (University of Arizona Press, 2020).

Welbanks, L., P. McGill, M. Line and N. Madhusudhan, "On the application of bayesian leave-one-out cross-validation to exoplanet atmospheric analysis", The Astronomical Journal 165, 3, 112 (2023).

Wong, M. L., A. Prabhu, J. Williams, S. M. Morrison and R. M. Hazen, "Toward network-based planetary biosignatures: atmospheric chemistry as unipartite, unweighted, undirected networks", Journal of Geophysical Research: Planets 128, 6, e2022JE007658 (2023).

Wordsworth, R. and R. Pierrehumbert, "Abiotic oxygen-dominated atmospheres on terrestrial habitable zone planets", The Astrophysical Journal Letters 785, 2, L20 (2014).

Zahnle, K., M. Claire and D. Catling, "The loss of mass-independent fractionation in sulfur due to a palaeoproterozoic collapse of atmospheric methane", Geobiology 4, 4, 271–283 (2006).

Chapter 3

INFERRING EXOPLANET DISEQUILIBRIA WITH MULTIVARIATE INFORMATION IN ATMOSPHERIC REACTION NETWORKS

3.1 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.

3.2 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 (Fujii *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 and Seager, 2012; Line *et al.*, 2015; Madhusudhan, 2019; Madhusudhan *et al.*, 2011). 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-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 (Centler and 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.*, 2016);(Krissansen-Totton *et al.*, 2019; Line *et al.*, 2013; Molaverdikhani *et al.*, 2019; Simoncini *et al.*, 2013). Network-theoretic and thermodynamic measures have been proposed as possible can-

didate 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 and 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 *et al.*, 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.

3.3 Methods

3.3.1 Atmospheric Modeling with VULCAN

We used the VULCAN modeling package (Tsai *et al.*, 2017) to simulate the H-C-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 *et al.*, 2014), 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

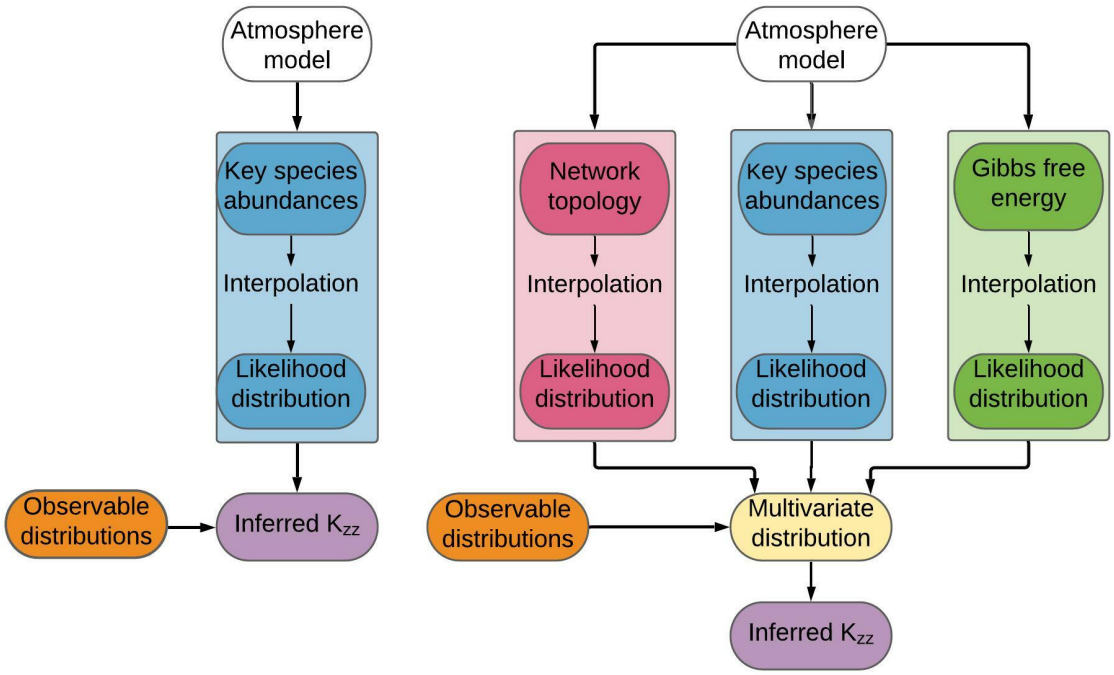


Figure 3.1: Left: "standard" analysis pipeline used to infer atmospheric K_{zz} by generating a likelihood distribution from the abundance of key species. Right: schematic of the analysis pipeline used for the multivariate approach to identify optimal predictors. Inverse modeling, based on observational 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.

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

$$\frac{\partial n_i}{\partial t} = p_i - L_i - \frac{\partial f}{\partial z} \tag{3.1}$$

where n_i equals the number density of the ith 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 ith species ($cm^{-3}s^{-1}$), respectively. The transport flux, f can be written as:

$$f = -K_{zz} n_{total} \frac{\partial X_i}{\partial z} \tag{3.2}$$

where X_i represents the mixing ratio of the ith 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).

3.3.2 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 and Douglas, 2010) and are distinguishable from the CRNs of biological systems (Jolley and Douglas, 2011). 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 CH4 and O2 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 $\langle C \rangle$ (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 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_{zz} are shown in Figure 3. Some of the edges in the network appear thicker

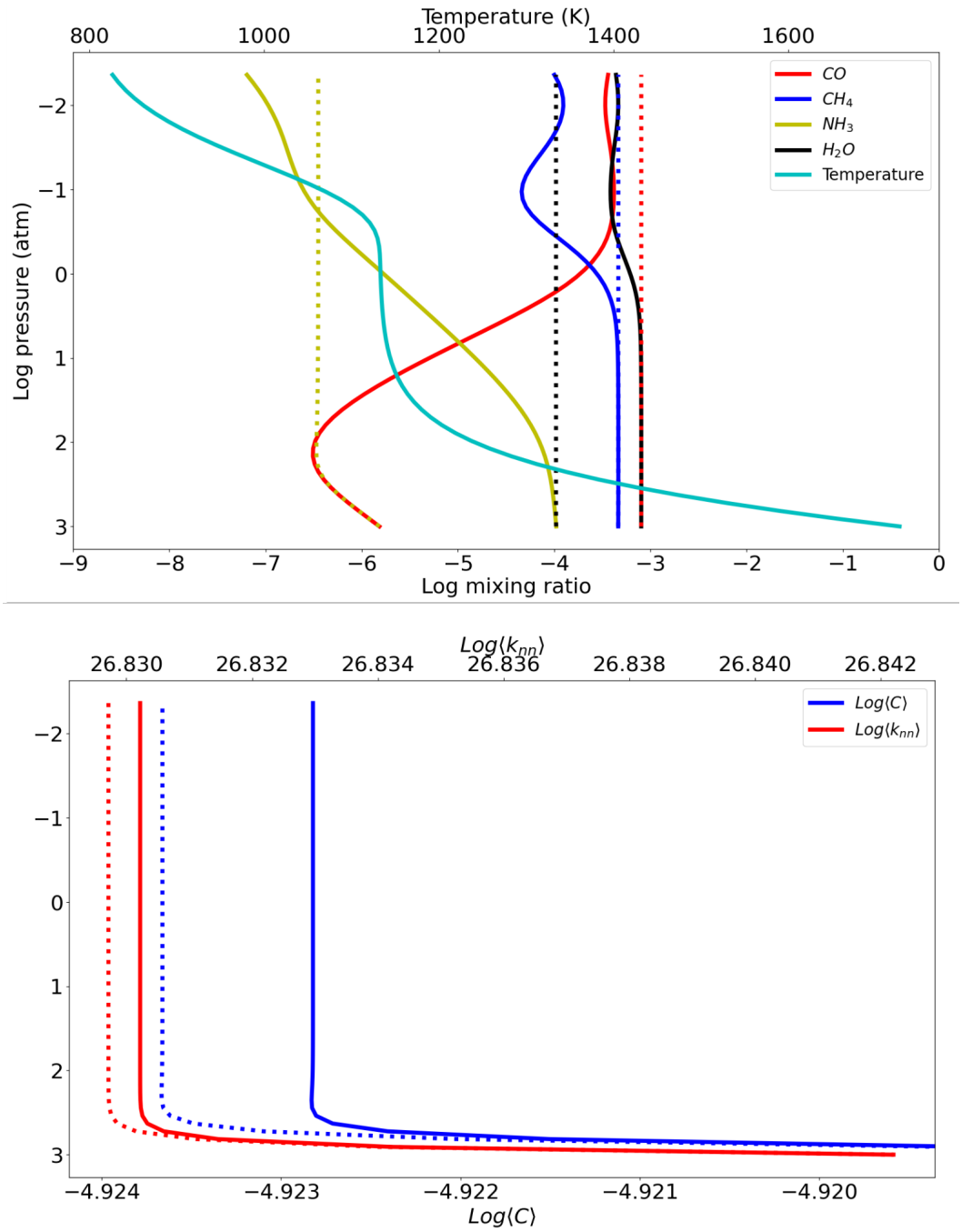


Figure 3.2: Solid lines represent the atmosphere at K_{zz} of zero, dashed lines an atmosphere at K_{zz} of 10^{10}. Top: the abundance of key molecular species as a function of pressure, at 1200 K and 1 solar metallicity. Bottom: network topological measures—average clustering coefficient, and average 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.

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 to the edges (because more than two species typically participate in any given reaction, a single reaction can correspond to more than one edge in this network projection, where each edge associated to 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 and Barabási (2002) and Newman (2003, 2010) for general cases and in Antoniou and 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:

$$k_i = \sum_{j \in V} a_{ij} w_{ij} \tag{3.3}$$

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 its all 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:

$$\langle k \rangle = \frac{1}{N} \sum_{i \in V} k_i \tag{3.4}$$

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

$$\ell(i,j) = \min_{\gamma(i,j) \in \Gamma(i,j)} \left[\sum_{u,v \in \gamma(i,j)} w_{uv} \right] \tag{3.5}$$

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 and Tsompa, 2008). Note 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 dependence amongst chemical species in a series of reactions where species i and j are participants in the end member reaction in the set each. 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

$$\langle \ell \rangle = \frac{1}{N(N-1)} \sum_{i,j \in V} \ell(i,j) \tag{3.6}$$

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:

$$C_u = \frac{1}{d_u \left(d_u - 1\right)} \sum_{vw} \left(\hat{w}_{uv} \hat{w}_{uw} \hat{w}_{vw}\right)^{1/3} \tag{3.7}$$

where d_u is the number of edges connected to a node u, and $\hat{w}_{uv} = w_{uv}/max(w)$ is normalized by the maximum weight in the network (Onnela *et al.*, 2005). C_u is assigned to 0 for $d_u < 2$. A group of nodes with high clustering coefficient will be tightly knit. Average clustering coefficient (Barabási and Pósfai, 2016),

$$\langle C \rangle = \frac{1}{N} \sum_{i \in V} C_i \tag{3.8}$$

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

$$k_{nn,i} = \frac{1}{k_i} \sum_{j \in N(i)} w_{ij} d_j \tag{3.9}$$

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:

$$g(v) = \sum_{s,t \in v} \frac{\sigma(s,t \mid v)}{\sigma(s,t)} \tag{3.10}$$

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 $v \sum s,t, \sigma(s,t|v) = 0$. 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, $\langle g(v) \rangle$, 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 it represents the number of shortest paths that transverse through edge, e, instead of a node:

$$g(e) = \sum_{s,t \in v} \frac{\sigma(s,t \mid e)}{\sigma(s,t)} \tag{3.11}$$

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.

3.3.3 Thermodynamic Analysis

A number of different thermodynamically-motivated measures of atmospheric disequilibria have been proposed (e.g., Line *et al.*, 2013; Krissansen-Totton *et al.*, 2016, 2018). Among these, the measure developed by Krissansen-Totton *et al.* (2016, 2018) has been widely discussed due 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_{zz}). Initially, we measured the thermochemical disequilibria by calculating the total available Gibbs free energy, or Φ, for each atmospheric model following Krissansen-Totton et al. (Krissansen-Totton *et al.*, 2016). To calculate Φ requires summing the absolute Gibbs free energy of all species at a given temperature, T and pressure, P_r:

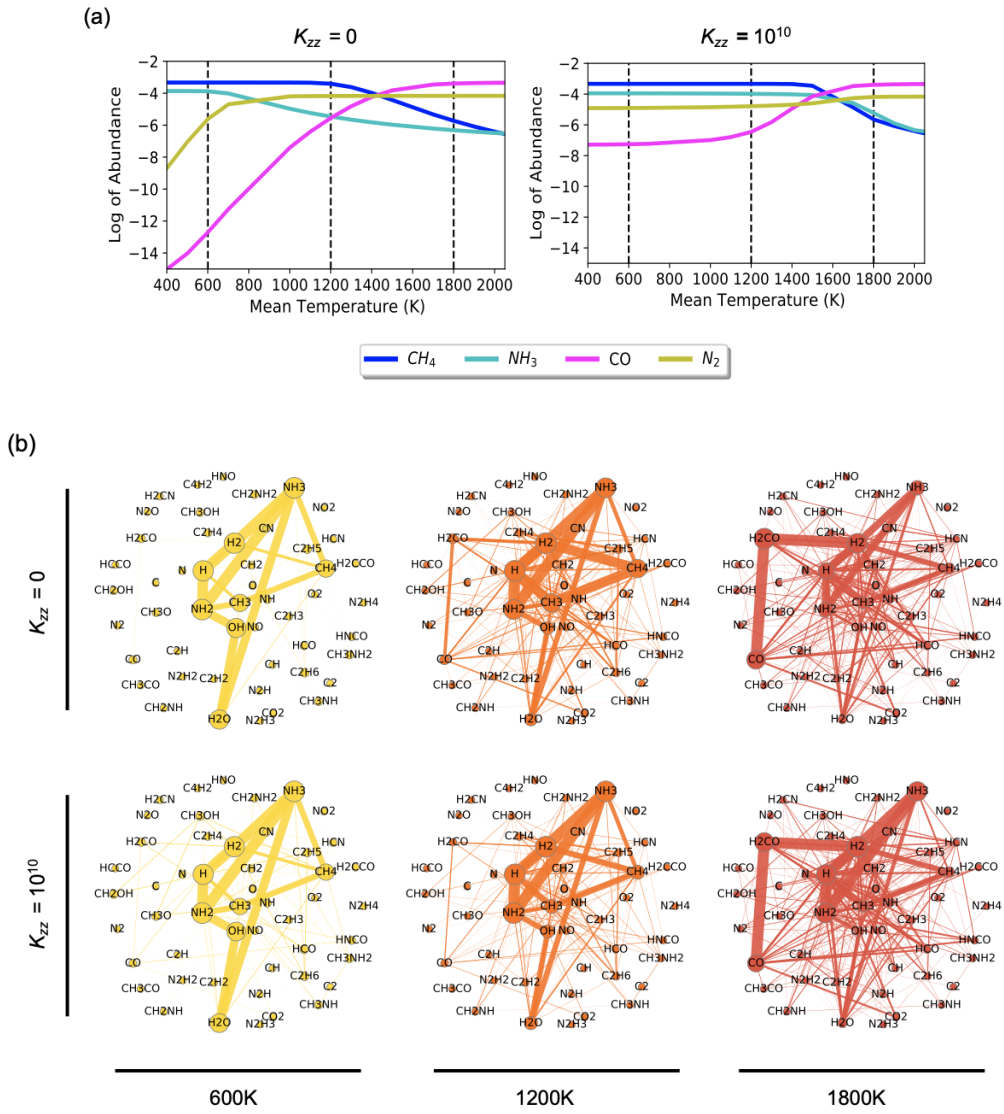


Figure 3.3: (a) The abundance of CH_4, CO, NH_3, and N_2 as a function of the temperature of the atmosphere at constant pressure of 100 mbar, demonstrating the transition from abundant reduced species (CH_4 and NH_3) to abundant oxidized species (CO and N_2) as temperatures increase. Increasing the K_{zz} 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}.

$$G_{(T,P)} = \sum_i n_i \left(G^0_{i(T,P_r}\right) + RT \ln\left(a_i\right) = \sum_i n_i \left(G^0_{i(T,P)}\right) + RT \ln\left(\frac{P n_i \gamma_{fi}}{n_T}\right) \quad (3.12)$$

where $\left(G^0_{i(T,P_r}\right)$ is the standard energy of formation for the ith species at temperature T and pressure P_r, P is standard pressure (1 bar), and n_i and f_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 (273.15K) 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 Φ, but is simpler to calculate.

To determine Φ, the total Gibbs free energy in Eq. (2.12) was calculated based on the mole fraction n_i of each species present at equilibrium ($K_{zz} = 0$), and then compared to the mole fraction calculated from simulated atmospheric data for each model in our grid, producing the calculated value of Φ for that model atmosphere, which represents the available Gibbs free energy in the atmosphere. A higher value of Φ indicates that the atmosphere is further away from equilibrium. Φ is given by the equation:

$$\Phi = \Delta G_{(T,P)}\left(n_i\right) - \Delta G_{(T,P)}\left(\bar{n}_i\right) \quad (3.13)$$

By construction, Φ, 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 Φ 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 $K_{zz} = 0 = 0$ 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 we could not make direct comparisons between the utility of Φ and other atmospheric variables because the former includes an arbitrary normalization, whereas the other measures do not. We therefore elected to use $\Delta G_{(T,P)}$, here after 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 Φ directly), because this did not require us to define a non-observable equilibrium case for comparison.

3.3.4 Generating Statistical Distributions via Interpolation of Chemical Species Abundances, Network Topology, and Gibbs 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 multi-dimensional 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 equilibrium temperature and metallicity input into

the VULCAN code. That is, when given a set of planetary parameters, for example, an equilibrium temperature distribution with a T_{mean} = 1200K and metallicity distribution of 0.5_{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 equilibrium temperature, metallicity, and pressure values. The centers of these distributions started at 400K, then increased at 100K increments until 2000K, 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 3.5 and 3.6, 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 ±50K to ±1000K were incorporated to generate the distributions. The griddata package cannot interpolate values outside the original range given (i.e., here between 400 and 3000K for the mean temperature of the normal distributions); therefore many of the conditions drawn from the ±1000K 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 towards the boundary of the data, it is not expected this will significantly impact the conclusions of our analysis.

3.4 Results

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

A key goal of our analyses was to determine if 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 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-Smirnoff test, or 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 zero and one representing the maximum distance between two cumulative distribution functions, along with its associated p value. The distance is given as

$$D_{n,m} = \sup_x \mid F_{1,n}(x) - F_{2,m}(x) \tag{3.14}$$

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

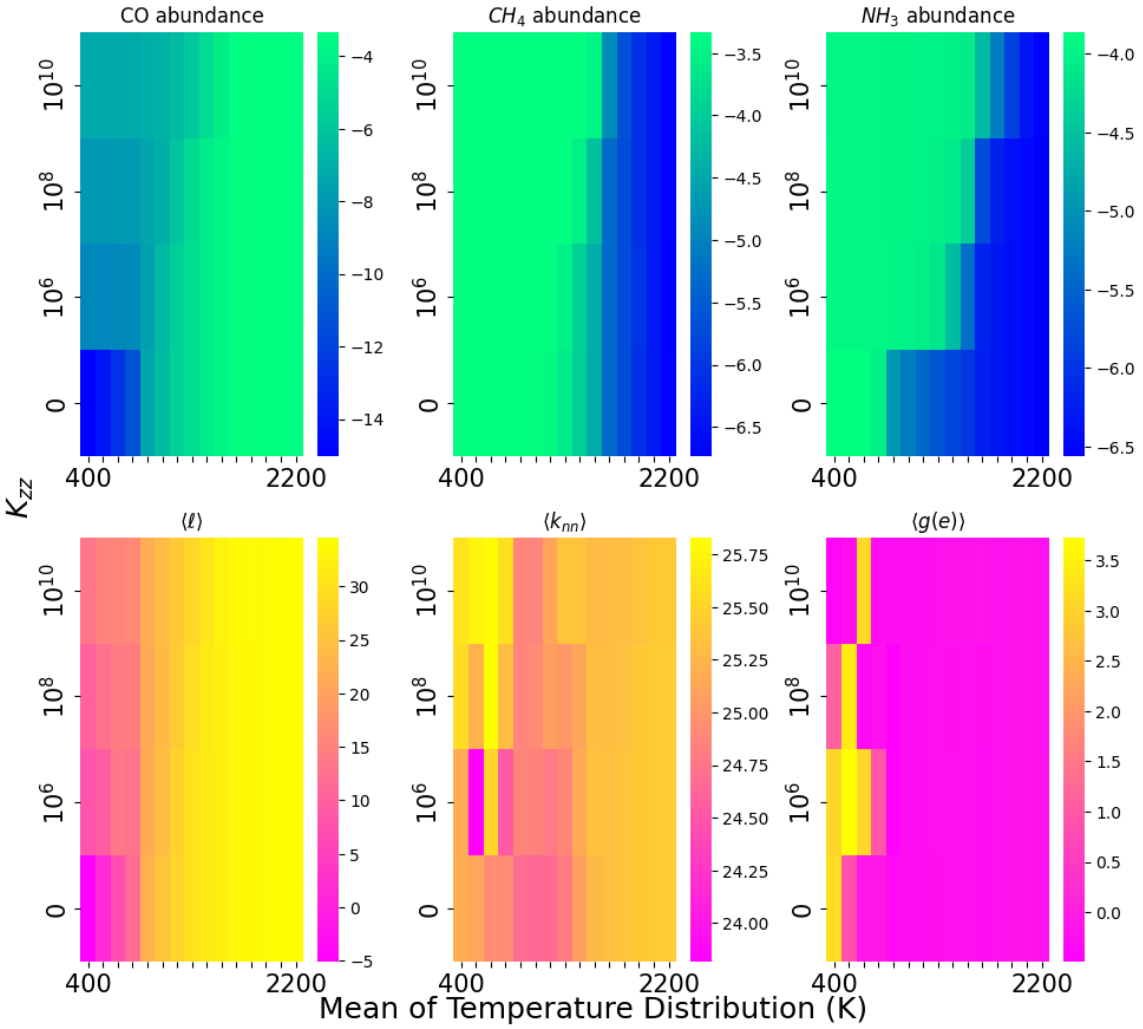


Figure 3.4: Output from the interpolation functions: contour plots of network measures and species abundance data as a function of the mean temperature of the input distribution and K_{zz}. Metallicity is 1 solar. The x-axis is the mean temperature in kelvin of the Gaussian distributions analyzed, the y-axis is K_{zz}, and the color scale represents the value of the plotted network measures or mixing ratio. These heat maps demonstrate that the interpolation function can yield useful insights into relationships between atmospheric parameters. For example, the abundance of CO appears to be positively correlated to high values of average shortest path length, whereas CH_4 appears to be negatively correlated. NH_3's relationship appears much less clear but does have a weak positive correlation with the average edge betweenness centrality. For the full set of contour plots detailing our results across multiple metallicity values, see Figures A1–A10 in the Appendix. For the description of the network measures in the figure, see Table 3.1 and Section 3.2.2.

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 Fig 3.7 and 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 3.7) 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 $\pm$250K (Figures 3.5 and 3.6). 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.

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, Φ 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 Φ as the primary thermodynamic measure in this work). 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 grew stronger as distance from equilibrium increased.

3.4.2 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 during either 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 due 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 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 more subtle 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_2O, and NH_3— and their associated reactions.

We found that perturbing the networks changed the distribution of network mea-

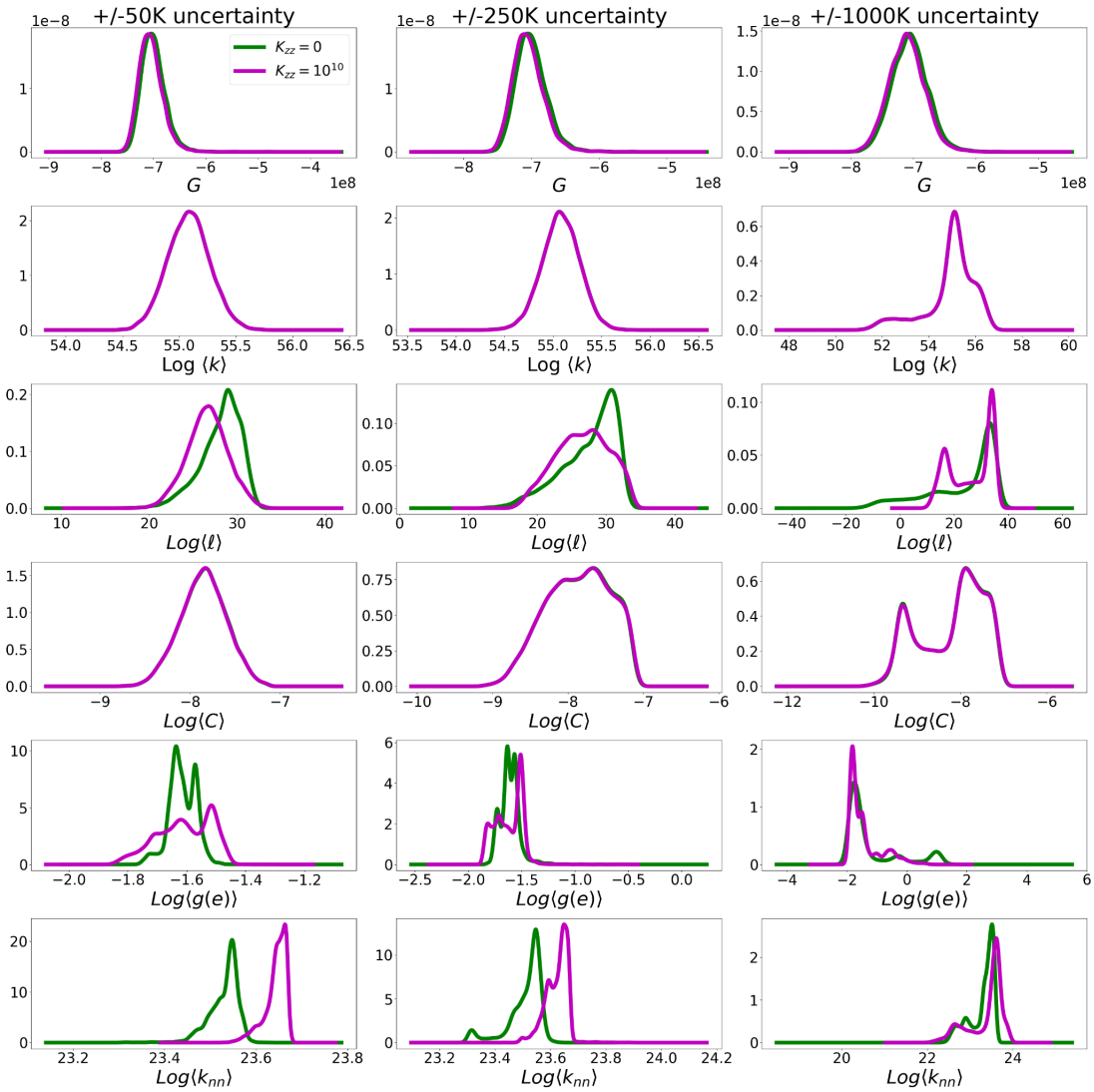


Figure 3.5: T = 1200K. Distributions of network measures and thermodynamic parameters, interpolated from a 10,000-point normal distribution of initial conditions centered on a mean temperature of 1200 K, at K_{zz} values of 0 and 10^{10}, and uncertainties in temperature of 50, 250, and 1000 K. The distributions 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 3.1 and Section 3.2.2.

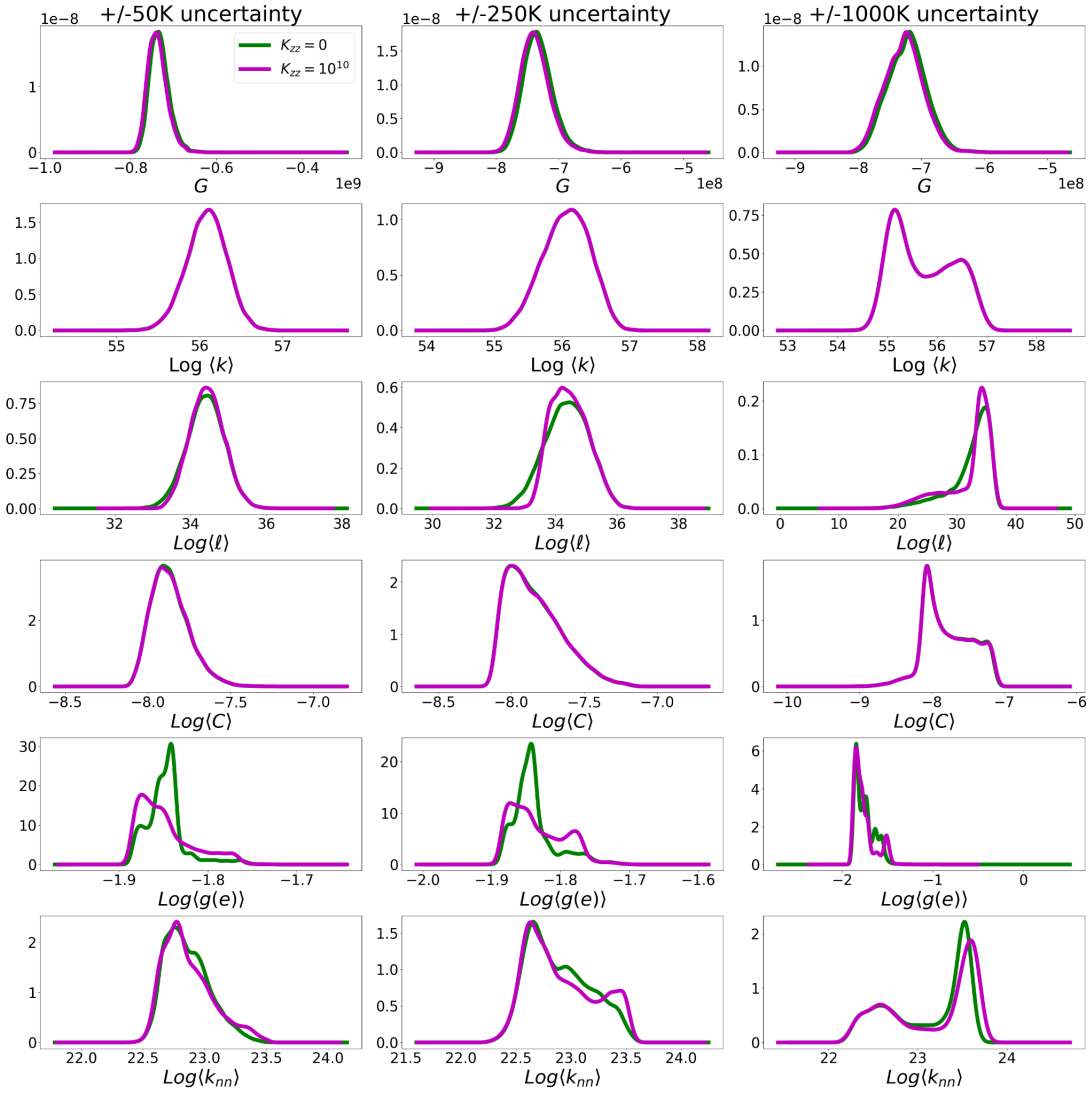


Figure 3.6: T = 2000K. Distributions of network topology, interpolated from a 10,000-point normal distribution of initial conditions centered on a mean temperature of 2000 K, at K_{zz} values of 0 and 10^{10}, and uncertainties in temperature of 50, 250, and 1000 K. Distributions with different K_{zz} values are less distinguishable from each other, likely because the faster thermal kinetics forces the system back to equilibrium more rapidly than chemical species can be dredged up via vertical mixing. This is in line with 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 bimodal distributions seen in the ± 1000 K case are an artifact of adding noise to the data. For the description of the network measures in the figure, see Table 3.1 and Section 3.2.2.

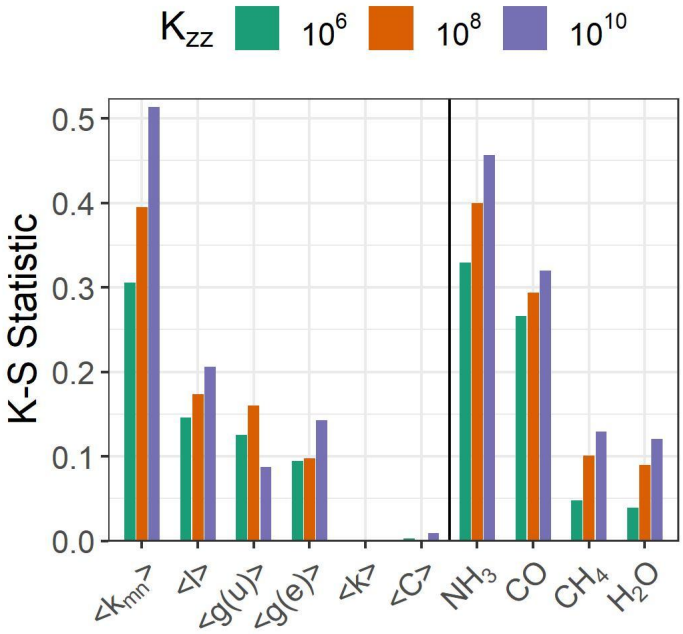


Figure 3.7: K-S metric values for the distributions of observable molecule abundance, and network topological measures. Distributions were calculated from a normal distribution of initial conditions, centered at 900 K, 50_{sol} metallicity, and 100 mb. Values were calculated with respect to the distribution at equilibrium ($K_{zz} = 0$). Uncertainty for the distributions was $\pm$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 3.1 and Section 3.2.2.

surements, even when only the CH_4-CO pathway was removed (see Figure 3.8). 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).

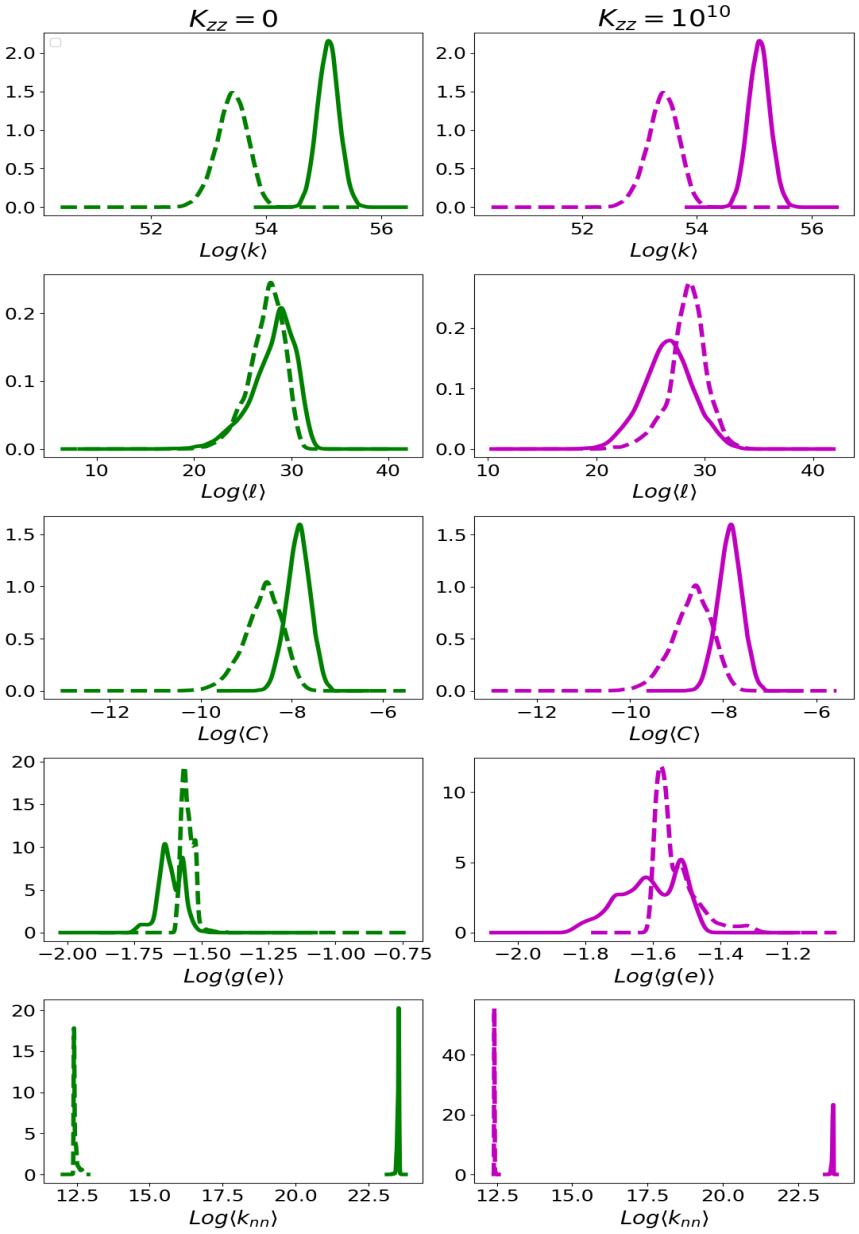


Figure 3.8: $T = 1200$ K w/ perturbations w/ pathway removed, uncertainty ± 50 K. Distributions of network measures when the CH_4-CO quenching pathway is removed from the network (dashed line) vs. complete networks (solid line), at 1200 K and 1 solar metallicity: clustering coefficient and average neighbor degree allow distinction between the two. For the description of the network measures in the figure, see Table 1 and Section 2.2.

3.4.3 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 and 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 atmosphere 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 dataset into training and testing sets with 80-20 ratio randomly using a Python package *sklearn.model_selection.train_test_split* and train *XGBClassifer* on the training sets with min_child_weight=10, max_depth=5, and n_estimators=1000.

Figure 3.9 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 3.9, 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.

3.4.4 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

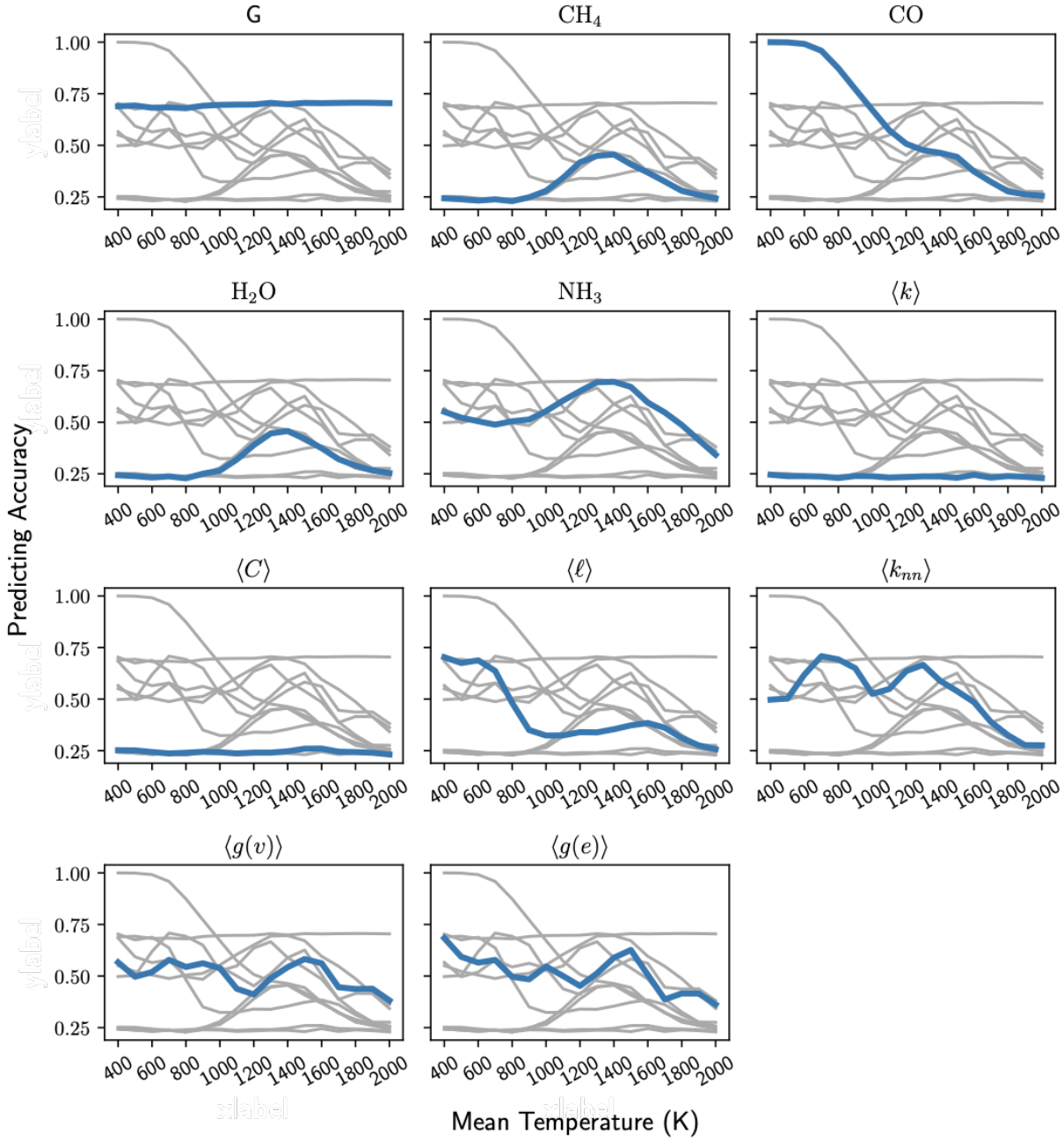


Figure 3.9: Predictive power of variables for the disequilibrium state of the atmospheric chemical reaction networks. Each panel of the plot shows the predicting accuracy of one individual variable denoted on the title (highlighted in blue) compared to the rest 10 variables shown in other panels (colored in gray) for atmospheric networks associated with $K_{zz} = 0$, 10^6, 10^8, and 10^{10} as a function of the mean temperature with uncertainty $\pm$50K (where grey lines show the predicting accuracy from the other variables 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_2O; and the average topological properties (See Table 3.1 and Section 3.2.2)

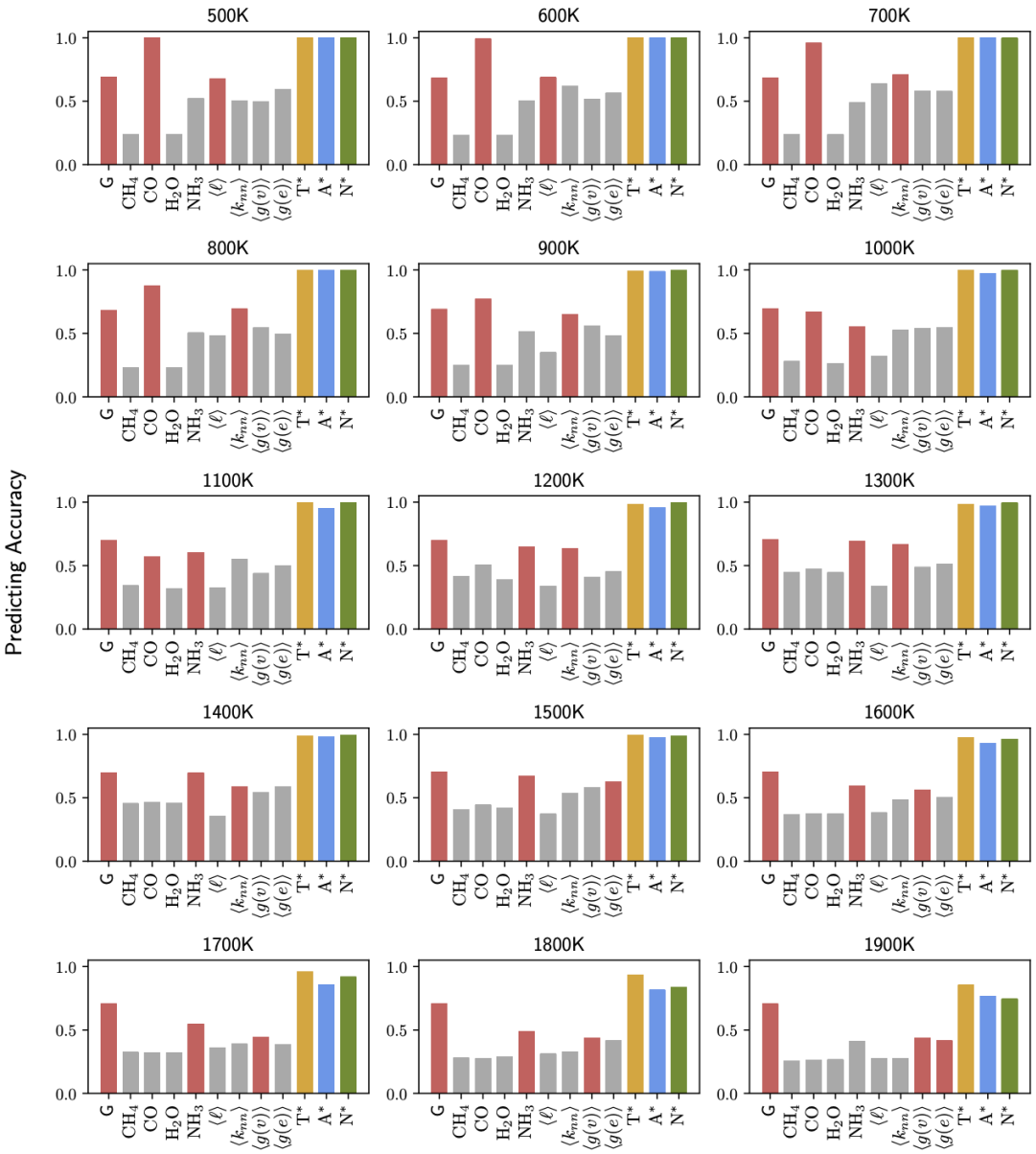


Figure 3.10: Boosting up the predicting accuracy for the disequilibrium state of the atmospheric chemical reaction networks with multivariate information. Each panel of the plot shows the predicting accuracy of variables for atmospheric networks associated with $K_{zz} = 0$, 10^6, 10^8, and 10^{10} at 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 3.1 and Section 3.2.2). 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 3.9, respectively. The three different multivariable sets demonstrated higher predicting accuracy than individual variables even in the cases where the highest predictors are not included.

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 be inferred from Figure 3.9. 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 3.10 also shows the predicting accuracy measured by using two other different multivariate groups: the set of the abundances of NH_3, CH_4, CO, and H_2O (denoted by A*); and the set of six network topological measures (denoted by N*). The two groups 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 ±250 K and ±500 K can be found in the Appendix (Figures A19-A21).

3.4.5 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_2O—was removed from networks and computed their predicting

accuracy for K_{zz}based on the perturbed network datasets. Figure 3.11 shows the predictive power of 4 individual topological measures – $\langle \ell \rangle$,$\langle k_{nn} \rangle$,$\langle g(v) \rangle$, and $\langle g(e) \rangle$ – and the set of topological measures, N*, before (gray) and after (green) removal of each of four chemical species, and Abundance, the set of the abundance of 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.

3.5 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

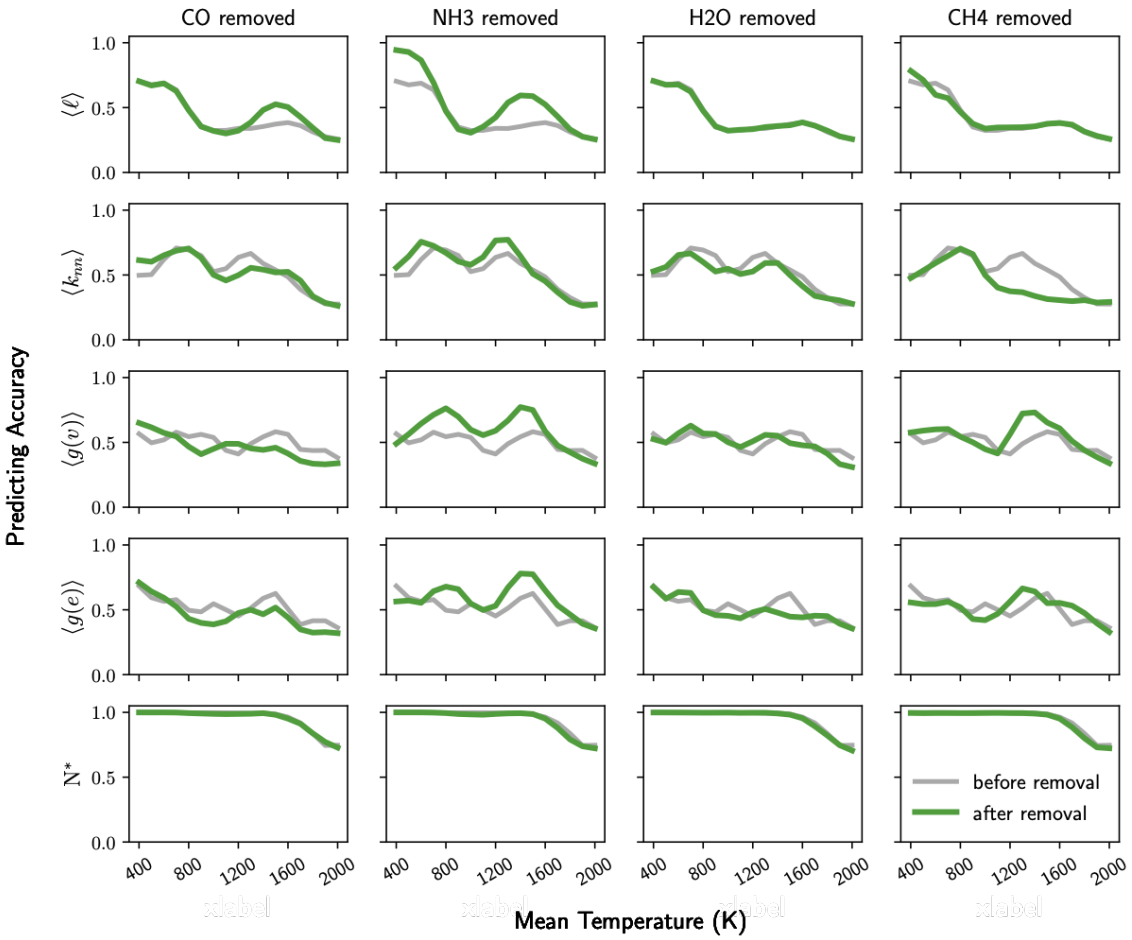


Figure 3.11: Robustness of network topology as predictions for the disequilibrium state of the atmospheric chemical reaction networks against perturbation. The predicting accuracy of topological variables and the set of them for K_{zz} is measured where one of four chemical species – NH_3 , CH_4 , CO, and H_2O – was knocked out. The uncertainty of temperature in the given dataset is $\pm$50K.

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 abun-

dances 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 (Centler and Dittrich, 2007; Estrada, 2012; Schwieterman *et al.*, 2018; Walker *et al.*, 2018). 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 use-

ful 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.

REFERENCES

Albert, R. and A.-L. Barabási, "Statistical mechanics of complex networks", Rev. Mod. Phys. 74, 1, 47–97 (2002).

Antoniou, I. E. and E. T. Tsompa, "Statistical analysis of weighted networks", Discrete Dynamics in Nature and Society 2008, e375452 (2008).

Barabási, A.-L. and M. Pósfai, *Network Science* (Cambridge University Press, Cambridge, United Kingdom, 2016).

Barrat, A., M. Barthélemy, R. Pastor-Satorras and A. Vespignani, "The architecture of complex weighted networks", Proceedings of the National Academy of Sciences 101, 11, 3747–3752 (2004).

Benneke, B. and S. Seager, "Atmospheric retrieval for super-Earths: uniquely constraining the atmospheric composition with transmission spectroscopy", ApJ 753, 2, 100 (2012).

Catling, D. C., J. Krissansen-Totton, N. Y. Kiang, D. Crisp, T. D. Robinson, S. DasSarma, A. J. Rushby, A. Del Genio, W. Bains and S. Domagal-Goldman, "Exoplanet biosignatures: a framework for their assessment", Astrobiology 18, 6, 709–738 (2018).

Centler, F. and P. Dittrich, "Chemical organizations in atmospheric photochemistries—A new method to analyze chemical reaction networks", Planetary and Space Science 55, 4, 413–428 (2007).

Chen, T. and C. Guestrin, *XGBoost* (2016), publication Title: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Cobb, A. D., M. D. Himes, F. Soboczenski, S. Zorzan, M. D. O'Beirne, A. G. Baydin, Y. Gal, S. D. Domagal-Goldman, G. N. Arney, D. Angerhausen and 2018 NASA FDL Astrobiology Team II, "An ensemble of Bayesian neural networks for exoplanetary atmospheric retrieval", AJS 158, 1, 33 (2019).

Crossfield, I. J. M., "Observations of exoplanet atmospheres", PASP 127, 956, 941 (2015).

Estrada, E., "Returnability as a criterion of disequilibrium in atmospheric reactions networks", Journal of Mathematical Chemistry 50, 6, 1363–1372 (2012).

Fortney, J. J., R. I. Dawson and T. D. Komacek, "Hot jupiters: origins, structure, atmospheres", J. Geophys. Res. Planets (2021).

Fujii, Y., D. Angerhausen, R. Deitrick, S. Domagal-Goldman, J. L. Grenfell, Y. Hori, S. R. Kane, E. Pallé, H. Rauer, N. Siegler, K. Stapelfeldt and K. B. Stevenson, "Exoplanet biosignatures: observational prospects", Astrobiology 18, 6, 739–778 (2018).

Gleiss, P. M., P. F. Stadler, A. Wagner and D. A. Fell, "Relevant cycles in chemical reaction networks", Advs. Complex Syst. 04, 02n03, 207–226 (2001).

Hayes, J. J. C., E. Kerins, S. Awiphan, I. McDonald, J. S. Morgan, P. Chuanraksasat, S. Komonjinda, N. Sanguansak, P. Kittara and (SPEARNET), "Optimizing exoplanet atmosphere retrieval using unsupervised machine-learning classification", Mon. Not. R. Astron. Soc. 494, 3, 4492–4508 (2020).

Heng, K., J. M. Mendonça and J.-M. Lee, "Analytical models of exoplanetary atmospheres. ii. radiative transfer via the two-stream approximation", The Astrophysical Journal Supplement Series 215, 1, 4 (2014).

Jeong, H., B. Tombor, R. Albert, Z. N. Oltvai and A.-L. Barabási, "The large-scale organization of metabolic networks", Nature 407, 6804, 651–654 (2000).

Jolley, C. and T. Douglas, "Topological biosignatures: large-scale structure of chemical networks from biology and astrochemistry", Astrobiology 12, 1, 29–39 (2011).

Jolley, C. C. and T. Douglas, "A network-theoretical approach to understanding interstellar chemistry", The Astrophysical Journal 722, 2, 1921–1931 (2010).

Kiang, N. Y., S. Domagal-Goldman, M. N. Parenteau, D. C. Catling, Y. Fujii, V. S. Meadows, E. W. Schwieterman and S. I. Walker, "Exoplanet biosignatures: at the dawn of a new era of planetary observations", Astrobiology 18, 6, 619–629 (2018).

Krissansen-Totton, J., G. N. Arney, D. C. Catling, R. Felton, J. Fortney, R. Garland, P. Irwin, R. Kopparapu, O. Lehmer, J. Lustig-Yaeger and others, "Atmospheric disequilibrium as an exoplanet biosignature: Opportunities for next generation telescopes", BAAS 51, 3, 158 (2019).

Krissansen-Totton, J., D. S. Bergsman and D. C. Catling, "On detecting biospheres from chemical thermodynamic disequilibrium in planetary atmospheres", Astrobiology 16, 1, 39–67 (2016).

Krissansen-Totton, J., S. Olson and D. C. Catling, "Disequilibrium biosignatures over Earth history and implications for detecting exoplanet life", Science Advances 4, 1, eaao5747 (2018).

Line, M. R., J. Teske, B. Burningham, J. J. Fortney and M. S. Marley, "Uniform atmospheric retrieval analysis of ultracool dwarfs. I. Characterizing benchmarks, Gl 570D and HD 3651B", The Astrophysical Journal 807, 2, 183 (2015).

Line, M. R., A. S. Wolf, X. Zhang, H. Knutson, J. A. Kammer, E. Ellison, Pieter Deroo, D. Crisp and Y. L. Yung, "A systematic retrieval analysis of secondary eclipse spectra. I. A comparison of atmospheric retrieval techniques", The Astrophysical Journal 775, 2, 137 (2013).

Madhusudhan, N., "Exoplanetary atmospheres: key insights, challenges and prospects", Annual Review of Astronomy and Astrophysics 57, 617–663 (2019).

Madhusudhan, N., J. Harrington, K. B. Stevenson, S. Nymeyer, C. J. Campo, P. J. Wheatley, D. Deming, J. Blecic, R. A. Hardy, N. B. Lust, D. R. Anderson, A. Collier-Cameron, C. B. T. Britt, W. C. Bowman, L. Hebb, C. Hellier, P. F. L. Maxted, D. Pollacco and R. G. West, "A high C/O ratio and weak thermal inversion in the atmosphere of exoplanet WASP-12b", Nature 469, 7328, 64–67 (2011).

Massey, F. J., "The Kolmogorov-Smirnov test for goodness of fit", Journal of the American Statistical Association 46, 68–78 (1951).

Meadows, V. S., "Reflections on O_2 as a biosignature in exoplanetary atmospheres", Astrobiology 17, 10, 1022–1052 (2017).

Molaverdikhani, K., T. Henning and P. Mollière, "From cold to hot irradiated gaseous exoplanets: fingerprints of chemical disequilibrium in atmospheric spectra", The Astrophysical Journal 883 (2019).

Mollière, P., R. v. Boekel, C. Dullemond, T. Henning and C. Mordasini, "Model atmospheres of irradiated exoplanets: the influence of stellar parameters, metallicity, and the C/O ratio", The Astrophysical Journal 813, 1, 47 (2015).

Montañez, R., M. A. Medina, R. V. Solé and C. Rodríguez-Caso, "When metabolism meets topology: reconciling metabolite and reaction networks", Bioessays 32, 3, 246–256 (2010).

Moses, J. I., C. Visscher, J. J. Fortney, A. P. Showman, N. K. Lewis, C. A. Griffith, S. J. Klippenstein, M. Shabram, A. J. Friedson, M. S. Marley and R. S. Freedman, "Disequilibrium carbon, oxygen, and nitrogen chemistry in the atmospheres of HD 189733b and HD 209458b", The Astrophysical Journal 737, 1, 15 (2011).

Newman, M., *Networks: An Introduction* (OUP Oxford, 2010).

Newman, M., A.-L. Barabási and D. J. Watts, *The Structure and Dynamics of Networks* (Princeton University Press, 2011).

Newman, M. E. J., "The structure and function of complex networks", SIAM Rev. 45, 2, 167–256 (2003).

Onnela, J.-P., J. Saramäki, J. Kertész and K. Kaski, "Intensity and coherence of motifs in weighted complex networks", Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 71, 6 Pt 2, 065103 (2005).

Schwieterman, E. W., N. Y. Kiang, M. N. Parenteau, C. E. Harman, S. DasSarma, T. M. Fisher, G. N. Arney, H. E. Hartnett, C. T. Reinhard, S. L. Olson, V. S. Meadows, C. S. Cockell, S. I. Walker, J. L. Grenfell, S. Hegde, S. Rugheimer, R. Hu and T. W. Lyons, "Exoplanet biosignatures: a review of remotely detectable signs of life", Astrobiology 18, 6, 663–708 (2018).

Seager, S., "Exoplanet habitability", Science 340, 6132, 577–581 (2013).

Seager, S. and D. Deming, "Exoplanet atmospheres", Annual Review of Astronomy and Astrophysics 48, 631–672 (2010).

Sharma, N., *XGBoost. The Extreme Gradient Boosting for Mining Applications* (GRIN Verlag, 2018).

Simoncini, E., N. Virgo and A. Kleidon, "Quantifying drivers of chemical disequilibrium: theory and application to methane in the Earth's atmosphere", Earth System Dynamics 4, 317–331 (2013).

Smith, H. B., H. Kim and S. I. Walker, "Scarcity of scale-free topology is universal across biochemical networks", Sci. Rep. 11, 1, 6542 (2021).

Solé, R. V. and A. Munteanu, "The large-scale organization of chemical reaction networks in astrophysics", Europhysics Letters (EPL) 68, 2, 170–176 (2004).

Thorngren, D., P. Gao and J. J. Fortney, "The intrinsic temperature and radiative–convective boundary depth in the atmospheres of hot Jupiters", The Astrophysical Journal 884, 1, L6 (2019).

Tsai, S.-M., J. R. Lyons, L. Grosheintz, P. B. Rimmer, D. Kitzmann and K. Heng, "VULCAN: an open-source, validated chemical kinetics Python code for exoplanetary atmospheres", The Astrophysical Journal Supplement Series 228, 2, 20 (2017).

Venot, O., E. Hébrard, M. Agúndez, M. Dobrijevic, F. Selsis, F. Hersant, N. Iro and R. Bounaceur, "A chemical model for the atmosphere of hot Jupiters", Astron. Astrophys. Suppl. Ser. 546, A43 (2012).

Walker, S. I., W. Bains, L. Cronin, S. DasSarma, S. Danielache, S. Domagal-Goldman, B. Kacar, N. Y. Kiang, A. Lenardic, C. T. Reinhard, W. Moore, E. W. Schwieterman, E. L. Shkolnik and H. B. Smith, "Exoplanet biosignatures: future directions", Astrobiology 18, 6, 779–824 (2018).

Zahnle, K., M. S. Marley, R. S. Freedman, K. Lodders and J. J. Fortney, "Atmospheric sulfur photochemistry on hot Jupiters", The Astrophysical Journal 701, 1, L20–L24 (2009).

Chapter 4

ASSESSING BIOSIGNATURES AND TECHNOSIGNATURES IN THE ATMOSPHERES OF TERRESTRIAL EXOPLANETS USING NETWORK THEORY

4.1 Abstract

The next decade holds great promises for the detection and characterization of terrestrial exoplanet atmospheres. With this comes the prospect of edging closer to the discovery of alien life from atmospheric data, but a significant challenge looms: the majority of candidate exoplanet biosignatures are susceptible to false positives that can confound claims of life detection. Within the astrobiology community, this has motivated a renewed search for resilient and agnostic atmospheric biosignatures that are unlikely to generate false positives. Here, I demonstrate how a systems-level approach, representing atmospheric chemistry as a complex network, can help distinguish between the atmospheres of planets without life and those with it. I also show that this approach enables a unified treatment to identify life generating both atmospheric biosignature and technosignatures.

I used atmos to simulate 30,000 terrestrial atmospheres, organized in two model sets: an Earth analogue with Archean atmospheric conditions for the investigation of biosignatures related to biogenic CH_4, and an Earth analogue with modern atmospheric conditions for the investigation of CFC-based technosignatures. Half of each set incorporates a flux variable: CH_4, for the Archean Earth analogues, or CFC-12, for the modern Earth analogues. I found network topological metrics of the large-scale organization of chemical reaction networks generated from the modeled atmospheres

allowed a better classification of the models than looking at CH_4 abundance alone. Additionally, network metrics could be used to detect the presence of CFC-12. These results suggest that the development of measurements and techniques targeted at assessing system-level features in the large-scale network organization of planetary atmospheric chemistry could prove useful in increasing confidence of life detection above current approaches that focus on gas abundances only.

4.2 Introduction

The James Webb Space Telescope (JWST) and the high-resolution spectroscopy in planned 30-meter-class ground-based observatories (*e.g.*, the Extremely Large Telescope) make the long sought-after goal of characterizing terrestrial exoplanet atmospheres increasingly likely (Fujii *et al.*, 2018). This creates a pressing need to identify the most effective method(s) for leveraging the data from these observatories to assess the likelihood that a given exoplanet harbors a biosphere or technosphere (Walker *et al.*, 2018; Sheikh *et al.*, 2021).

As discussed in Section 1.1, early approaches relied on the identification of "smoking guns", viewed as indisputable evidence for life (Meadows, 2017). These include observationally accessible molecules, produced in abundance on Earth by biology. An example is O_2 (Tremblay *et al.*, 2019), particularly in concert with CH_4 (Tremblay *et al.*, 2019; Kaltenegger *et al.*, 2020). As knowledge of exoplanets increased and the immense diversity of worlds became apparent, atmospheric molecular biosignatures – most prevalent among them O_2 - have come under more scrutiny due to the risks for false positives (Domagal-Goldman *et al.*, 2014; Meadows *et al.*, 2018; Harman and Domagal-Goldman, 2018) and false negatives (Wordsworth and Pierrehumbert, 2014; Reinhard *et al.*, 2017). CH_4, another species of interest, is also vulnerable to false positives (Guzmán-Marmolejo *et al.*, 2013). Even the well-studied O_2-CH_4

chemical disequilibrium (Sagan *et al.*, 1993) can be mimicked if the atmosphere of an exoplanet's unseen moon contaminates the spectra of its host exoplanet, leading to the misleading appearance of a single body with a strong atmospheric disequilibrium (Rein *et al.*, 2014).

The atmospheric composition of exoplanets is also a target in the search for extraterrestrial technology (Wright, 2018; Haqq-Misra *et al.*, 2022c). Several industrial gasses have been posited, including N_2O (Kopparapu *et al.*, 2021), as well as overall changes in the mix of nitrogenous gasses (Haqq-Misra *et al.*, 2022a). Chlorofluorocarbons (CFCs) are also of interest, due to their distinctive spectral signatures Haqq-Misra *et al.* (2022b) and longevity in an atmosphere Balbi and Ćirković (2021). However, they are not without their weaknesses, as some species of halocarbons can be produced by volcanic activity rather than technology (Visscher *et al.*, 2004; Broadley *et al.*, 2018; Klobas and Wilmouth, 2019). Additionally, CFC spectral detection is only barely within near-term capabilities; it is estimated that CFC concentration would have to be an order of magnitude or more greater than the peak of anthropogenic CFC concentration of 0.225-0.515 ppb, to be readily detectable by JWST (Lin *et al.*, 2014; Haqq-Misra *et al.*, 2022b).

The observational challenges and confounding abiotic factors confronting exoplanet life detection efforts suggest new approaches will be necessary to build confidence in the likelihood of a planet's atmosphere being influenced by biology or technology (Walker *et al.*, 2018).

A possible approach is the analysis of the topology of atmospheric chemical reaction networks (CRNs), which probe larger scale statistical features of a planetary atmosphere's chemistry (Schwieterman *et al.*, 2018; Walker *et al.*, 2020) than studying individual gas abundances alone. In my prior work, I showed how network metrics could be reliable indicators of the distance from thermochemical equilibria in hot

Jupiter atmospheres over a wide range of temperatures (Fisher *et al.*, 2022). Building on this, and previous work examining the CRN topology of individual atmospheres (Solé and Munteanu, 2004; Wong *et al.*, 2023)(see Section 1.3.1 for more detail), I now turn to evaluating the effectiveness of network topology as a discriminator of biological or technological activity for populations of Earth-like atmospheres. I expect that abiotic, biological or technological sources could mediate surface fluxes and thus drive disequilibria in the atmospheres of exoplanets.

These studies suggest features of the large-scale structure of planetary atmospheric chemistry, as characterized by its global network topology, could provide a quantitative framework for assessing differences between planets with life (or technology) and those without it, potentially providing a solution to the dilemma of false positives and negatives. However, to assess this possibility requires 1) care in constructing network representations that accurately capture relevant atmospheric physics and chemistry, and 2) that the networks are constructed in the same way across likely abiotic, biological and technological models in order to compare them.

Thus, the question is raised: can I solve the dilemma of false positives and negatives and distinguish between populations of terrestrial atmospheres that have a biosphere or technosphere, and those without using network topology metrics? And if so, do these metrics performs better than looking at the abundance of potentially relevant species (such as CH_4) alone?

In what follows, I attempt to answer this question, and test the hypothesis that planetary atmospheric reaction network structure can discriminate between abiotic, biotic and technological surface-flux driven disequilibria. I use CH_4 as a case study gas produced by biology on planets like the Archean Earth, and CFC-12 as a case study gas emitted by technological industry on planets resembling modern Earth. In the Methods section, I provide an overview of my analysis pipeline, from atmospheric

models to chemical reaction networks to statistical analysis of network structure. In the Results section, I provide a metric-by-metric assessment of these metrics ability to distinguish between biotic and abiotic atmospheres, and technological and non-industrial atmospheres. Lastly, in the Discussion section, I consider why some features of network structure appear to be better discriminators of biological or technological driven fluxes than others, and the strengths and limitations of this approach to guide development of remotely observable biosignatures.

4.3 Methods

4.3.1 Archean Earth-like Atmospheres with CH_4

I modeled 20,000 200-layer Archean Earth models using atmos, a Fortran atmospheric modeling package (Arney *et al.*, 2016) (see Section 2.3.2 for more detail on atmos and its internal workings). Each model had differing values for CH_4 surface flux, and varied whether or not a biological methanogenesis reaction was included depending on which population of models a model was part of, but were otherwise identical in composition (see Table C.1 in the Appendix). To make running large numbers of models easier, I used PyATMOS, a Python wrapper for atmos that allows massive batch runs of the atmospheric modeling package (Bella *et al.*, 2018). CH_4 was chosen as my target biosignature gas for this study, both due its important role in the Archean Earth's atmosphere (Kasting and Siefert, 2002; Catling and Zahnle, 2020) and its potential spectral detectability (Arney *et al.*, 2016). The atmospheres were modeled as orbiting around a red dwarf star, Gliese 576, with an eye towards future observations, as terrestrial worlds around red dwarf stars are likely to be the first to have their atmospheres analyzed (Fujii *et al.*, 2018; Lincowski *et al.*, 2019; Arney, 2019).

The Archean Earth-like models were subdivided into four populations, based on the abundance and source of CH_4 in the atmosphere:

1. *Abiotic flux from serpentinization.* CH_4 is introduced as an surface flux from the reaction of CO_2 with H_2 generated from the transformation of ferromagnesian minerals into serpentinite via water-rock reactions (Thompson *et al.*, 2022). This provides a potential confounding false-positive case for comparison to the models with biological surface fluxes. CH_4 fluxes are drawn from a Gaussian distribution centered around 10^6 molecules/cm^2/s, which yields flux values with a range between 10^{-2} and 10^8 molecules/cm^2/s, based on Etiope and Sherwood Lollar (2013).

2. *Biotic surface flux.* In these models, CH_4 is introduced as a surface flux, and a biological methanogenesis reaction used by Terrestrial methanogens was explicitly included in the reaction network (Catling and Zahnle, 2020). This scenario represents a true positive case for life detection. CH_4 fluxes are drawn from a Gaussian distribution centered around 10^{11} molecules/cm^2/s, which yields flux values with a range between 10^3 and 10^{13} molecules/cm^2/s, based on Archean ecosystem models from Catling and Zahnle (2020).

3. *Anomalous surface flux.* CH_4 is modeled as a surface flux with the same levels as the biotic surface flux models, but the reaction network does not include the biological methanogenesis reactions. This group is referred to as the "anomalous high flux group" since the source of the flux is undefined. This population of models was included to present a challenge for distinguishing between an anomalous source and and an explicitly biological source. CH_4 flux values are drawn from a Gaussian distribution centered around 10^{11} molecules/cm^2/s, which yields flux values with a range between 10^3 and 10^{13} molecules/cm^2/s,

the same as the biological surface flux models.

4. *Abiotic atmospheric steady state.* Abiotic CH_4 was modeled as part of an atmospheric steady state, with the same abundances as the biotic model. This population of models represents a worst-case scenario for a false positive, where the CH_4 abundances are tuned to be close to identical with the biotic scenarios. Instead of a being modeled as a flux, the CH_4 concentration is held constant at an initial value ranging from 10^{-4} to 10^4 ppm. This scenario is probably the least realistic, as atmospheres with a fixed mixing ratio of CH_4 rarely exist in nature, and atmospheres with CH_4-rich secondary atmospheres are difficult to produce (Tian and Heng, 2023), but is included for completeness.

As mentioned in Section 2.3.2, due to limitations in atmos, the methanogenesis reaction

$$CO_2 + 4H_2 \rightarrow CH_4 + 2H_2O \tag{4.1}$$

in biological surface flux models was included in the CRN as a three-step reaction, involving several fictitious short-lived intermediate species:

$$H_2 + H_2 \rightarrow \text{`}H_4\text{'} \tag{4.2}$$

$$\text{`}H_4\text{'} + CO_2 \rightarrow \text{`}H_4C_2\text{'} \tag{4.3}$$

$$\text{`}H_4\text{'} + \text{`}H_4CO_2\text{'} \rightarrow CH_4 + H_2O + H_2O \tag{4.4}$$

These intermediate species are considered to be short-lived species, which are modeled separately in atmos than long-lived species, and are not included in the model's Jacobian. While they are included in the network, their effect on the network topology is expected to be minimal.

The kinetics of methanogenesis are modeled by an Arrhenius equation, using rate constants for anaerobic microbes in the form of

$$m_E = 3.3 * e^{\frac{-6.94\times10^4}{R} - \frac{1}{T} - \frac{1}{298}} \tag{4.5}$$

(Tijhuis *et al.*, 1993; Seager *et al.*, 2013)

Where m_E refers to the maintenance energy required to sustain the population of methanogenic microbes.

For the purposes of perturbation testing of the model, I also ran a set of models where acetylene (C_2H_2) was removed. This species plays a key role in the formation of hazes, thus eliminating it removes a (poorly constrained) hydrocarbon sink. This significantly changes the behavior of the chemistry of the atmosphere by leading to much higher levels of long-chain hydrocarbons. Detection of this change by the network analysis pipeline would help build confidence that my findings are reflective of the modeled atmospheric chemistry (see Section 2.4.2 for more detail on perturbation testing to validate atmospheric models).

4.3.2 Modern Earth-like Atmospheres with CFC-12

To test the impact of CFC emissions on atmospheric CRN topology, I used the existing dataset generated by the NASA Frontiers Development Lab using the PyATMOS package. As a test of PyATMOS' capabilities, FDL modeled a large set of conditions with a resulting dataset composed of over 120,000 modeled modern-Earth analogue atmospheres of varying composition (FDL, 2019). From this dataset, I randomly selected 5,000 models, and re-ran them with the same otherwise initial conditions (see see Table 4.1, as well as C.2 in the Appendix for a full list) after incorporating a surface flux of CCl_2F_2, better known as CFC-12 or Freon^{TM}. CFC-12 is notorious for its depleting effect on atmospheric ozone; when exposed to UV light,

Table 4.1: Species concentrations that are free parameters in the modern Earth analogue population of models

Species	Concentration (fractional)
CH_4	1.63×10^{-6}-0.13
CO_2	8×10^{-8}-0.4
H_2	6×10^{-8}-0.19
O_2	0.02-0.5
CFC-12	4.2×10^{-11}-2.3×10^{-6}

it gives rise to a Cl radical:

$$CCl_2F_2 + h\nu \rightarrow CClF_2 + Cl^{\cdot} \tag{4.6}$$

This radical can in turn catalyze the conversion of O_3 into other oxygen species, via

$$Cl^{\cdot} + O_3 \rightarrow ClO + O_2 \tag{4.7}$$

and

$$ClO + O_3 \rightarrow ClOO + O_2 \tag{4.8}$$

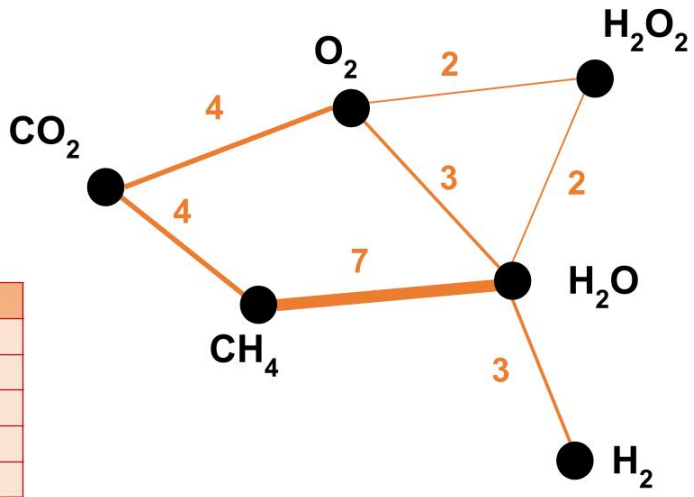
In order to simplify atmospheric chemistry for the purposes of constraining detectability, CFC-12 abundances were modeled as a steady state, with values ranging from 0.05 to 5 ppb, based on the peak abundance of industrial CFC-12 in the Earth's atmosphere of 0.515 ppb, as modeled by Haqq-Misra *et al.* (2022b).

4.3.3 Constructing and Measuring Networks

The models were run until they converged on a steady state. To get as large a sample size of models as possible, I implemented a maximum simulation time cut-off of 5×10^{10} simulated years, though most converged before reaching that point. In the interest of consistency, I pared down the populations of models in the Archean Earth-like set that converged to a steady state (41% to 95% of the models run, depending on the model category) until they all had the same number of models (721); models with high CH_4 flux were preferentially removed, as they were considered to be less realistic. In the case of the modern Earth-like atmospheres, virtually all the models converged, and such paring was not necessary.

Once these models converged, I created a unipartite network representation of the chemical reactions in the atmosphere drawn from the reaction list found in the output files of each model, resulting in a population of atmospheric CRNs. To construct networks from these atmospheres, I represented each species present in the atmosphere mathematically as a node, within the network, with a node connected to another node if they co-participate in the same reaction; this connection, or edge, is weighted by the reaction rate (see Fig.4.1).

As the reactions involved were almost identical across the populations of modeled atmospheres, the edges were weighted according to the modeled reaction rates in order to be able to better understand how the presence or absence of life effects the flux across the network, and make the populations potentially more distinguishable from each other. The reaction rates were calculated by multiplying the temperature-dependent rate constants (in the form of their Arrhenius equation) by the abundance of the reactants; see Arney *et al.* (2016) for a full list of the reactions and their respective rate constants for the Archean-Earth like atmospheres, and the Modern



Edge	Weight
CO_2 - O_2	4
CO_2 - CH_4	4
CH_4 - H_2O	7
H_2O - O_2	3
H_2O - H_2O_2	2
H_2O - H_2	2
O_2 - H_2O_2	3

Figure 4.1: Simple example of a chemical network, composed of the reactions $CH_4 + O_2 \rightarrow H_2O + CO_2$, $H_2 + O_2 \rightarrow H_2O$, and $H_2O_2 \rightarrow H_2 + O_2$. Each species present in the reactions is represented by a node; reactant nodes are connected via edges to the nodes representing the products of the reactions they participate in.

Earth template for atmos for the modern Earth-like atmospheres.

The resulting network for the Archean Earth analogue had 74 nodes with 704 edges (reactions) for the abiotic case, and 77 nodes with 711 edges for the biotic case.

The modern Earth analogue models yielded a network with 77 nodes with 562 edges for the baseline case, and 78 nodes with 574 edges in the case including CFC-12. Species abundances were measured at a geopotential height of $z = 6.38 \times 10^6$ meters, which corresponds to a pressure of about 0.2 mbars.

I measured topological features of the created graphical representations of the atmosphere to characterize the structure of these networks using the NetworkX package (Hagberg *et al.*, 2008).

4.3.4 Analyzing CRN Topology

Once the graphical representation of an atmospheric model has been constructed, its network topology can be measured. Topological measures implemented in this study include mean degree, clustering coefficient, average shortest path length, node betweenness centrality, edge betweenness centrality, and average nearest neighbor degree (see Table 2.1 and Section 2.3.1 for detailed descriptions of these metrics). Thus, each population of models gives rise to populations of metrics that are then converted into probability distributions by using SciPy's *gaussian_kde* function (Virtanen *et al.*, 2020) and evaluating the area under the resulting curve using the minimum and maximum values of the population as bounds. These distributions, now normalized so that probability falls between 0 and 1, can then be analyzed in terms of distinguishability (see Figure 2.2, and Section 2.4.1 for more detail on quantifying distinguishability between distributions).

To do so, I calculated the pair-wise percentage overlap between distributions (see Figure 4.3) and the Kolmogorov-Smirnoff metric (see Figure 4.4). The former is simply just the percentage of values between two distributions that overlap; the smaller the value, the less overlap there is, and presumably the more distinguishable the two distributions. In practice, this means it's more likely that a given observed metric will definitively signify either the presence of absence of life; high overlap means that that the range of distribution observed metric values from abiotic and biotic atmospheres coincides, increasing the ambiguity of an observed value of a metric as to whether or not life is present. Essentially, a lower overlap means a lower chance of a false positive or negative.

The latter was calculated using the *ks_2samp* from SciPy.stats (Virtanen *et al.*, 2020), and is a measure of the distance between two distributions. The closer a

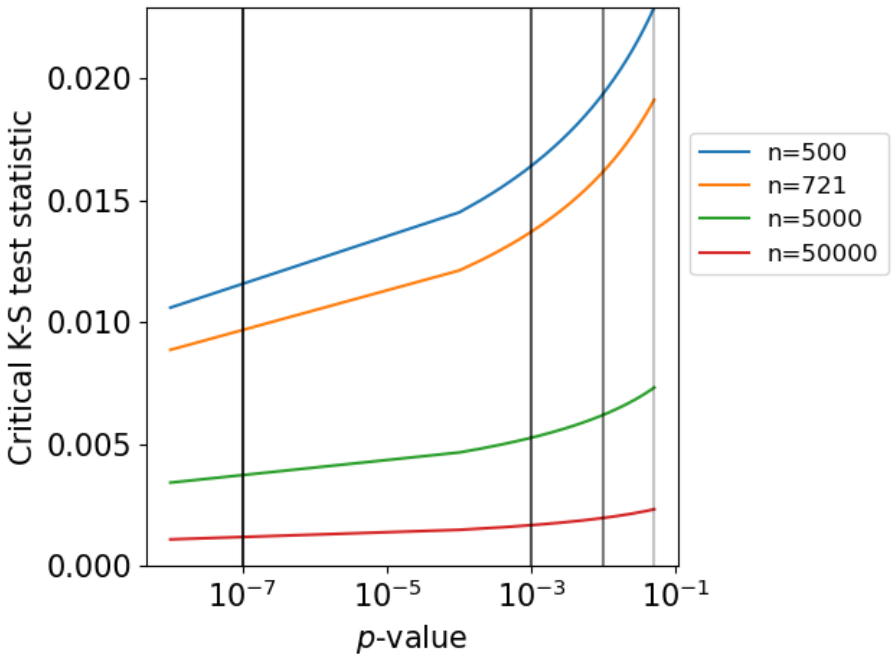


Figure 4.2: The threshold for a K-S test statistic to be considered significant for the distributions with n=721 and a p-value below 10^{-7} is 0.011.

K-S value is to 1, the greater the distance between the distributions, and the more distinguishable they are. For my population of 721 models per distribution, K-S scores must be at 0.01 to be considered useful (see Figure 4.2).

In the ideal scenario, the distributions of metrics from abiotic and biotic populations of atmospheres would have zero overlap, and a K-S value close to 1 and a p-value close to zero, as this would allow the metric value to easily indicate the presence or absence of a biosphere, and provide high confidence in the detection.

4.4 Results

Network measures yielded promising results for detecting both biological and technological signatures in terrestrial atmospheres, though some measures were more effective at distinguishing between populations of atmospheres.

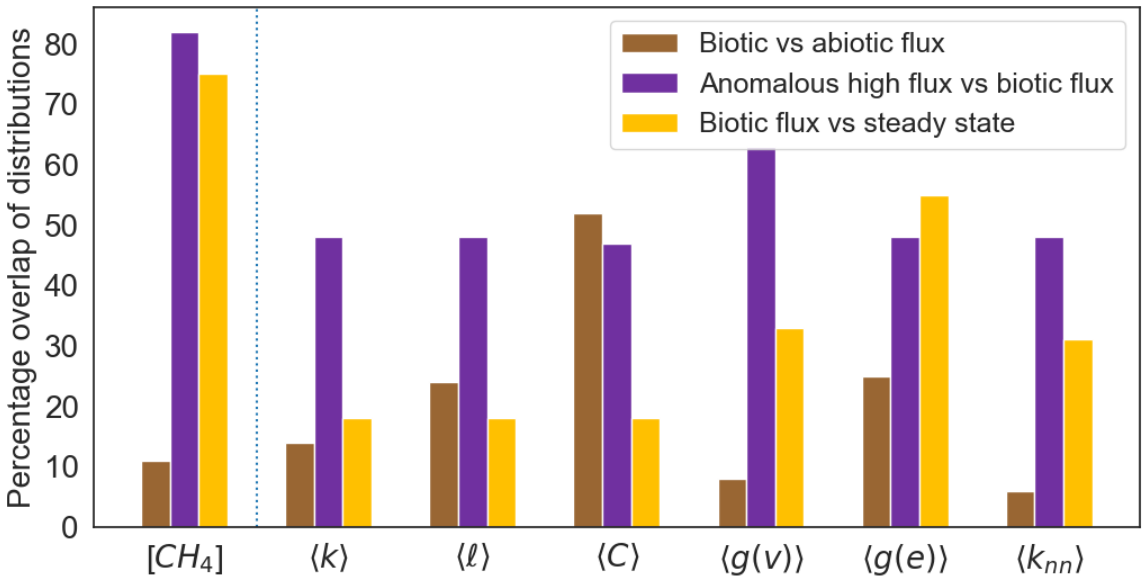


Figure 4.3: The percentage of overlap between distributions varies considerably on the source and abundance of CH_4 in the population of models; lower overlap indicates a lower risk of a false positive or negative. While CH_4 had less overlap when comparing the biotic flux and abiotic flux populations, network metrics generally had less overlap than CH_4 abundance when comparing between the biotic flux populations to the false positive populations, suggesting a lower chance of false positives or negatives. The blue dotted line is to visually separate CH_4 from network topological metrics.

4.4.1 Archean Earth-like Atmospheres and CH_4

The performance of network measures in correctly distinguishing abiotic and biotic scenarios varied depending on the source of CH_4, and the metric used, with some performing better than others.

CH_4 abundance. CH_4 abundance will be the metric most familiar to exoplanet scientists, and the only one that can be directly observed (see Figure 4.5). It performs well when comparing the abiotic flux population to the biotic flux, with an overlap percentage of only 11% and a K-S score of 0.86, p=0. However, it performs much more poorly when faced with the false positive scenarios, with an overlap percentage of 82% when comparing the biotic flux population to the anomalous high flux population and 75% when comparing the biotic flux population to the abiotic steady state

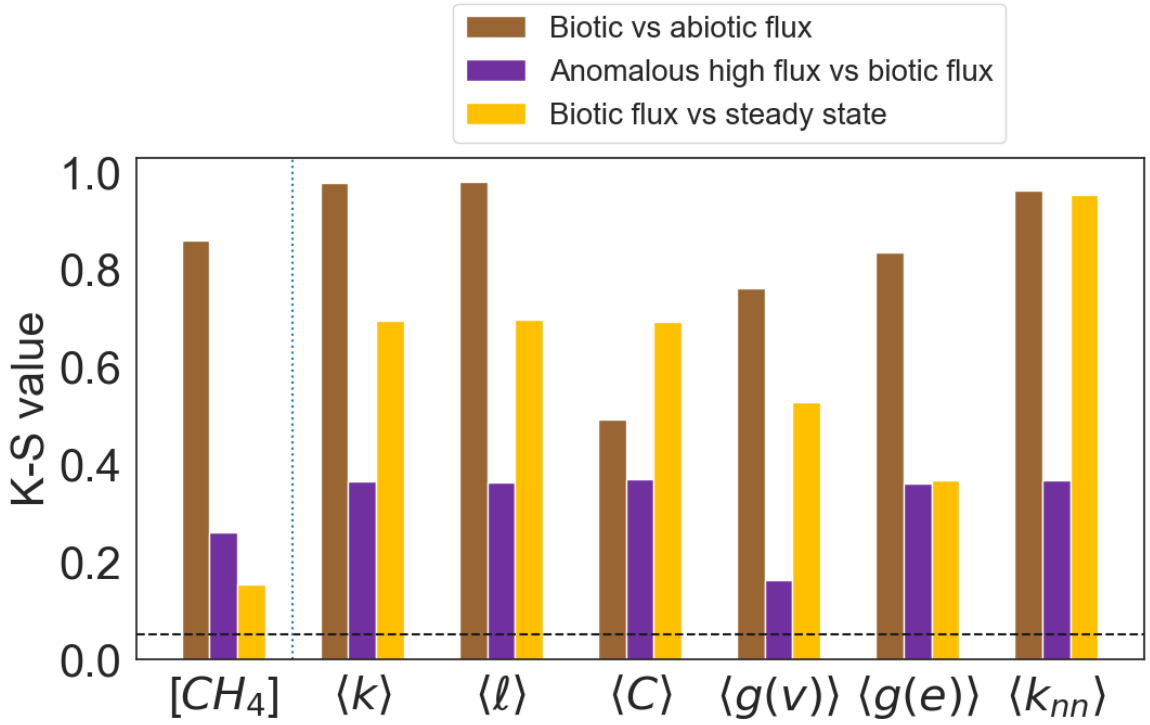


Figure 4.4: K-S scores of the biotic flux scenario as compared to the three abiotic models (abiotic flux, anomalous high flux, and abiotic steady state). Each bar indicates the discrimination of the corresponding metric in distinguishing between different modeled scenarios. The black dashed line indicates the minimum critical value for K-S scores, and the blue dashed line visually separates CH_4 abundance from network metrics. Network metrics performed well compared to CH_4 when distinguishing the biotic flux population from the false positive populations.

population. The K-S scores against the false positive scenarios were 0.26, p=0, and 0.15, p=9.8×10^{-8}, respectively, further demonstrating the difficulty in distinguishing between a true signal of life and a false positive using CH_4 abundance alone.

Mean degree. Mean degree (see Figure 4.6) exhibited similar behavior to CH_4 abundance, performing well when distinguishing between the biotic and abiotic flux scenarios; the two distributions only overlaped 14%, and the K-S test yielded a very high score of 0.98, p=0. However, mean degree was weaker in distinguishing between the distributions from the biotic flux populations and the anomalous high flux, overlapping 48% and having a K-S score of 0.36, p=0–though still better than CH_4 abundance. The metric performed better when comparing the biotic flux and abiotic

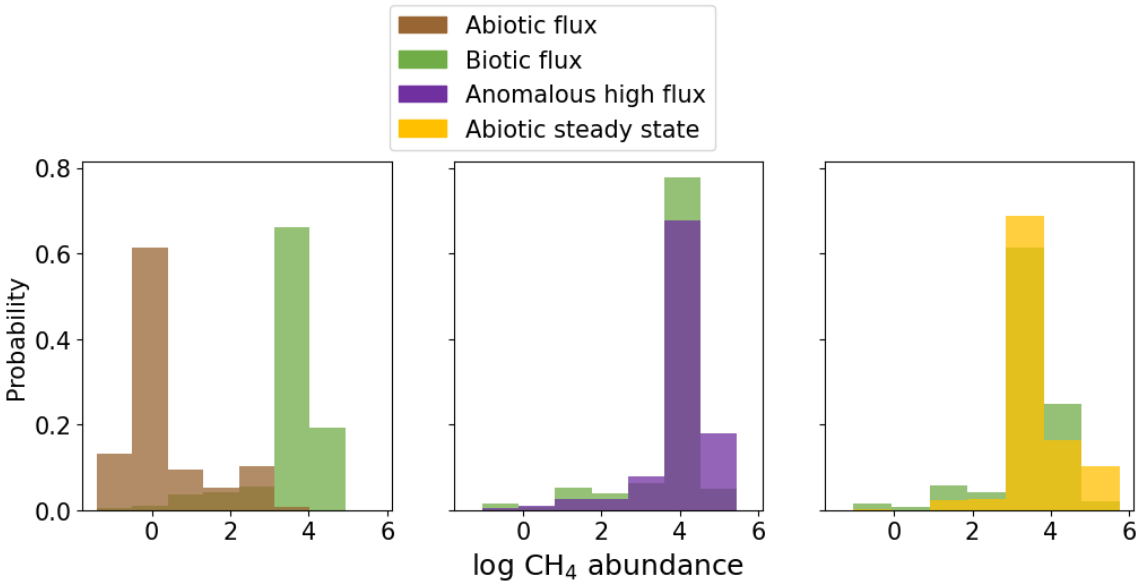


Figure 4.5: From left to right, the probability distributions of log CH_4 abundance from abiotic flux, anomalous high flux, and abiotic steady state populations of Archean Earth-like models, compared to the distribution from the biotic flux population. CH_4 abundance is useful in distinguishing between the abiotic and biotic flux scenarios, but harder to distinguish when comparing the biotic flux scenario to the false positive scenarios.

steady state, however, overlapping 18% and scoring a K-S metric of 0.69, p=0. These results are likely the result of mean degree's sensitivity to nodes with high reaction rates–in this case, CH_4. In situations where the CH_4 flux from the surface is high, this drives the mean degree value upwards, regardless of whether or not life is present.

Average shortest path length. Average shortest path length (see Figure 4.7), like mean degree, is sensitive to nodes involved in reactions with high rates, and thus displayed good distinguishingability when comparing the biotic flux population with the abiotic flux population (distribution overlap 24%, K-S score 0.98, p=0) and the abiotic steady state population (overlap 18%, K-S score 0.70, p=0), but did not distinguish as well between the biotic flux population and the anomalous high flux population (overlap 47%, K-S score, p=0). However, it's worth noting that average

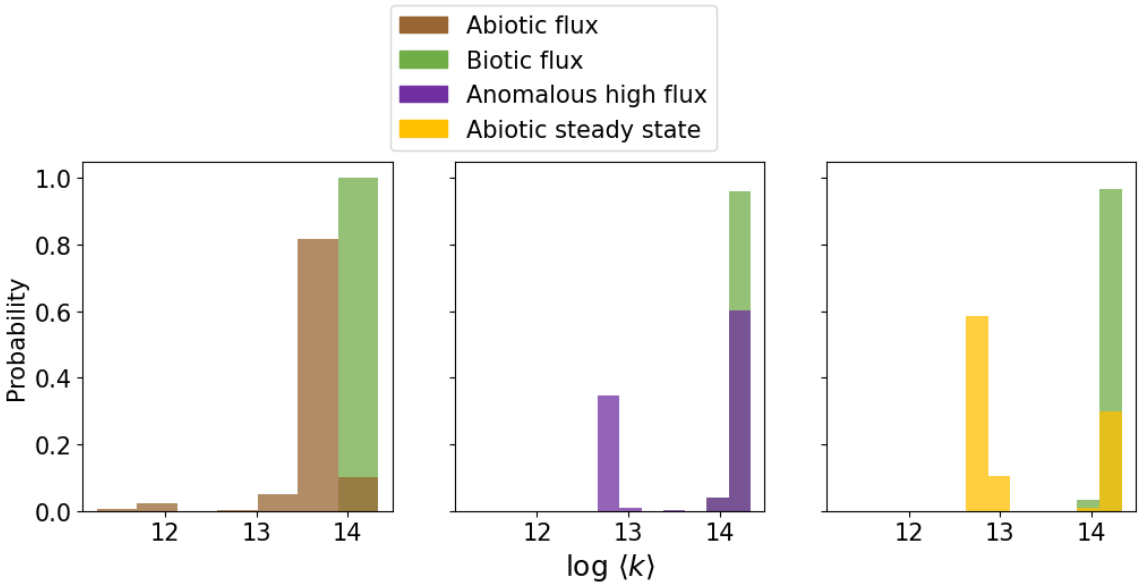


Figure 4.6: From left to right, the probability distributions of log mean degree from abiotic flux, anomalous high flux, and abiotic steady state populations of Archean Earth-like models, compared to the distribution from the biotic flux population. Mean degree is useful in distinguishing between the biotic scenario and the abiotic flux and steady state scenarios, but doesn't perform as well when comparing the biotic flux population to the anomalous high flux population. Nonetheless, it still outperforms CH_4 abundance alone.

shortest path length often performs worse than CH_4 abundance alone.

Average clustering coefficient. Average clustering coefficient (see Figure 4.8) did not perform particularly well when comparing the biotic flux population to the abiotic flux and anomalous high flux populations, having a distribution overlap percentage of 52% and 47%, respectively, with K-S scores of 0.49, p=0, and 0.37, p=0. It did slightly better when comparing the biotic flux scenarios to the abiotic steady state scenarios, having an overlap percentage of only 18% and a K-S score of 0.69, p=0. When compared to CH_4, this distinguishability was on par with mean degree.

Node betweenness centrality. Node betweenness centrality (see Figure 4.9) showed high distinguishing power between the abiotic and biotic flux populations, with an overlap of 8% and a K-S score of 0.76, p=0. However, it performed more worse than any other metric save CH_4 when comparing between the biotic flux population and

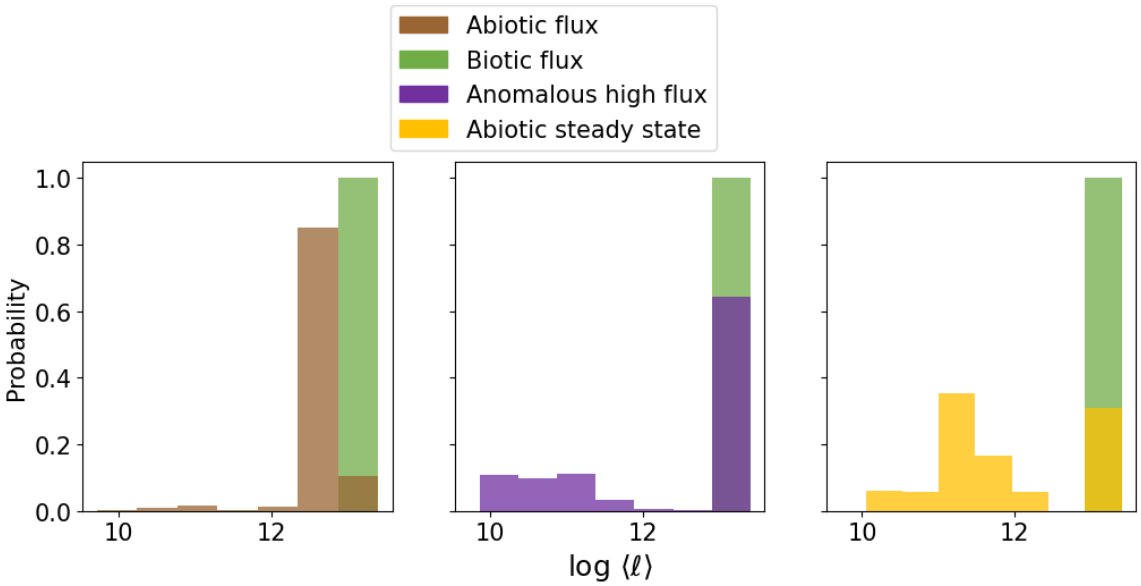


Figure 4.7: From left to right, the probability distributions of log average shortest path length from abiotic flux, anomalous high flux, and abiotic steady state populations of Archean Earth-like models, compared to the distribution from the biotic flux population. Like mean degree, average shortest path length is useful in distinguishing between the biotic scenario and the abiotic flux and steady state scenarios, but doesn't perform as well when comparing the biotic flux population to the anomalous high flux population. Furthermore, it performs worse than CH_4 abundance in the former two scerarios.

the anomalous high flux population, with an overlap of 63% and a K-S score of 0.16, p=7.6×10^{-9}. Node betweenness centrality did a little better when comparing biotic flux to abiotic steady state, with an overlap of 33% and a K-S score of 0.52, p=0, though still does not provide the same degree of distinguishability as mean degree or average shortest path length.

Edge betweenness centrality. Edge betweenness centrality (see Figure 4.10) performed slightly better compared to node betweenness centrality, with an overlap of 15% and a K-S score of 0.83, p=0, distinguishing between the abiotic and biotic flux cases; 47% and 0.36, p=0, between the anomalous high flux and biotic flux cases; and 54% and 0.37,p=0 between the abiotic steady state and biotic flux cases.

Average neighbor degree. Average neighbor degree was one of the best performing

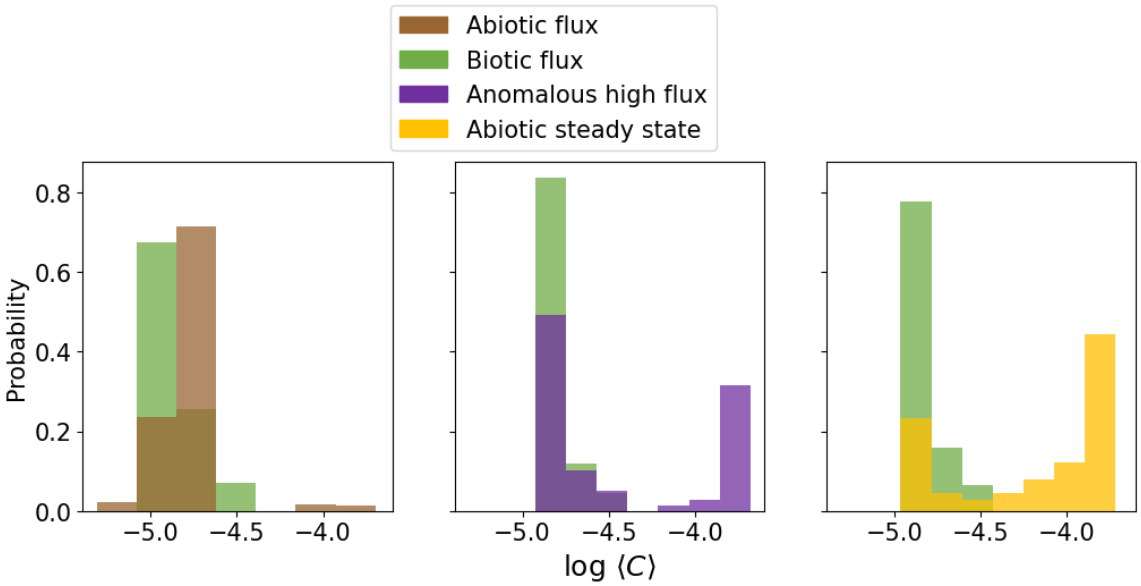


Figure 4.8: From left to right, the probability distributions of log average clustering coefficient from abiotic flux, anomalous high flux, and abiotic steady state populations of Archean Earth-like models, compared to the distribution from the biotic flux population. Clustering coefficient performs relatively well when distinguishing between the biotic flux and abiotic steady state populations, but does not perform very well when comparing between the other populations; overall, its performance is comparable to mean degree when compared to that of CH_4.

metrics when comparing the abiotic flux and biotic flux populations, with an overlap of only 6% and K-S score of 0.96, p=0. It exhibited the lowest amount of overlap of any metric when comparing the distributions from the anomalous high flux and biotic flux populations, at 18%, though its K-S score was only 0.36, p=0. In the case of distinguishing between abiotic steady state and biotic flux, this was reversed, as the distributions has 31% overlap, but a K-S score of 0.95, p=0. Nonetheless, this suggests that average neighbor degree may be particularly useful compared to CH_4 abundance when trying to distinguish between biotic and false positive scenarios using atmospheric CRN metrics, particularly versus the abiotic steady state case.

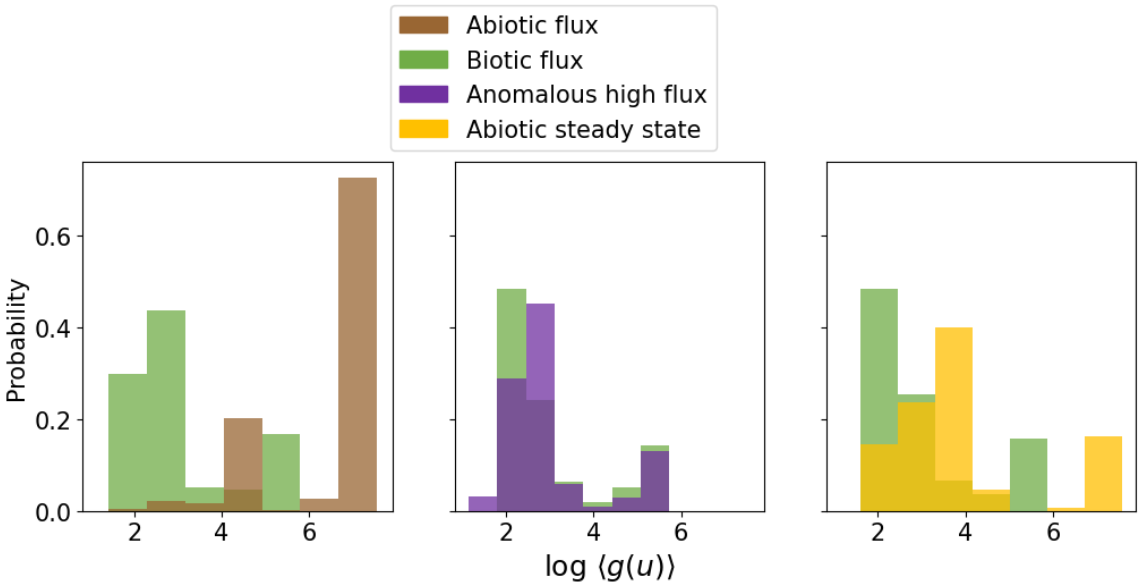


Figure 4.9: From left to right, the probability distributions of log node betweenness centrality from abiotic flux, anomalous high flux, and abiotic steady state populations of Archean Earth-like models, compared to the distribution from the biotic flux population. Node betweenness centrality performed poorly compared to other metrics, regarding comparison between the biotic flux and anomalous high flux populations, though it still performed better than CH_4 when comparing the biotic flux and anomalous high flux and abiotic steady state populations.

4.4.2 Perturbation Testing

In the perturbation test cases, where C_2H_2 was specifically removed from the abiotic flux models, network metrics consistently had a lower overlap of distribution and higher K-S score than CH_4; the differences were even more distinct when comparing the biotic flux case to the perturbed biotic flux case, where the overlap was as low as zero (see Figures 4.12 and 4.13, as well as Figures C.4-C.10 in the Appendix). The K-S scores were also quite high, especially between the perturbed and non-perturbed biotic populations, where mean degree and average shortest path length both score 1.0, p=0.

The fact that distinguishability is stronger regarding the biotic cases is likely due to the increased flux of CH_3 derived from the higher surface flux of CH_4, which

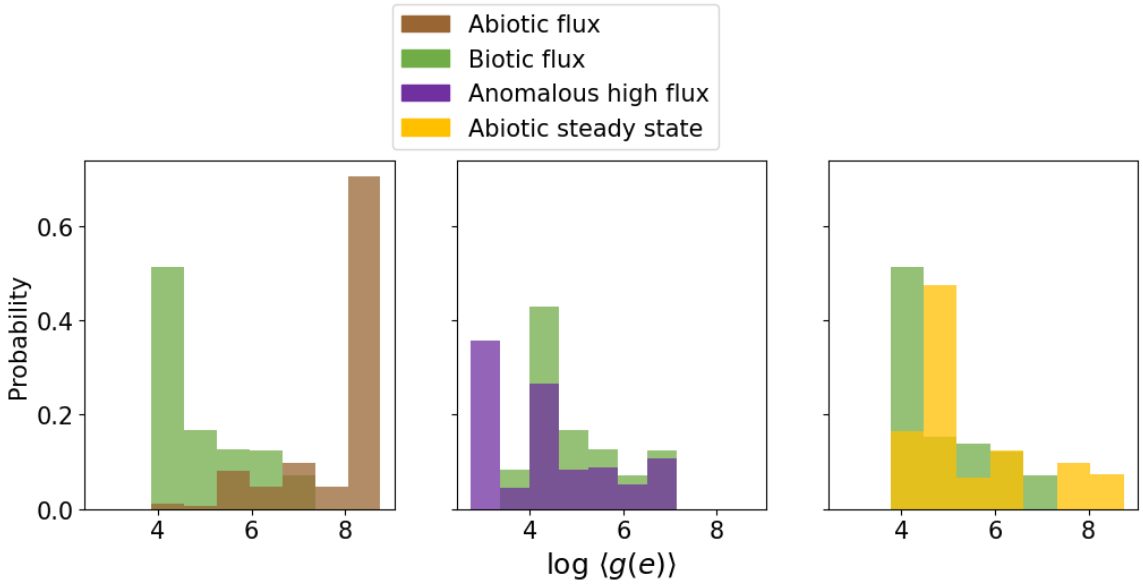


Figure 4.10: From left to right, the probability distributions of log edge betweenness centrality from abiotic flux, anomalous high flux, and abiotic steady state populations of Archean Earth-like models, compared to the distribution from the biotic flux population. Edge betweenness also performed poorly compared to the other metrics (save CH_4).

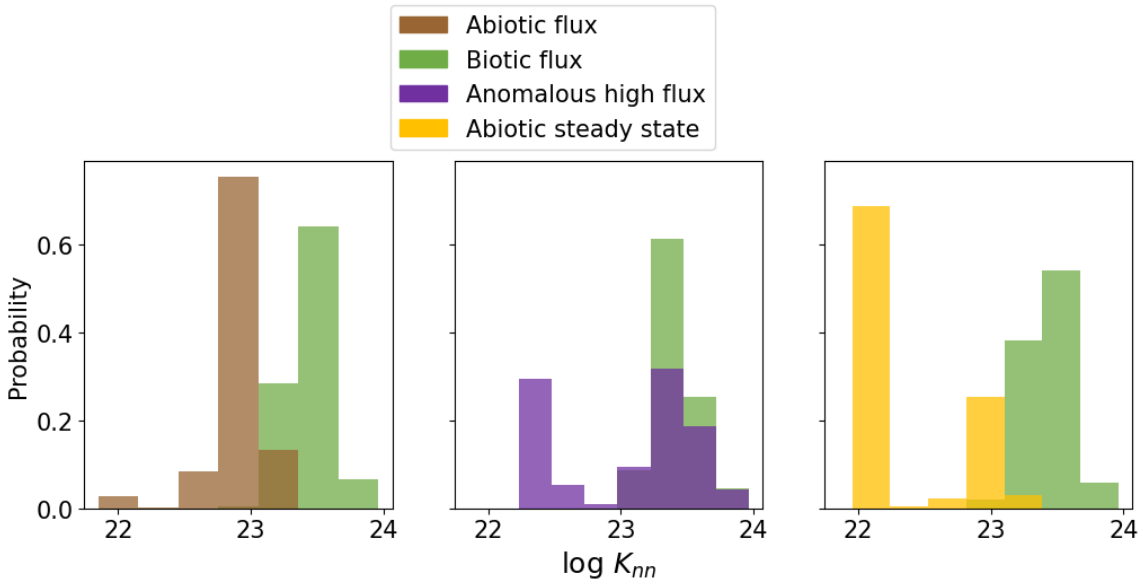


Figure 4.11: From left to right, the probability distributions of log average neighbor degree from abiotic flux, anomalous high flux, and abiotic steady state populations of Archean Earth-like models, compared to the distribution from the biotic flux population. Average neighbor degree seems to perform particularly well when comparing biotic and false positive scenarios when compared to CH_4.

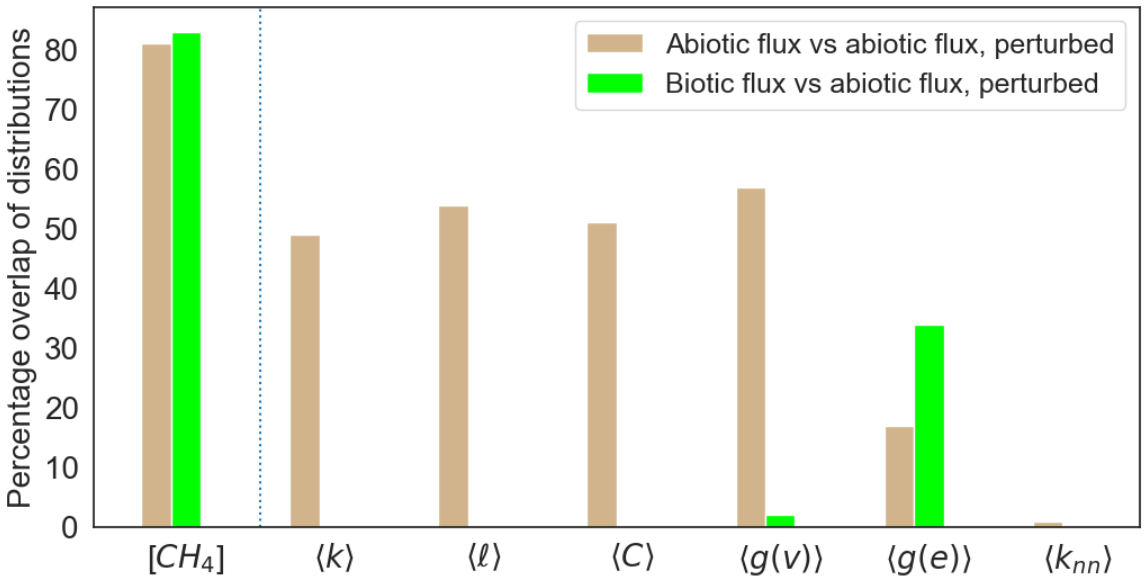


Figure 4.12: The percentage of overlap of the abiotic and biotic flux distributions versus the distributions drawn from the perturbed versions of their models. Network metrics had considerably less overlap than CH_4–in many cases, as low as zero percent overlap.

is no longer being removed by conversion into C_2H_2 and then to haze. While the CH_4 concentrations were largely the same regardless of whether the model had been perturbed or not, likely due to another sink of CH_4 coming into play to replace those related to C_2H_2, the knock-on effects on reaction rates caused by increased CH_3 and different sinks led to considerable changes in the weighted topology of the network.

This perturbation testing demonstrates that not only network metrics indicative of changes in the chemical system of an atmosphere (or otherwise they would be unable to differentiate between the perturbed and non-perturbed models), but they can detect these changes even when the abundance and distribution of a species under observation, such as CH_4, is seemingly unaffected by them. This further builds confidence in the usefulness of network metrics for analyzing planetary atmospheres.

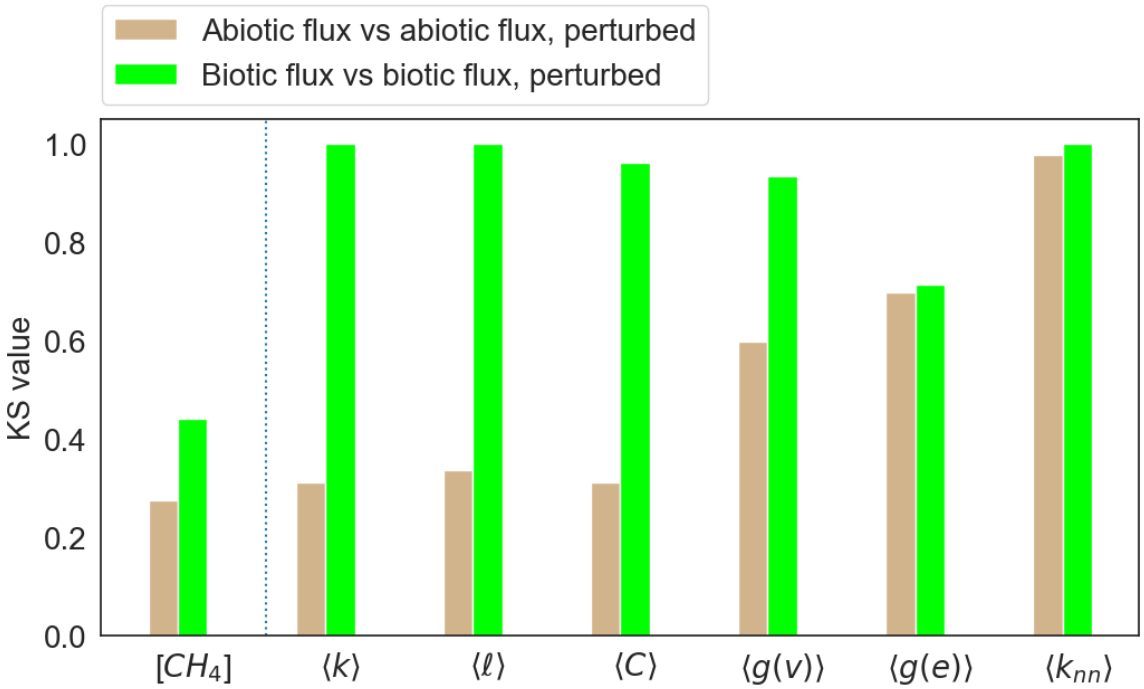


Figure 4.13: K-S scores of the abiotic and biotic flux distributions compared to the distributions drawn from the perturbed versions of their models. Each bar indicates the discrimination of the corresponding metric in distinguishing between different modeled scenarios. Network metrics performed considerably better than CH_4 abundance alone.

4.4.3 Modern Earth-like Atmospheres and CFC-12

Network metrics also showed promise in distinguishing between modern Earth analogue atmospheres with and without CFC-12 (see Figures 4.14 and 4.15); mean degree performed particularly well.

Mean degree. Mean degree (see Figure 4.16) was the best performing metric when comparing populations with CFC-12 versus those without, with an overlap of only 16% and a K-S score of 0.85, p=0. This may be due to the high reactivity of the Cl radical–even a small amount can affect multiple other reactions and their rates.

Average shortest path length. Curiously, average shortest path length (see Figure 4.17) did not perform as well as mean degree in distinguishing between populations with CFC-12 and those without, with an overlap between the distributions of 29%

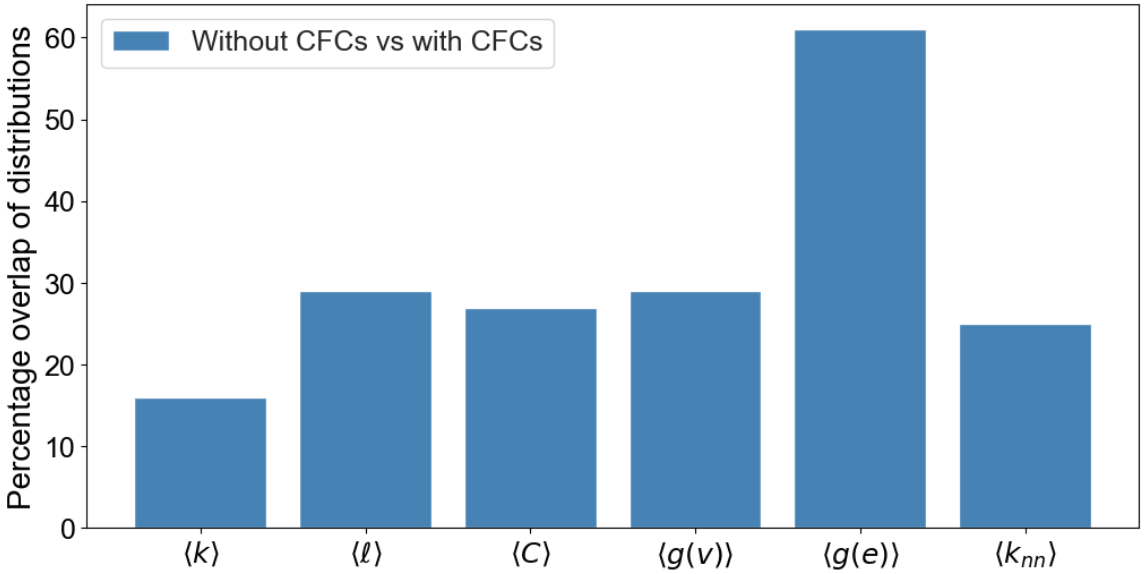


Figure 4.14: The percentage of overlap between distributions varies considerably, though mean degree and average neighbor degree showed the lowest overlap.

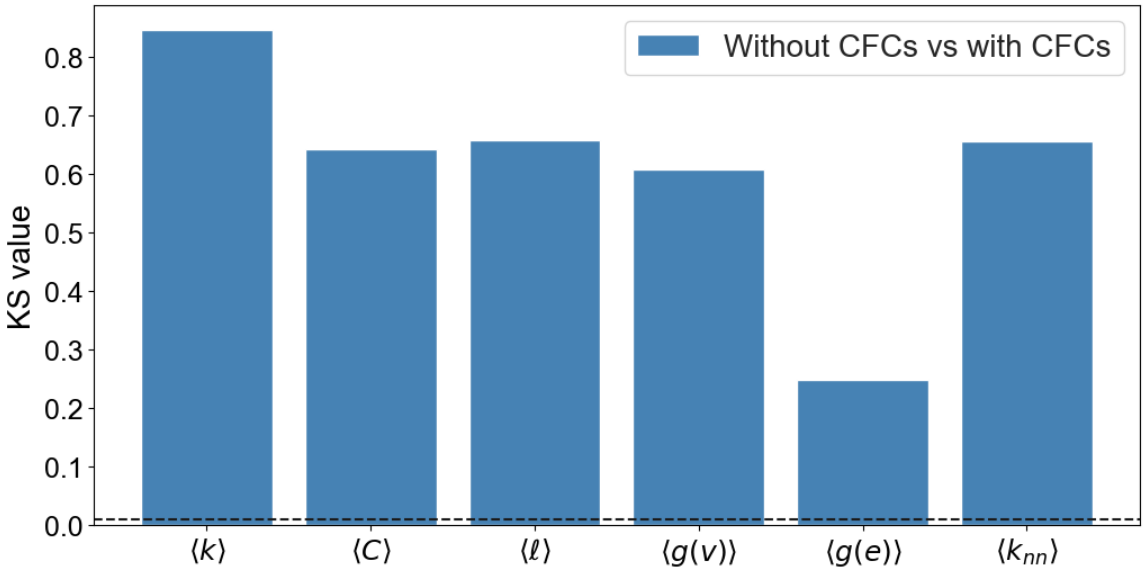


Figure 4.15: K-S scores quantifying the distinguishability between network metric distributions drawn from modern Earth-like atmosphere populations, with and without CFC-12. Mean degree, in particular, performed especially well.

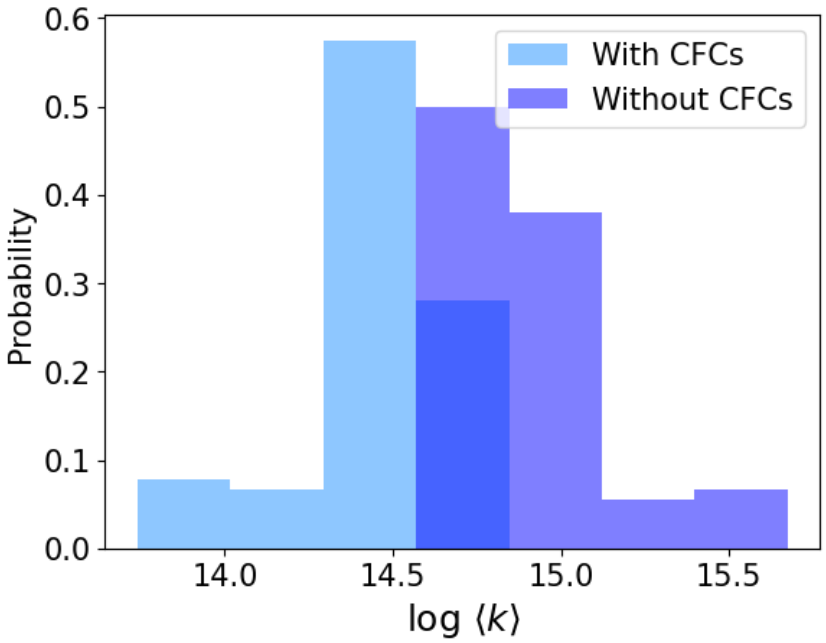


Figure 4.16: The probability distributions of mean degree from population of modern Earth-like atmospheres, with and without the addition of CFC-12, showing mean degree's strong performance in distinguishing between the two.

and a K-S score of 0.66, p=0.

Average clustering coefficient and *node betweenness centrality.* Average clustering coefficient (see Figure 4.18) and node betweenness centrality (see Figure 4.19) showed similar performance, with an overlap between distributions of 27 and 29%, respectively, and a K-S score of 0.63, p=0, and 0.61, p=0.

Edge betweenness centrality. Edge betweenness centrality (see Figure 4.20) performed the worst out of the metrics used, with a percentage overlap of 61% and a K-S score of only 0.25, p=0.

Average neighbor degree. Average neighbor degree (see Figure 4.21) performed slightly better than average clustering coefficient and node betweenness centrality, with an overlap of 25% and a K-S score of 0.66, p=0.

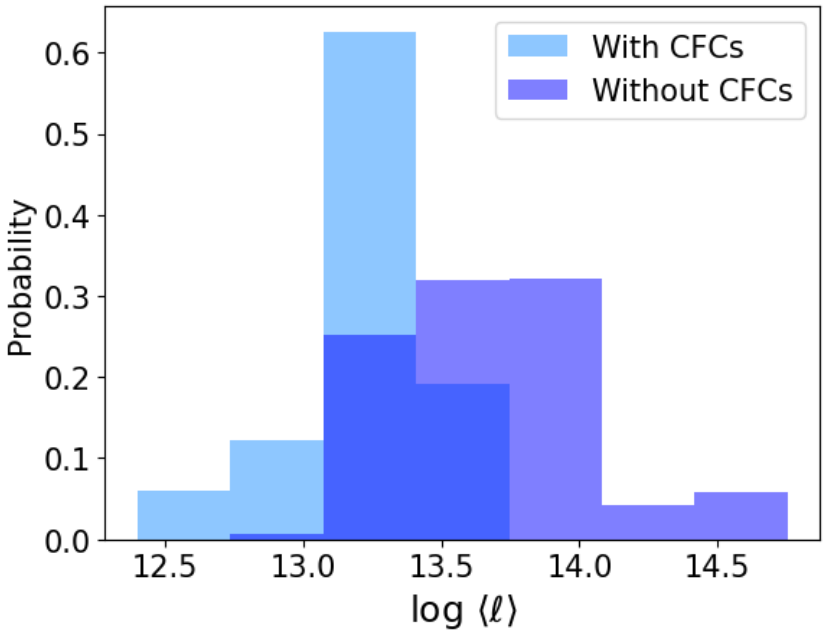


Figure 4.17: The probability distributions of average shortest path length from population of modern Earth-like atmospheres, with and without the addition of CFC-12. Average shortest path length offered middling performance compared to mean degree.

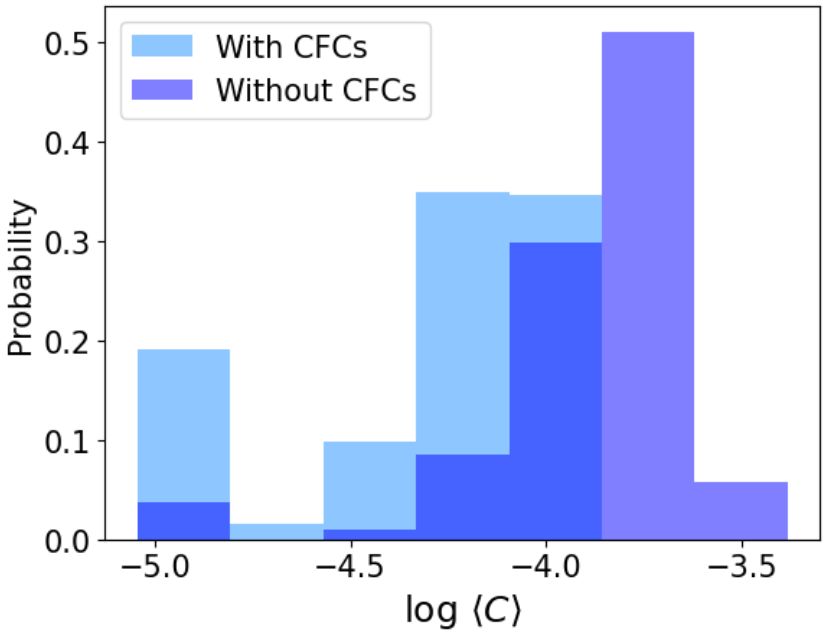


Figure 4.18: The probability distributions of average clustering coefficient from population of modern Earth-like atmospheres, with and without the addition of CFC-12. Like average shortest path length, average clustering coefficient offered mediocre performance compared to mean degree.

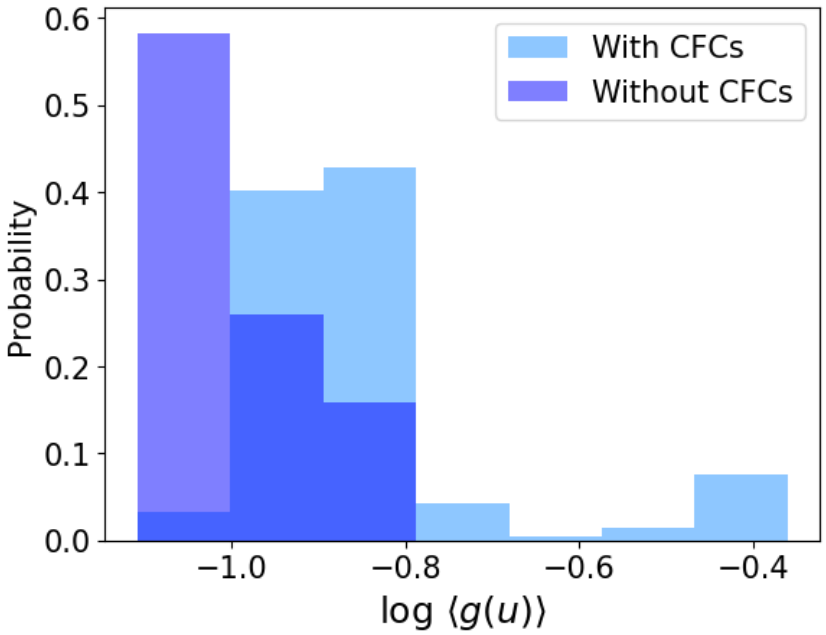


Figure 4.19: The probability distributions of node betweenness centrality from population of modern Earth-like atmospheres, with and without the addition of CFC-12. Node betweenness centrality performed similarly to average clustering coefficient.

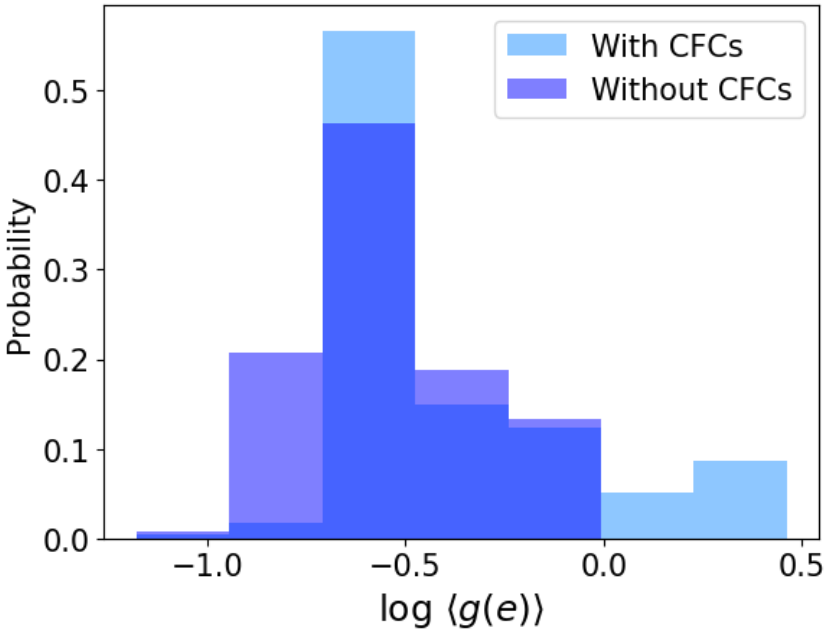


Figure 4.20: The probability distributions of edge betweenness centrality from population of modern Earth-like atmospheres, with and without the addition of CFC-12. Of all the metrics, edge betweenness centrality performed the worst.

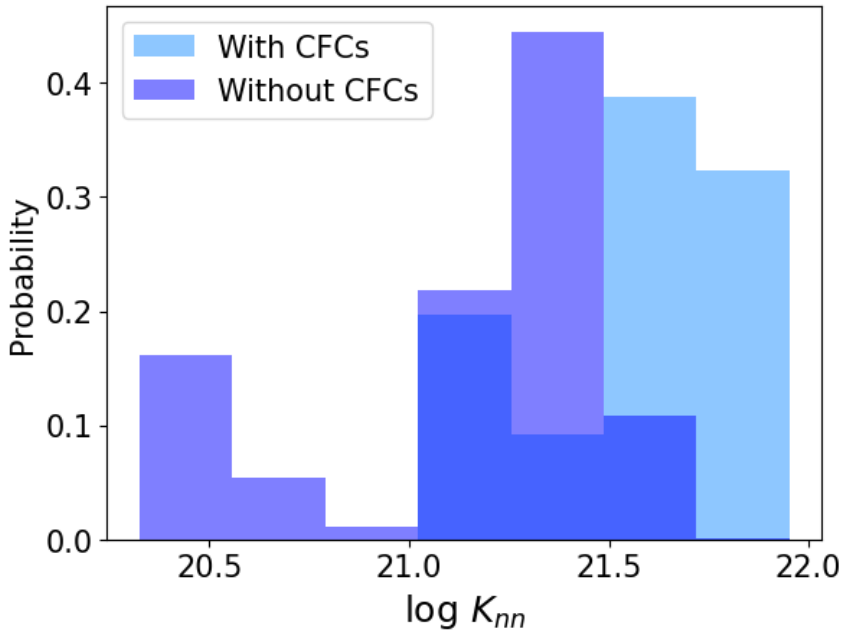


Figure 4.21: The probability distributions of average neighbor degree from population of modern Earth-like atmospheres, with and without the addition of CFC-12. Average neighbor degree performed slightly better than average clustering coefficient and node betweenness centrality.

4.5 Discussion

In both the Archean Earth and modern Earth models, the topology of the chemical reaction exhibited quantitative changes to the introduction of a biogenic or technogenic reaction pathway. In both cases, the metrics that performed the best in distinguishing between atmospheres with biotic or technological components and those without, mean degree and average neighbor degree, were the most sensitive to changes in chemical rates. However, it is clear that it's not just the rates that the metrics are indicative of, as they were still able to distinguish between abiotic scenarios with similar CH_4 fluxes or concentrations as the biotic scenario with more accuracy than CH_4 alone.

Introducing network topology analysis to atmospheric characterization pipelines could prove useful to complement in assessing exoplanet biosignatures. This is partic-

ularly true in cases where the underlying biosphere or technosphere may be unknown or poorly characterized and I have no choice but to rely on limited observational data. I hope to investigate further the utility of this approach by testing it on a wider variety of atmospheres and biospheres, such as planets with an abiotic O_2 flux (Kleinböhl *et al.*, 2018), or the complex chemistry of Titan-like worlds (Willacy *et al.*, 2016).

The next step is to determine whether the network topology of an atmosphere can be directly inferred from exoplanet spectra, without needing to construct a network and analyze it. If the presence, absence, and/or attenuation of specific spectral lines can be tied to network topology, then these spectral signatures – which may not be directly dependent on the observed abundances of biologically relevant gasses – would be a powerful addition to the toolbox of exoplanet-focused astrobiologists. Developing new tools for the analysis and interpretation of exoplanet atmospheric data becomes particularly important in the scenarios investigated in this paper: if I assume a 10 ppm-noise floor for JWST Near-Infrared Spectrometer or Mid-Infrared Instrument, none of the output (CH_4 or CFCs) concentrations from the simulations reach 1 ppm, demanding a large number of co-added transit time (100 h according to Haqq-Misra et al. (2022) in the case of CFCs), or relegating direct abundance inferences to long-term future observatories such as LUVOIR or Habex (or a hybrid between the two, as suggested by the ASTRO 2020 Decadal Survey).

If network topology does not affect observed spectral lines directly and such inference is not possible, then the next challenge is to determine what the minimum network size (both in terms of species identified, and uncertainties in abundance) is required to meaningfully assess an exoplanet's atmosphere characteristics. This, in turn, can provide constraints for future observation campaigns, such as minimum spectral resolution or optimal combinations of observations required to detect the a sufficient number of species and their abundances to construct a useful chemical

reaction network.

Beyond the implications for biosignature detection, the influence of biology on atmospheric chemical reaction network topology may provide a window into the physics of life itself. If there is, as discussed in Section 1.3.1, a 'universal biology' dictated by the physical constraints of the universe (Mariscal and Fleming, 2018), one manifestation of it may be in its network topology (Solé *et al.*, 2002). Given how tightly coupled the biosphere is to the atmosphere on Earth, it is not unreasonable for the CRN topology of the biosphere to influence the CRN topology of the atmosphere. For instance, one might model planetary evolution as a multi-layer network, where each layer represents the chemistry in the geosphere, biosphere, or atmosphere (and technosphere for technological planets). The impact of industrial emissions on CRN topology is harder to explain in this manner, though similarities in CRN topology between biological and technological systems are not unprecedented (Solé *et al.*, 2011). In any case, further investigation into atmospheric CRN topology is warranted, with interesting implications in a variety of fields and investigation topics; see Section 5.1.2 for more detail on these possibilities.

4.6 Conclusion

I have demonstrated that the topology of atmospheric chemical reaction network may provide the tools for higher confidence in the detection of atmospheric biosignatures and technosignatures than relying on atmospheric species alone. While further research is required, these findings could greatly reduce the risk of false positives and negatives when searching for life beyond my solar system.

REFERENCES

Arney, G., S. D. Domagal-Goldman, V. S. Meadows, E. T. Wolf, E. Schwieterman, B. Charnay, M. Claire, E. Hébrard and M. G. Trainer, "The pale orange dot: the

spectrum and habitability of hazy Archean Earth", Astrobiology 16, 11, 873–899 (2016).

Arney, G. N., "The k dwarf advantage for biosignatures on directly imaged exoplanets", The Astrophysical Journal Letters 873, 1, L7 (2019).

Balbi, A. and M. M. Ćirković, "Longevity is the key factor in the search for technosignatures", The Astronomical Journal 161, 5, 222 (2021).

Bella, A., A. Choprab, W. Fawcettc and R. Talebid, "NASA Frontier Development Lab technical memorandum", (2018).

Broadley, M. W., P. H. Barry, C. J. Ballentine, L. A. Taylor and R. Burgess, "End-permian extinction amplified by plume-induced release of recycled lithospheric volatiles", Nature Geoscience 11, 9, 682–687 (2018).

Catling, D. C. and K. J. Zahnle, "The archean atmosphere", Science Advances 6, 9, eaax1420 (2020).

Domagal-Goldman, S. D., A. Segura, M. W. Claire, T. D. Robinson and V. S. Meadows, "Abiotic ozone and oxygen in atmospheres similar to prebiotic Earth", The Astrophysical Journal 792, 2, 90 (2014).

Etiope, G. and B. Sherwood Lollar, "Abiotic methane on earth", Reviews of Geophysics 51, 2, 276–299 (2013).

FDL, N., "NASA Frontier Development Labs (FDL) Project", URL https://exoplanetarchive.ipac.caltech.edu/cgi-bin/FDL/nph-fdl?atmos (2019).

Fisher, T., H. Kim, C. Millsaps, M. Line and S. I. Walker, "Inferring exoplanet disequilibria with multivariate information in atmospheric reaction networks", The Astronomical Journal 164, 2, 53 (2022).

Fujii, Y., D. Angerhausen, R. Deitrick, S. Domagal-Goldman, J. L. Grenfell, Y. Hori, S. R. Kane, E. Pallé, H. Rauer, N. Siegler, K. Stapelfeldt and K. B. Stevenson, "Exoplanet biosignatures: observational prospects", Astrobiology 18, 6, 739–778 (2018).

Guzmán-Marmolejo, A., A. Segura and E. Escobar-Briones, "Abiotic production of methane in terrestrial planets", Astrobiology 13, 6, 550–559 (2013).

Hagberg, A., P. Swart and D. S Chult, "Exploring network structure, dynamics, and function using NetworkX", Tech. rep., Los Alamos National Lab.(LANL), Los Alamos, NM (United States) (2008).

Haqq-Misra, J., T. J. Fauchez, E. W. Schwieterman and R. Kopparapu, "Disruption of a planetary nitrogen cycle as evidence of extraterrestrial agriculture", The Astrophysical Journal Letters 929, 2, L28 (2022a).

Haqq-Misra, J., R. Kopparapu, T. J. Fauchez, A. Frank, J. T. Wright and M. Lingam1, "Detectability of chlorofluorocarbons in the atmospheres of habitable M-dwarf planets", arXiv:2202.05858 ArXiv: 2202.05858 (2022b).

Haqq-Misra, J., E. W. Schwieterman, H. Socas-Navarro, R. Kopparapu, D. Angerhausen, T. G. Beatty, S. Berdyugina, R. Felton, S. Sharma, G. G. De la Torre and D. Apai, "Searching for technosignatures in exoplanetary systems with current and future missions", Acta Astronautica 198, 194–207 (2022c).

Harman, C. E. and S. Domagal-Goldman, "Biosignature false positives", in "Handbook of Exoplanets", edited by H. J. Deeg and J. A. Belmonte, pp. 1–22 (Springer International Publishing, Cham, 2018).

Kaltenegger, L., Z. Lin and S. Rugheimer, "Finding signs of life on transiting Earthlike planets: high-resolution transmission spectra of Earth through time around FGKM host stars", The Astrophysical Journal 904, 1, 10 (2020).

Kasting, J. F. and J. L. Siefert, "Life and the evolution of earth's atmosphere", Science 296, 5570, 1066–1068 (2002).

Kleinböhl, A., K. Willacy, A. J. Friedson, P. Chen and M. R. Swain, "Buildup of abiotic oxygen and ozone in moist atmospheres of temperate terrestrial exoplanets and its impact on the spectral fingerprint in transit observations", The Astrophysical Journal 862, 2, 92 (2018).

Klobas, J. E. and D. Wilmouth, "Volcanogenic chlorofluorocarbons and the recent cfc anomalies", (2019).

Kopparapu, R., G. Arney, J. Haqq-Misra, J. Lustig-Yaeger and G. Villanueva, "Nitrogen dioxide pollution as a signature of extraterrestrial technology", The Astrophysical Journal 908, 2, 164 (2021).

Lin, H. W., G. G. Abad and A. Loeb, "Detecting industrial pollution in the atmospheres of Earth-like exoplanets", The Astrophysical Journal 792, 1, L7 (2014).

Lincowski, A. P., J. Lustig-Yaeger and V. S. Meadows, "Observing isotopologue bands in terrestrial exoplanet atmospheres with the James Webb Space Telescope: implications for identifying past atmospheric and ocean loss", The Astronomical Journal 158, 1, 26 (2019).

Mariscal, C. and L. Fleming, "Why we should care about universal biology", Biological Theory 13, 2, 121–130 (2018).

Meadows, V. S., "Reflections on O_2 as a biosignature in exoplanetary atmospheres", Astrobiology 17, 10, 1022–1052 (2017).

Meadows, V. S., C. T. Reinhard, G. N. Arney, M. N. Parenteau, E. W. Schwieterman, S. D. Domagal-Goldman, A. P. Lincowski, K. R. Stapelfeldt, H. Rauer, S. DasSarma, S. Hegde, N. Narita, R. Deitrick, J. Lustig-Yaeger, T. W. Lyons, N. Siegler and J. L. Grenfell, "Exoplanet biosignatures: understanding oxygen as a biosignature in the context of its Environment", Astrobiology 18, 6, 630–662 (2018).

Rein, H., Y. Fujii and D. S. Spiegel, "Some inconvenient truths about biosignatures involving two chemical species on Earth-like exoplanets", Proceedings of the National Academy of Sciences 111, 19, 6871–6875 (2014).

Reinhard, C. T., S. L. Olson, E. W. Schwieterman and T. W. Lyons, "False negatives for remote life detection on ocean-bearing planets: lessons from the early Earth", Astrobiology 17, 4, 287–297 (2017).

Sagan, C., W. R. Thompson, R. Carlson, D. Gurnett and C. Hord, "A search for life on Earth from the Galileo spacecraft", Nature 365, 6448, 715–721 (1993).

Schwieterman, E. W., N. Y. Kiang, M. N. Parenteau, C. E. Harman, S. DasSarma, T. M. Fisher, G. N. Arney, H. E. Hartnett, C. T. Reinhard, S. L. Olson, V. S. Meadows, C. S. Cockell, S. I. Walker, J. L. Grenfell, S. Hegde, S. Rugheimer, R. Hu and T. W. Lyons, "Exoplanet biosignatures: a review of remotely detectable signs of life", Astrobiology 18, 6, 663–708 (2018).

Seager, S., W. Bains and R. Hu, "A biomass-based model to estimate the plausibility of exoplanet biosignature gases", The Astrophysical Journal 775, 2, 104 (2013).

Sheikh, S., A. Berea, R. Davis, G. G. De la Torre, J. DeMarines, T. Fisher, S. Foote, D. Gelino, D. Grinspoon, G. Profitiliotis *et al.*, "Technosignatures as a priority in planetary science", Bulletin of the American Astronomical Society 53, 4, 427 (2021).

Solé, R. V., S. A. Levin, J. H. Brown, V. K. Gupta, B.-L. Li, B. T. Milne, C. Restrepo and G. B. West, "The fractal nature of nature: power laws, ecological complexity and biodiversity", Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 357, 1421, 619–626 (2002).

Solé, R. V. and A. Munteanu, "The large-scale organization of chemical reaction networks in astrophysics", Europhysics Letters (EPL) 68, 2, 170–176 (2004).

Solé, R. V., S. Valverde and C. Rodriguez-Caso, "Convergent evolutionary paths in biological and technological networks", Evolution: Education and Outreach 4, 3, 415–426 (2011).

Thompson, M. A., J. Krissansen-Totton, N. Wogan, M. Telus and J. J. Fortney, "The case and context for atmospheric methane as an exoplanet biosignature", Proceedings of the National Academy of Sciences 119, 14, e2117933119 (2022).

Tian, M. and K. Heng, "Atmospheric chemistry of secondary and hybrid atmospheres of super earths and sub-neptunes", arXiv preprint arXiv:2301.10217 (2023).

Tijhuis, L., M. C. Van Loosdrecht and J. Heijnen, "A thermodynamically based correlation for maintenance gibbs energy requirements in aerobic and anaerobic chemotrophic growth", Biotechnology and Bioengineering 42, 4, 509–519 (1993).

Tremblay, L., M. R. Line, K. B. Stevenson, T. Kataria, R. T. Zellem, J. J. Fortney and C. V. Morley, "The detectability and constraints of biosignature gases in the near & mid-infrared from transit transmission spectroscopy", arXiv:1912.10939 [astro-ph] (2019).

Virtanen, P., R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt and SciPy 1.0 Contributors, "SciPy 1.0: fundamental algorithms for scientific computing in Python", Nature Methods 17, 261–272 (2020).

Visscher, H., C. V. Looy, M. E. Collinson, H. Brinkhuis, J. H. Van Konijnenburg-Van Cittert, W. M. Kürschner and M. A. Sephton, "Environmental mutagenesis during the end-permian ecological crisis", Proceedings of the National Academy of Sciences 101, 35, 12952–12956 (2004).

Walker, S. I., W. Bains, L. Cronin, S. DasSarma, S. Danielache, S. Domagal-Goldman, B. Kacar, N. Y. Kiang, A. Lenardic, C. T. Reinhard, W. Moore, E. W. Schwieterman, E. L. Shkolnik and H. B. Smith, "Exoplanet biosignatures: future directions", Astrobiology 18, 6, 779–824 (2018).

Walker, S. I., L. Cronin, A. Drew, S. Domagal-Goldman, T. Fisher, M. Line and C. Millsaps, "Probabilistic biosignature frameworks", in "Planetary Astrobiology", p. 477 (University of Arizona Press, 2020).

Willacy, K., M. Allen and Y. Yung, "A new astrobiological model of the atmosphere of titan", The Astrophysical Journal 829, 2, 79 (2016).

Wong, M. L., A. Prabhu, J. Williams, S. M. Morrison and R. M. Hazen, "Toward network-based planetary biosignatures: atmospheric chemistry as unipartite, unweighted, undirected networks", Journal of Geophysical Research: Planets 128, 6, e2022JE007658 (2023).

Wordsworth, R. and R. Pierrehumbert, "Abiotic oxygen-dominated atmospheres on terrestrial habitable zone planets", The Astrophysical Journal Letters 785, 2, L20 (2014).

Wright, J. T., "Exoplanets and SETI", in "Handbook of Exoplanets", pp. 3405–3412 (Springer International Publishing, 2018).

Chapter 5

CONCLUSION

Can atmospheric network topology be used to distinguish between populations of abiotic and biotic worlds better than atmospheric composition alone? This has been the central question motivating this work, and thus far, the answer appears to be "yes." First, I have demonstrated how the topology of chemical reaction networks can be analyzed and quantified, and how these topological metrics can then be applied to populations of modeled atmospheres. I then confirmed that network metrics are indicative of the chemical and physical characteristics of an atmosphere, using hot jupiters as a proof of concept. Finally, I determined that the metrics can distinguish between populations of atmospheres interfacing with a biosphere or technosphere from those that aren't, and, in the case of the Archean Earth analogues, due so better than the abundance of atmospheric species alone.

However, there is still much room for improvement and exploration using this technique.

5.1 Future Directions

5.1.1 Connecting Atmospheric CRN Topology to Observational Data

The next great challenge for network metrics as tools to understand exoplanets is linking network topology to actual observations of exoplanet atmospheres. This is easier said than done–it is likely that a limited number of species will be detectable in exoplanet atmospheres (Tremblay *et al.*, 2019), and it must be determined if the species that *can* be detected is sufficient to build a network that usefully distinguish between populations of atmospheres. In essence, we need to know the minimal useful

size of a network, and there are at least two ways ascertain this quantity.

The first might be called the *top-down* approach. Here, I start with two fully modeled atmosphere, one with life, one without, build networks from them, and measure the networks' topological characteristics. I then begin removing the same species from each of the networks, measuring their metrics with every removal. The point at which the networks are no longer distinguishable using topological metrics, then, is the minimum useful network size required for our purposes.

The second might referred to as the *bottom-up* approach. Instead of starting with a fully modeled network, I begin with the species that *have* been observed, and construct a network or networks out of them, perhaps via network expansion, as has been done elsewhere in astrobiology (Smith *et al.*, 2021). I then compare the topological metrics of these networks to ones derived from fully modeled atmospheres. If the two sets of networks are too dissimilar, I add another species to the reduced network (ideally, the one that is the next most easily observed in the near-term future), and measure the network metrics again. When the metrics of the reduced network and the full network start to converge, I will know I have found the minimum useful network.

A third option might be to disregard building networks from observed species altogether, and instead directly infer the network structure from observed spectral features, such as the presence or absence of specific absorption lines. It remains to be seen if this is even feasible, but statistical analyses such as latent class analysis using finite mixture models–techniques developed specifically to detect subclasses and structures within populations that can't be observed directly within the data–may allow it (Nylund-Gibson *et al.*, 2019; Ferguson *et al.*, 2020). Should this prove to be a viable approach, then network topology analysis could potentially be integrated into existing analysis pipelines for observational data.

5.1.2 Potential Applications

If atmospheric CRN topology can be practically applied to observed exoplanets, than a range of exiting possible uses open up.

Some of the first applications, as alluded in Section 4.5, would be distinguishing between worlds with abiotically produced O_2 (Kleinböhl *et al.*, 2018), and those where O_2 is produced by biological processes like photosynthesis. Ideally, this would help mitigate the risk of false positives and negatives associated with species of oxygen in exoplanet atmospheres (Meadows, 2017; Harman and Domagal-Goldman, 2018), in much the same way I've done for CH_4 in this work. Another set of targets are planets with complex chemical atmospheres, along the lines of the deeply photochemisty-influenced atmosphere of Titan (Willacy *et al.*, 2016); an important test of atmospheric CRNs is whether they still can meaningfully distinguish between worlds with and without life in these circumstances.

It would also be interesting to see how network metrics change over time–either over geological time, through evolutionary transitions of the Earth's atmosphere such as that Great Oxidation Event (Schirrmeister *et al.*, 2013; Blaustein, 2016), or in on much shorter terms, such as annual seasonality. The former could help us understand how our planet's biosphere and atmosphere have co-evolved, and the latter could provide an additional biosignature in of itself (Olson *et al.*, 2018; Schwieterman, 2018).

Beyond that, however, we also must engage with the fact that biospheres are unlikely to be as simple as I've modeled thus far–as Section 1.3 reminds us, we must consider their biosignatures in the context of their ecology. An Archean Earth-like planet will have a biosphere with more than just methanogens–from our own planet's history, it is known that there was a wide variety of organisms and metabolisms in play, from heterotrophs to sulfate reducers to fermentative bacteria (Kharecha

et al., 2005; Canfield *et al.*, 2006). Incorporating these ecological dynamics into the network analysis pipeline could enhance their utility even further, and also help better constrain the probability space of biosignature detection and boosting confidence in any detection, as is already being investigated for biosignatures within our solar system (Affholder *et al.*, 2021).

More generally speaking, network metrics could easily be incorporated into a more Bayesian approach towards exoplanet biosignatures (Walker *et al.*, 2018, 2020), where instead of assuming a totally unambiguous detection of life beyond Earth, I conceptualize my findings in terms of likelihood and confidence, *e.g.*,

$$P(\text{life} \mid D, C) = \frac{P(D \mid C,\ \text{life}\ )P(\ \text{life}\ \mid C)}{P(D \mid C,\ \text{life}\ )P(\ \text{life}\ \mid C) + P(D \mid C,\ \text{no life}\ )P(\ \text{no life}\ \mid C)} \tag{5.1}$$

where D and C represent the data and context, respectively, of an exoplanet being evaluated for the potential presence of life (Catling *et al.*, 2018). Network metrics could provide a powerful prior to this framework for assessing $P(\text{life} \mid D, C)$–the probability of an accurate detection of extraterrestrial life.

If, as suggested in Section 1.3.1, network topology represents a "universal biosignatures", than even more exciting prospects open up towards the search for life wholly unlike our own. Network-based biosignatures could be applied to, for example, the nuclear chemical reaction networks found on the surface of stars in the search for putative nuclear-based "stellivore" organisms (Vidal, 2016, 2020), or lifeforms with even more bizarre metabolisms (Greffenstette *et al.*, 2023).

Lastly, closer to home, there exists in context of biochemical reaction networks the concept of the "control kernel"–a node whose behavior has an out-sized influence on the behavior of the entire network (Kim *et al.*, 2013; Walker *et al.*, 2016; Borriello and Daniels, 2021). If similar control kernels can be identified within atmospheres, then

this knowledge could help assess potential avenues towards climate change mitigation here on Earth (VijayaVenkataRaman *et al.*, 2012; Panepinto *et al.*, 2021).

5.2 Final Remarks

The detection of life beyond Earth would, without a doubt, be one of the greatest scientific discoveries in the history of humanity. It is due to this magnitude of importance that we must establish as much confidence in any potential detection, even if it requires letting go of our previous assumptions. Network-based biosignatures could play a powerful–perhaps even necessary–role in the re-evaluation of the hunt for extraterrestrial life.

The universe is, indeed, far queerer than we could suppose, and we will need to update our approaches accordingly.

REFERENCES

Affholder, A., F. Guyot, B. Sauterey, R. Ferrière and S. Mazevet, "Bayesian analysis of enceladus's plume data to assess methanogenesis", Nature Astronomy 5, 8, 805–814 (2021).

Blaustein, R., "The great oxidation event: evolving understandings of how oxygenic life on earth began", BioScience 66, 3, 189–195 (2016).

Borriello, E. and B. C. Daniels, "The basis of easy controllability in boolean networks", Nature Communications 12, 1, 5227 (2021).

Canfield, D. E., M. T. Rosing and C. Bjerrum, "Early anaerobic metabolisms", Philosophical Transactions of the Royal Society B: Biological Sciences 361, 1474, 1819–1836 (2006).

Catling, D. C., J. Krissansen-Totton, N. Y. Kiang, D. Crisp, T. D. Robinson, S. DasSarma, A. J. Rushby, A. Del Genio, W. Bains and S. Domagal-Goldman, "Exoplanet biosignatures: a framework for their assessment", Astrobiology 18, 6, 709–738 (2018).

Ferguson, S. L., E. W. G. Moore and D. M. Hull, "Finding latent groups in observed data: A primer on latent profile analysis in mplus for applied researchers", International Journal of Behavioral Development 44, 5, 458–468 (2020).

Greffenstette, N., L. Chou, S. Colon-Santos, T. M. Fisher, C. Nural, P. Sinhadc, V. Mierzejewski, M. Vidaurri, L. Vincent and M. M. Weng, "Astrobiology Primer 3.0, Ch. 9: Life as we don't know it", Astrobiology (in review) (2023).

Harman, C. E. and S. Domagal-Goldman, "Biosignature false positives", in "Handbook of Exoplanets", edited by H. J. Deeg and J. A. Belmonte, pp. 1–22 (Springer International Publishing, Cham, 2018).

Kharecha, P., J. Kasting and J. Siefert, "A coupled atmosphere–ecosystem model of the early archean earth", Geobiology 3, 2, 53–76 (2005).

Kim, J., S.-M. Park and K.-H. Cho, "Discovery of a kernel for controlling biomolecular regulatory networks", Scientific Reports 3, 1, 2223 (2013).

Kleinböhl, A., K. Willacy, A. J. Friedson, P. Chen and M. R. Swain, "Buildup of abiotic oxygen and ozone in moist atmospheres of temperate terrestrial exoplanets and its impact on the spectral fingerprint in transit observations", The Astrophysical Journal 862, 2, 92 (2018).

Meadows, V. S., "Reflections on O_2 as a biosignature in exoplanetary atmospheres", Astrobiology 17, 10, 1022–1052 (2017).

Nylund-Gibson, K., R. P. Grimm and K. E. Masyn, "Prediction from latent classes: A demonstration of different approaches to include distal outcomes in mixture models", Structural Equation Modeling: A Multidisciplinary Journal 26, 6, 967–985 (2019).

Olson, S. L., E. W. Schwieterman, C. T. Reinhard, A. Ridgwell, S. R. Kane, V. S. Meadows and T. W. Lyons, "Atmospheric seasonality as an exoplanet biosignature", The Astrophysical Journal Letters 858, 2, L14 (2018).

Panepinto, D., V. A. Riggio and M. Zanetti, "Analysis of the emergent climate change mitigation technologies", International Journal of Environmental Research and Public Health 18, 13, 6767 (2021).

Schirrmeister, B. E., J. M. de Vos, A. Antonelli and H. C. Bagheri, "Evolution of multicellularity coincided with increased diversification of cyanobacteria and the great oxidation event", Proceedings of the National Academy of Sciences 110, 5, 1791–1796 (2013).

Schwieterman, E. W., "Surface and temporal biosignatures", Handbook of Exoplanets p. 69 (2018).

Smith, H. B., A. Drew, J. F. Malloy and S. I. Walker, "Seeding biochemistry on other worlds: Enceladus as a case study", Astrobiology 21, 2, 177–190 (2021).

Tremblay, L., M. R. Line, K. B. Stevenson, T. Kataria, R. T. Zellem, J. J. Fortney and C. V. Morley, "The detectability and constraints of biosignature gases in the near & mid-infrared from transit transmission spectroscopy", arXiv:1912.10939 [astro-ph] (2019).

Vidal, C., "Stellivore extraterrestrials? Binary stars as living systems", Acta Astronautica 128, 251–256 (2016).

Vidal, C., "Energy rate density as a technosignature: The case for stellivores", in "Life in the Universe: Big History, SETI and the Future of Humankind. IBHA & INAF-IASF MI Symposium 15-16 July 2019", (2020).

VijayaVenkataRaman, S., S. Iniyan and R. Goic, "A review of climate change, mitigation and adaptation", Renewable and Sustainable Energy Reviews 16, 1, 878–897 (2012).

Walker, S. I., W. Bains, L. Cronin, S. DasSarma, S. Danielache, S. Domagal-Goldman, B. Kacar, N. Y. Kiang, A. Lenardic, C. T. Reinhard, W. Moore, E. W. Schwieterman, E. L. Shkolnik and H. B. Smith, "Exoplanet biosignatures: future directions", Astrobiology 18, 6, 779–824 (2018).

Walker, S. I., L. Cronin, A. Drew, S. Domagal-Goldman, T. Fisher, M. Line and C. Millsaps, "Probabilistic biosignature frameworks", in "Planetary Astrobiology", p. 477 (University of Arizona Press, 2020).

Walker, S. I., H. Kim and P. C. Davies, "The informational architecture of the cell", Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374, 2063, 20150057 (2016).

Willacy, K., M. Allen and Y. Yung, "A new astrobiological model of the atmosphere of titan", The Astrophysical Journal 829, 2, 79 (2016).

Adam, Z. R., A. C. Fahrenbach, S. M. Jacobson, B. Kacar and D. Y. Zubarev, "Radiolysis generates a complex organosynthetic chemical network", Scientific Reports 11, 1, 1–10 (2021).

Affholder, A., A. Bixel, B. Sauterey, R. Ferrière, D. Apai and S. Mazevet, "Predicting quantitative biosignature patterns from populations of exoplanets", in "AGU Fall Meeting Abstracts", vol. 2021, pp. P55D–1969 (2021a).

Affholder, A., F. Guyot, B. Sauterey, R. Ferrière and S. Mazevet, "Bayesian analysis of enceladus's plume data to assess methanogenesis", Nature Astronomy 5, 8, 805–814 (2021b).

Ahrer, E.-M., L. Alderson, N. M. Batalha, N. E. Batalha, J. L. Bean, T. G. Beatty, T. J. Bell, B. Benneke, Z. K. Berta-Thompson, A. L. Carter, I. J. M. Crossfield, N. Espinoza, A. D. Feinstein, J. J. Fortney, N. P. Gibson, J. M. Goyal, E. M.-R. Kempton, J. Kirk, L. Kreidberg, M. López-Morales, M. R. Line, J. D. Lothringer, S. E. Moran, S. Mukherjee, K. Ohno, V. Parmentier, C. Piaulet, Z. Rustamkulov, E. Schlawin, D. K. Sing, K. B. Stevenson, H. R. Wakeford, N. H. Allen, S. M. Birkmann, J. Brande, N. Crouzet, P. E. Cubillos, M. Damiano, J.-M. Désert, P. Gao, J. Harrington, R. Hu, S. Kendrew, H. A. Knutson, P.-O. Lagage, J. Leconte, M. Lendl, R. J. MacDonald, E. M. May, Y. Miguel, K. Molaverdikhani, J. I. Moses, C. A. Murray, M. Nehring, N. K. Nikolov, D. J. M. Petit dit de la Roche, M. Radica, P.-A. Roy, K. G. Stassun, J. Taylor, W. C. Waalkes, P. Wachiraphan, L. Welbanks, P. J. Wheatley, K. Aggarwal, M. K. Alam, A. Banerjee, J. K. Barstow, J. Blecic, S. L. Casewell, Q. Changeat, K. L. Chubb, K. D. Colón, L.-P. Coulombe, T. Daylan, M. de Val-Borro, L. Decin, L. A. Dos Santos, L. Flagg, K. France, G. Fu, A. García Muñoz, J. E. Gizis, A. Glidden, D. Grant, K. Heng, T. Henning, Y.-C. Hong, J. Inglis, N. Iro, T. Kataria, T. D. Komacek, J. E. Krick, E. K. H. Lee, N. K. Lewis, J. Lillo-Box, J. Lustig-Yaeger, L. Mancini, A. M. Mandell, M. Mansfield, M. S. Marley, T. Mikal-Evans, G. Morello, M. C. Nixon, K. Ortiz Ceballos, A. A. A. Piette, D. Powell, B. V. Rackham, L. Ramos-Rosado, E. Rauscher, S. Redfield, L. K. Rogers, M. T. Roman, G. M. Roudier, N. Scarsdale, E. L. Shkolnik, J. Southworth, J. J. Spake, M. E. Steinrueck, X. Tan, J. K. Teske, P. Tremblin, S.-M. Tsai, G. S. Tucker, J. D. Turner, J. A. Valenti, O. Venot, I. P. Waldmann, N. L. Wallack, X. Zhang, S. Zieba and JWST Transiting Exoplanet Community Early Release Science Team, "Identification of carbon dioxide in an exoplanet atmosphere", Nature 614, 7949 (2023).

Akins, A. B., A. P. Lincowski, V. S. Meadows and P. G. Steffes, "Complications in the alma detection of phosphine at venus", The Astrophysical Journal Letters 907, 2, L27 (2021).

Albert, R. and A.-L. Barabási, "Statistical mechanics of complex networks", Rev. Mod. Phys. 74, 1, 47–97 (2002).

Alderson, L., H. R. Wakeford, M. K. Alam, N. E. Batalha, J. D. Lothringer, J. Adams Redai, S. Barat, J. Brande, M. Damiano, T. Daylan *et al.*, "Early release science of the exoplanet wasp-39b with jwst nirspec g395h", Nature 614, 7949, 664–669 (2023).

Amaral, L. A. and J. M. Ottino, "Complex networks: Augmenting the framework for the study of complex systems", The European Physical Journal B 38, 147–162 (2004).

Anand, M., A. Gonzalez, F. Guichard, J. Kolasa and L. Parrott, "Ecological systems as complex systems: challenges for an emerging science", Diversity 2, 3, 395–410 (2010).

Angeli, D., "A tutorial on chemical reaction networks dynamics", in "2009 European Control Conference (ECC)", pp. 649–657 (IEEE, 2009).

Antoniou, I. E. and E. T. Tsompa, "Statistical analysis of weighted networks", Discrete Dynamics in Nature and Society 2008, e375452 (2008).

Arney, G., S. D. Domagal-Goldman, V. S. Meadows, E. T. Wolf, E. Schwieterman, B. Charnay, M. Claire, E. Hébrard and M. G. Trainer, "The pale orange dot: the spectrum and habitability of hazy Archean Earth", Astrobiology 16, 11, 873–899 (2016).

Arney, G. N., "The k dwarf advantage for biosignatures on directly imaged exoplanets", The Astrophysical Journal Letters 873, 1, L7 (2019).

Bains, W., J. J. Petkowski, S. Seager, S. Ranjan, C. Sousa-Silva, P. B. Rimmer, Z. Zhan, J. S. Greaves and A. M. Richards, "Phosphine on venus cannot be explained by conventional processes", Astrobiology 21, 10, 1277–1304 (2021).

Bains, W., S. Seager and A. Zsom, "Photosynthesis in hydrogen-dominated atmospheres", Life 4, 4, 716–744 (2014).

Balbi, A. and M. M. Ćirković, "Longevity is the key factor in the search for technosignatures", The Astronomical Journal 161, 5, 222 (2021).

Barabasi, A.-L. and Z. N. Oltvai, "Network biology: understanding the cell's functional organization", Nature Reviews Genetics 5, 2, 101–113 (2004).

Barabási, A.-L. and M. Pósfai, *Network Science* (Cambridge University Press, Cambridge, United Kingdom, 2016).

Barrat, A., M. Barthélemy, R. Pastor-Satorras and A. Vespignani, "The architecture of complex weighted networks", Proceedings of the National Academy of Sciences 101, 11, 3747–3752 (2004).

Bascompte, J., "Networks in ecology", Basic and Applied Ecology 8, 6, 485–490 (2007).

Bella, A., A. Choprab, W. Fawcettc and R. Talebid, "NASA Frontier Development Lab technical memorandum", (2018).

Benneke, B. and S. Seager, "Atmospheric retrieval for super-Earths: uniquely constraining the atmospheric composition with transmission spectroscopy", ApJ 753, 2, 100 (2012).

Blaustein, R., "The great oxidation event: evolving understandings of how oxygenic life on earth began", BioScience 66, 3, 189–195 (2016).

Borriello, E. and B. C. Daniels, "The basis of easy controllability in boolean networks", Nature Communications 12, 1, 5227 (2021).

Broadley, M. W., P. H. Barry, C. J. Ballentine, L. A. Taylor and R. Burgess, "End-permian extinction amplified by plume-induced release of recycled lithospheric volatiles", Nature Geoscience 11, 9, 682–687 (2018).

Broido, A. D. and A. Clauset, "Scale-free networks are rare", Nature Communications 10, 1, 1017 (2019).

Burger, J. R., C. Hou, C. Hall and J. R. Brown, "Universal rules of life: metabolic rates, biological times and the equal fitness paradigm", bioRxiv (2020).

Butcher, J. C., *Numerical methods for ordinary differential equations* (John Wiley & Sons, 2016).

Cabrol, N. A., "The coevolution of life and environment on Mars: An ecosystem perspective on the robotic exploration of biosignatures", Astrobiology 18, 1, 1–27 (2018).

Canfield, D. E., M. T. Rosing and C. Bjerrum, "Early anaerobic metabolisms", Philosophical Transactions of the Royal Society B: Biological Sciences 361, 1474, 1819–1836 (2006).

Catling, D. C., J. Krissansen-Totton, N. Y. Kiang, D. Crisp, T. D. Robinson, S. DasSarma, A. J. Rushby, A. Del Genio, W. Bains and S. Domagal-Goldman, "Exoplanet biosignatures: a framework for their assessment", Astrobiology 18, 6, 709–738 (2018).

Catling, D. C. and K. J. Zahnle, "The archean atmosphere", Science Advances 6, 9, eaax1420 (2020).

Ceja, A. Y. and S. Kane, "The search for extraterrestrial life: An astro-ecological modeling approach for characterizing exoplanet habitability", in "American Astronomical Society Meeting Abstracts# 233", vol. 233, pp. 346–01 (2019).

Centler, F. and P. Dittrich, "Chemical organizations in atmospheric photochemistries—A new method to analyze chemical reaction networks", Planetary and Space Science 55, 4, 413–428 (2007).

Chan, M. A., N. W. Hinman, S. L. Potter-McIntyre, K. E. Schubert, R. J. Gillams, S. M. Awramik, P. J. Boston, D. M. Bower, D. J. Des Marais, J. D. Farmer, T. Z. Jia, P. L. King, R. M. Hazen, R. J. Léveillé, D. Papineau, K. R. Rempfert, M. Sánchez-Román, J. R. Spear, G. Southam, J. C. Stern and H. J. Cleaves, "Deciphering biosignatures in planetary contexts", Astrobiology 19, 9, 1075–1102 (2019).

Charbonneau, D., T. M. Brown, R. W. Noyes and R. L. Gilliland, "Detection of an extrasolar planet atmosphere", The Astrophysical Journal 568, 1, 377 (2002).

Chela-Flores, J., "From systems chemistry to systems astrobiology: life in the universe as an emergent phenomenon", International Journal of Astrobiology 12, 1, 8–16 (2013).

Chen, T. and C. Guestrin, *XGBoost* (2016), publication Title: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Clarke, B. L., "Graph theoretic approach to the stability analysis of steady state chemical reaction networks", The Journal of Chemical Physics 60, 4, 1481–1492 (1974).

Cobb, A. D., M. D. Himes, F. Soboczenski, S. Zorzan, M. D. O'Beirne, A. G. Baydin, Y. Gal, S. D. Domagal-Goldman, G. N. Arney, D. Angerhausen and 2018 NASA FDL Astrobiology Team II, "An ensemble of Bayesian neural networks for exoplanetary atmospheric retrieval", AJS 158, 1, 33 (2019).

Conradi, C., J. Saez-Rodriguez, E.-D. Gilles and J. Raisch, "Chemical reaction network theory... a tool for systems biology", Proceedings of the 5th MATHMOD (2006).

Coustenis, A., J. Schneider, R. Wittemberg, E. Chazsefiere, T. Guillot, A. Penny, T. Greene, H. Rauer and D. Bockelee-Morvan, "High resolution ground-based spectroscopy of 51 Peg b: Search for atmospheric signatures", in "Brown Dwarfs and Extrasolar Planets", vol. 134, p. 296 (1998).

Cronin, L., "Exploring the transition from prebiotic chemistry to biology using molecular assembly theory", in "AGU Fall Meeting Abstracts", vol. 2020, pp. P001–01 (2020).

Crossfield, I. J. M., "Observations of exoplanet atmospheres", PASP 127, 956, 941 (2015).

Daniels, B. C., H. Kim, D. Moore, S. Zhou, H. Smith, B. Karas, S. A. Kauffman and S. I. Walker, "Logic and connectivity jointly determine criticality in biological gene regulatory networks", Physical Review Letters 121, 13 (2018).

Davis, R., "Searching for extraterrestrial intelligence by locating potential ET communication networks in space", arXiv:1907.05259 [physics] (2019).

Domagal-Goldman, S. D., A. Segura, M. W. Claire, T. D. Robinson and V. S. Meadows, "Abiotic ozone and oxygen in atmospheres similar to prebiotic Earth", The Astrophysical Journal 792, 2, 90 (2014).

Elser, J. J., "Biological stoichiometry: a theoretical framework connecting ecosystem ecology, evolution, and biochemistry for application in astrobiology", International Journal of Astrobiology 2, 3, 185–193 (2003).

Estrada, E., "Returnability as a criterion of disequilibrium in atmospheric reactions networks", Journal of Mathematical Chemistry 50, 6, 1363–1372 (2012).

Etiope, G. and B. Sherwood Lollar, "Abiotic methane on earth", Reviews of Geophysics 51, 2, 276–299 (2013).

FDL, N., "NASA Frontier Development Labs (FDL) Project", URL https://exoplanetarchive.ipac.caltech.edu/cgi-bin/FDL/nph-fdl?atmos (2019).

Ferguson, S. L., E. W. G. Moore and D. M. Hull, "Finding latent groups in observed data: A primer on latent profile analysis in mplus for applied researchers", International Journal of Behavioral Development 44, 5, 458–468 (2020).

Fisher, T., H. Kim, C. Millsaps, M. Line and S. I. Walker, "Inferring exoplanet disequilibria with multivariate information in atmospheric reaction networks", The Astronomical Journal 164, 2, 53 (2022).

Fisher, T. M. and D. Schulze-Makuch, "Nutrient and population dynamics in a subglacial reservoir: a simulation case study of the Blood Falls ecosystem with implications for astrobiology", International Journal of Astrobiology 12, 4, 304–311 (2013).

Ford, F. A., *Modeling the Environment: An Introduction to System Dynamics Models of Environmental Systems* (Island press, 1999).

Fortney, J. J., R. I. Dawson and T. D. Komacek, "Hot jupiters: origins, structure, atmospheres", J. Geophys. Res. Planets (2021).

Fujii, Y., D. Angerhausen, R. Deitrick, S. Domagal-Goldman, J. L. Grenfell, Y. Hori, S. R. Kane, E. Pallé, H. Rauer, N. Siegler, K. Stapelfeldt and K. B. Stevenson, "Exoplanet biosignatures: observational prospects", Astrobiology 18, 6, 739–778 (2018).

Gebauer, S., I. Vilović, J. L. Grenfell, F. Wunderlich, F. Schreier and H. Rauer, "Influence of biomass emissions upon habitability, biosignatures and detectability in Earth-like atmospheres", arXiv p. arXiv:2102.00220 (2021).

Gleiss, P. M., P. F. Stadler, A. Wagner and D. A. Fell, "Relevant cycles in chemical reaction networks", Advs. Complex Syst. 04, 02n03, 207–226 (2001).

Grasby, S. E. and K. L. Londry, "Biogeochemistry of hypersaline springs supporting a mid-continent marine ecosystem: an analogue for martian springs?", Astrobiology 7, 4, 662–683 (2007).

Greaves, J. S., A. Richards, W. Bains, P. B. Rimmer, H. Sagawa, D. L. Clements, S. Seager, J. J. Petkowski, C. Sousa-Silva, S. Ranjan *et al.*, "Phosphine gas in the cloud decks of venus", Nature Astronomy 5, 7, 655–664 (2021).

Green, J., T. Hoehler, M. Neveu, S. Domagal-Goldman, D. Scalice and M. Voytek, "Call for a framework for reporting evidence for life beyond earth", Nature 598, 7882, 575–579 (2021).

Greffenstette, N., L. Chou, S. Colon-Santos, T. M. Fisher, C. Nural, P. Sinhadc, V. Mierzejewski, M. Vidaurri, L. Vincent and M. M. Weng, "Astrobiology Primer 3.0, Ch. 9: Life as we don't know it", Astrobiology (in review) (2023).

Grenfell, J. L., "A review of exoplanetary biosignatures", Physics Reports 713, 1–17 (2017).

Guzmán-Marmolejo, A., A. Segura and E. Escobar-Briones, "Abiotic production of methane in terrestrial planets", Astrobiology 13, 6, 550–559 (2013).

Haas, J. R., "The potential feasibility of chlorinic photosynthesis on exoplanets", Astrobiology 10, 9, 953–963 (2010).

Hagberg, A., P. Swart and D. S Chult, "Exploring network structure, dynamics, and function using NetworkX", Tech. rep., Los Alamos National Lab.(LANL), Los Alamos, NM (United States) (2008).

Haldane, J. B. S., *Possible Worlds and Other Essays* (Chatto and Windus, London, 1927), 1st edn.

Haqq-Misra, J., T. J. Fauchez, E. W. Schwieterman and R. Kopparapu, "Disruption of a planetary nitrogen cycle as evidence of extraterrestrial agriculture", The Astrophysical Journal Letters 929, 2, L28 (2022a).

Haqq-Misra, J., R. Kopparapu, T. J. Fauchez, A. Frank, J. T. Wright and M. Lingam1, "Detectability of chlorofluorocarbons in the atmospheres of habitable M-dwarf planets", arXiv:2202.05858 ArXiv: 2202.05858 (2022b).

Haqq-Misra, J., E. W. Schwieterman, H. Socas-Navarro, R. Kopparapu, D. Angerhausen, T. G. Beatty, S. Berdyugina, R. Felton, S. Sharma, G. G. De la Torre and D. Apai, "Searching for technosignatures in exoplanetary systems with current and future missions", Acta Astronautica 198, 194–207 (2022c).

Harman, C. E. and S. Domagal-Goldman, "Biosignature false positives", in "Handbook of Exoplanets", edited by H. J. Deeg and J. A. Belmonte, pp. 1–22 (Springer International Publishing, Cham, 2018).

Hayes, J. J. C., E. Kerins, S. Awiphan, I. McDonald, J. S. Morgan, P. Chuanraksasat, S. Komonjinda, N. Sanguansak, P. Kittara and (SPEARNET), "Optimizing exoplanet atmosphere retrieval using unsupervised machine-learning classification", Mon. Not. R. Astron. Soc. 494, 3, 4492–4508 (2020).

Heng, K., J. M. Mendonça and J.-M. Lee, "Analytical models of exoplanetary atmospheres. ii. radiative transfer via the two-stream approximation", The Astrophysical Journal Supplement Series 215, 1, 4 (2014).

Jeong, H., B. Tombor, R. Albert, Z. N. Oltvai and A.-L. Barabási, "The large-scale organization of metabolic networks", Nature 407, 6804, 651–654 (2000).

Jolley, C. and T. Douglas, "Topological biosignatures: large-scale structure of chemical networks from biology and astrochemistry", Astrobiology 12, 1, 29–39 (2011).

Jolley, C. C. and T. Douglas, "A network-theoretical approach to understanding interstellar chemistry", The Astrophysical Journal 722, 2, 1921–1931 (2010).

Jørgensen, S. E., B. C. Patten and M. Straškraba, "Ecosystems emerging: toward an ecology of complex systems in a complex future", Ecological Modelling 62, 1-3, 1–27 (1992).

Kaltenegger, L., Z. Lin and S. Rugheimer, "Finding signs of life on transiting Earthlike planets: high-resolution transmission spectra of Earth through time around FGKM host stars", The Astrophysical Journal 904, 1, 10 (2020).

Kasting, J. F., S. Liu and T. Donahue, "Oxygen levels in the prebiological atmosphere", Journal of Geophysical Research: Oceans 84, C6, 3097–3107 (1979).

Kasting, J. F. and J. L. Siefert, "Life and the evolution of earth's atmosphere", Science 296, 5570, 1066–1068 (2002).

Kharecha, P., J. Kasting and J. Siefert, "A coupled atmosphere–ecosystem model of the early archean earth", Geobiology 3, 2, 53–76 (2005).

Kiang, N. Y., S. Domagal-Goldman, M. N. Parenteau, D. C. Catling, Y. Fujii, V. S. Meadows, E. W. Schwieterman and S. I. Walker, "Exoplanet biosignatures: at the dawn of a new era of planetary observations", Astrobiology 18, 6, 619–629 (2018).

Kiang, N. Y., A. Segura, G. Tinetti, R. E. Blankenship, M. Cohen, J. Siefert, D. Crisp, V. S. Meadows and others, "Spectral signatures of photosynthesis: Coevolution with other stars and the atmosphere on extrasolar worlds", Astrobiology 7, 3, 486–7 (2007).

Kim, H., H. B. Smith, C. Mathis, J. Raymond and S. I. Walker, "Universal scaling across biochemical networks on Earth", Science Advances 5, 1, eaau0149 (2019).

Kim, J., S.-M. Park and K.-H. Cho, "Discovery of a kernel for controlling biomolecular regulatory networks", Scientific Reports 3, 1, 2223 (2013).

Kleinböhl, A., K. Willacy, A. J. Friedson, P. Chen and M. R. Swain, "Buildup of abiotic oxygen and ozone in moist atmospheres of temperate terrestrial exoplanets and its impact on the spectral fingerprint in transit observations", The Astrophysical Journal 862, 2, 92 (2018).

Klobas, J. E. and D. Wilmouth, "Volcanogenic chlorofluorocarbons and the recent cfc anomalies", (2019).

Kopparapu, R., G. Arney, J. Haqq-Misra, J. Lustig-Yaeger and G. Villanueva, "Nitrogen dioxide pollution as a signature of extraterrestrial technology", The Astrophysical Journal 908, 2, 164 (2021).

Krissansen-Totton, J., G. N. Arney, D. C. Catling, R. Felton, J. Fortney, R. Garland, P. Irwin, R. Kopparapu, O. Lehmer, J. Lustig-Yaeger and others, "Atmospheric disequilibrium as an exoplanet biosignature: Opportunities for next generation telescopes", BAAS 51, 3, 158 (2019).

Krissansen-Totton, J., D. S. Bergsman and D. C. Catling, "On detecting biospheres from chemical thermodynamic disequilibrium in planetary atmospheres", Astrobiology 16, 1, 39–67 (2016).

Krissansen-Totton, J., S. Olson and D. C. Catling, "Disequilibrium biosignatures over Earth history and implications for detecting exoplanet life", Science Advances 4, 1, eaao5747 (2018).

Kuikka, J. T., "Scaling laws in physiology: relationships between size, function, metabolism and life expectancy", International Journal of Nonlinear Sciences and Numerical Simulation 4, 4, 317–328 (2003).

Lente, G., *Deterministic Kinetics in Chemistry and Systems Biology: the Dynamics of Complex Reaction Networks* (Springer, 2015).

Lin, H. W., G. G. Abad and A. Loeb, "Detecting industrial pollution in the atmospheres of Earth-like exoplanets", The Astrophysical Journal 792, 1, L7 (2014).

Lincowski, A. P., J. Lustig-Yaeger and V. S. Meadows, "Observing isotopologue bands in terrestrial exoplanet atmospheres with the James Webb Space Telescope: implications for identifying past atmospheric and ocean loss", The Astronomical Journal 158, 1, 26 (2019).

Line, M. R., J. Teske, B. Burningham, J. J. Fortney and M. S. Marley, "Uniform atmospheric retrieval analysis of ultracool dwarfs. I. Characterizing benchmarks, Gl 570D and HD 3651B", The Astrophysical Journal 807, 2, 183 (2015).

Line, M. R., A. S. Wolf, X. Zhang, H. Knutson, J. A. Kammer, E. Ellison, Pieter Deroo, D. Crisp and Y. L. Yung, "A systematic retrieval analysis of secondary eclipse spectra. I. A comparison of atmospheric retrieval techniques", The Astrophysical Journal 775, 2, 137 (2013).

Lovelock, J. E. and L. Margulis, "Atmospheric homeostasis by and for the biosphere: the Gaia hypothesis", Tellus 26, 1-2, 2–10 (1974).

Lustig-Yaeger, J., V. S. Meadows and A. P. Lincowski, "The detectability and characterization of the TRAPPIST-1 exoplanet atmospheres with JWST", The Astronomical Journal 158, 1, 27 (2019).

Madhusudhan, N., "Exoplanetary atmospheres: key insights, challenges and prospects", Annual Review of Astronomy and Astrophysics 57, 617–663 (2019).

Madhusudhan, N., J. Harrington, K. B. Stevenson, S. Nymeyer, C. J. Campo, P. J. Wheatley, D. Deming, J. Blecic, R. A. Hardy, N. B. Lust, D. R. Anderson, A. Collier-Cameron, C. B. T. Britt, W. C. Bowman, L. Hebb, C. Hellier, P. F. L. Maxted, D. Pollacco and R. G. West, "A high C/O ratio and weak thermal inversion in the atmosphere of exoplanet WASP-12b", Nature 469, 7328, 64–67 (2011).

Mariscal, C. and L. Fleming, "Why we should care about universal biology", Biological Theory 13, 2, 121–130 (2018).

Marquet, P. A., R. A. Quiñones, S. Abades, F. Labra, M. Tognelli, M. Arim and M. Rivadeneira, "Scaling and power-laws in ecological systems", Journal of Experimental Biology 208, 9, 1749–1769 (2005).

Marshall, S. M., C. Mathis, E. Carrick, G. Keenan, G. J. Cooper, H. Graham, M. Craven, P. S. Gromski, D. G. Moore, S. I. Walker *et al.*, "Identifying molecules as biosignatures with assembly theory and mass spectrometry", Nature Communications 12, 1, 3033 (2021).

Massey, F. J., "The Kolmogorov-Smirnov test for goodness of fit", Journal of the American Statistical Association 46, 68–78 (1951).

May, R. M., "Network structure and the biology of populations", Trends in Ecology & Evolution 21, 7, 394–399 (2006).

Mayor, M. and D. Queloz, "A Jupiter-mass companion to a solar-type star", Nature 378, 6555, 355–359 (1995).

Meadows, V. S., "Reflections on O_2 as a biosignature in exoplanetary atmospheres", Astrobiology 17, 10, 1022–1052 (2017).

Meadows, V. S., C. T. Reinhard, G. N. Arney, M. N. Parenteau, E. W. Schwieterman, S. D. Domagal-Goldman, A. P. Lincowski, K. R. Stapelfeldt, H. Rauer, S. DasSarma, S. Hegde, N. Narita, R. Deitrick, J. Lustig-Yaeger, T. W. Lyons, N. Siegler and J. L. Grenfell, "Exoplanet biosignatures: understanding oxygen as a biosignature in the context of its Environment", Astrobiology 18, 6, 630–662 (2018).

Mitchell, M., "Complex systems: Network thinking", Artificial Intelligence 170, 18, 1194–1212 (2006).

Molaverdikhani, K., T. Henning and P. Mollière, "From cold to hot irradiated gaseous exoplanets: fingerprints of chemical disequilibrium in atmospheric spectra", The Astrophysical Journal 883 (2019).

Mollière, P., R. v. Boekel, C. Dullemond, T. Henning and C. Mordasini, "Model atmospheres of irradiated exoplanets: the influence of stellar parameters, metallicity, and the C/O ratio", The Astrophysical Journal 813, 1, 47 (2015).

Montañez, R., M. A. Medina, R. V. Solé and C. Rodríguez-Caso, "When metabolism meets topology: reconciling metabolite and reaction networks", Bioessays 32, 3, 246–256 (2010).

Moses, J. I., C. Visscher, J. J. Fortney, A. P. Showman, N. K. Lewis, C. A. Griffith, S. J. Klippenstein, M. Shabram, A. J. Friedson, M. S. Marley and R. S. Freedman, "Disequilibrium carbon, oxygen, and nitrogen chemistry in the atmospheres of HD 189733b and HD 209458b", The Astrophysical Journal 737, 1, 15 (2011).

Moutou, C., A. Coustenis, J. Schneider, R. St Gilles, M. Mayor, D. Queloz and A. Kaufer, "Search for spectroscopical signatures of transiting HD 209458b's exosphere", Astronomy & Astrophysics 371, 1, 260–266 (2001).

National Academy of Sciences, "Decadal survey on astronomy and astrophysics 2020", (2021).

Newman, M., *Networks: An Introduction* (OUP Oxford, 2010).

Newman, M., A.-L. Barabási and D. J. Watts, *The Structure and Dynamics of Networks* (Princeton University Press, 2011).

Newman, M. E. J., "The structure and function of complex networks", SIAM Rev. 45, 2, 167–256 (2003).

Nicholson, A., S. Daines, N. Mayne, J. Eager-Nash, T. Lenton and K. Kohary, "Predicting biosignatures for nutrient-limited biospheres", Monthly Notices of the Royal Astronomical Society 517, 1, 222–239 (2022).

Nylund-Gibson, K., R. P. Grimm and K. E. Masyn, "Prediction from latent classes: A demonstration of different approaches to include distal outcomes in mixture models", Structural Equation Modeling: A Multidisciplinary Journal 26, 6, 967–985 (2019).

Olson, S. L., E. W. Schwieterman, C. T. Reinhard, A. Ridgwell, S. R. Kane, V. S. Meadows and T. W. Lyons, "Atmospheric seasonality as an exoplanet biosignature", The Astrophysical Journal Letters 858, 2, L14 (2018).

Onnela, J.-P., J. Saramäki, J. Kertész and K. Kaski, "Intensity and coherence of motifs in weighted complex networks", Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 71, 6 Pt 2, 065103 (2005).

O'Malley-James, J. T. and L. Kaltenegger, "Biofluorescent worlds: global biological fluorescence as a biosignature", Monthly Notices of the Royal Astronomical Society 481, 2, 2487–2496 (2018).

O'Malley-James, J. T. and L. Kaltenegger, "Expanding the timeline for Earth's photosynthetic red edge biosignature", The Astrophysical Journal Letters 879, 2, L20 (2019).

Panepinto, D., V. A. Riggio and M. Zanetti, "Analysis of the emergent climate change mitigation technologies", International Journal of Environmental Research and Public Health 18, 13, 6767 (2021).

Proulx, S. R., D. E. Promislow and P. C. Phillips, "Network thinking in ecology and evolution", Trends in Ecology & Evolution 20, 6, 345–353 (2005).

Rein, H., Y. Fujii and D. S. Spiegel, "Some inconvenient truths about biosignatures involving two chemical species on Earth-like exoplanets", Proceedings of the National Academy of Sciences 111, 19, 6871–6875 (2014).

Reinhard, C. T., S. L. Olson, E. W. Schwieterman and T. W. Lyons, "False negatives for remote life detection on ocean-bearing planets: lessons from the early Earth", Astrobiology 17, 4, 287–297 (2017).

Sagan, C., W. R. Thompson, R. Carlson, D. Gurnett and C. Hord, "A search for life on Earth from the Galileo spacecraft", Nature 365, 6448, 715–721 (1993).

Sandora, M. and J. Silk, "Biosignature surveys to exoplanet yields and beyond", Monthly Notices of the Royal Astronomical Society 495, 1, 1000–1015 (2020).

Sanromá, E., E. Pallé, M. Parenteau, N. Kiang, A. Gutiérrez-Navarro, R. López and P. Montañés-Rodríguez, "Characterizing the purple Earth: modeling the globally integrated spectral variability of the Archean Earth", The Astrophysical Journal 780, 1, 52 (2013).

Saxena, P., "Photobombing earth 2.0: diffraction-limit-related contamination and uncertainty in habitable planet spectra", The Astrophysical Journal Letters 934, 2, L32 (2022).

Schirrmeister, B. E., J. M. de Vos, A. Antonelli and H. C. Bagheri, "Evolution of multicellularity coincided with increased diversification of cyanobacteria and the great oxidation event", Proceedings of the National Academy of Sciences 110, 5, 1791–1796 (2013).

Schneider, J., A. Léger, M. Fridlund, G. J. White, C. Eiroa, T. Henning, T. Herbst, H. Lammer, R. Liseau, F. Paresce and others, "The far future of exoplanet direct characterization", Astrobiology 10, 1, 121–126 (2010).

Schwieterman, E. W., "Surface and temporal biosignatures", Handbook of Exoplanets p. 69 (2018).

Schwieterman, E. W., N. Y. Kiang, M. N. Parenteau, C. E. Harman, S. DasSarma, T. M. Fisher, G. N. Arney, H. E. Hartnett, C. T. Reinhard, S. L. Olson, V. S. Meadows, C. S. Cockell, S. I. Walker, J. L. Grenfell, S. Hegde, S. Rugheimer, R. Hu and T. W. Lyons, "Exoplanet biosignatures: a review of remotely detectable signs of life", Astrobiology 18, 6, 663–708 (2018).

Seager, S., "Exoplanet habitability", Science 340, 6132, 577–581 (2013).

Seager, S., "The search for habitable planets with biosignature gases framed by a 'biosignature Drake Equation"', International Journal of Astrobiology 17, 4, 294–302 (2018).

Seager, S. and W. Bains, "The search for signs of life on exoplanets at the interface of chemistry and planetary science", Science Advances 1, 2, e1500047 (2015).

Seager, S., W. Bains and R. Hu, "A biomass-based model to estimate the plausibility of exoplanet biosignature gases", The Astrophysical Journal 775, 2, 104 (2013a).

Seager, S., W. Bains and R. Hu, "Biosignature gases in H2-dominated atmospheres on rocky exoplanets", The Astrophysical Journal 777, 2, 95 (2013b).

Seager, S., W. Bains and J. Petkowski, "Toward a list of molecules as potential biosignature gases for the search for life on exoplanets and applications to terrestrial Biochemistry", Astrobiology 16, 6, 465–485 (2016).

Seager, S. and D. Deming, "Exoplanet atmospheres", Annual Review of Astronomy and Astrophysics 48, 631–672 (2010).

Seager, S., E. L. Turner, J. Schafer and E. B. Ford, "Vegetation's red edge: a possible spectroscopic biosignature of extraterrestrial plants", Astrobiology 5, 3, 372–390 (2005).

Seeburger, R., P. M. Higgins, N. P. Whiteford and C. S. Cockell, "Linking methanogenesis in low-temperature hydrothermal vent systems to planetary spectra: Methane biosignatures on an archean-earth-like exoplanet", Astrobiology 23, 4, 415–430 (2023).

Shalizi, C. R., "Methods and techniques of complex systems science: An overview", Complex Systems Science in Biomedicine pp. 33–114 (2006).

Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, "Cytoscape: a software environment for integrated models of biomolecular interaction networks", Genome Research 13, 11, 2498–2504 (2003).

Sharma, N., *XGBoost. The Extreme Gradient Boosting for Mining Applications* (GRIN Verlag, 2018).

Sheikh, S., A. Berea, R. Davis, G. G. De la Torre, J. DeMarines, T. Fisher, S. Foote, D. Gelino, D. Grinspoon, G. Profitiliotis *et al.*, "Technosignatures as a priority in planetary science", Bulletin of the American Astronomical Society 53, 4, 427 (2021).

Siegenfeld, A. F. and Y. Bar-Yam, "An introduction to complex systems science and its applications", Complexity 2020, 1–16 (2020).

Simoncini, E., N. Virgo and A. Kleidon, "Quantifying drivers of chemical disequilibrium: theory and application to methane in the Earth's atmosphere", Earth System Dynamics 4, 317–331 (2013).

Smith, H. B., A. Drew, J. F. Malloy and S. I. Walker, "Seeding biochemistry on other worlds: Enceladus as a case study", Astrobiology 21, 2, 177–190 (2021a).

Smith, H. B., H. Kim and S. I. Walker, "Scarcity of scale-free topology is universal across biochemical networks", Sci. Rep. 11, 1, 6542 (2021b).

Smith, H. B. and C. Mathis, "The futility of exoplanet biosignatures", arXiv:2205.07921 (2022).

Solé, R. V., S. A. Levin, J. H. Brown, V. K. Gupta, B.-L. Li, B. T. Milne, C. Restrepo and G. B. West, "The fractal nature of nature: power laws, ecological complexity and biodiversity", Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 357, 1421, 619–626 (2002).

Solé, R. V. and A. Munteanu, "The large-scale organization of chemical reaction networks in astrophysics", Europhysics Letters (EPL) 68, 2, 170–176 (2004).

Solé, R. V., S. Valverde and C. Rodriguez-Caso, "Convergent evolutionary paths in biological and technological networks", Evolution: Education and Outreach 4, 3, 415–426 (2011).

Sousa, T., T. Domingos and S. Kooijman, "From empirical patterns to theory: a formal metabolic theory of life", Philosophical Transactions of the Royal Society B: Biological Sciences 363, 1502, 2453–2464 (2008).

Sousa-Silva, C., S. Seager, S. Ranjan, J. J. Petkowski, Z. Zhan, R. Hu and W. Bains, "Phosphine as a biosignature gas in exoplanet atmospheres", Astrobiology 20, 2, 235–268 (2020).

Sparks, W. B., M. N. Parenteau, R. E. Blankenship, T. A. Germer, C. Patty, K. M. Bott, C. M. Telesco and V. S. Meadows, "Spectropolarimetry of primitive phototrophs as global surface biosignatures", arXiv:2010.02133 (2020).

Suissa, G., A. M. Mandell, E. T. Wolf, G. L. Villanueva, T. Fauchez and R. kumar Kopparapu, "Dim prospects for transmission spectra of ocean Earths around M stars", The Astrophysical Journal 891, 1, 58 (2020).

Sumner, D. Y., "Microbial communities: Tracing growth processes from Antarctic lakes to early earth to other planets", in "AGU Fall Meeting Abstracts", vol. 2014, pp. B41P–01 (2014).

Swain, M. R., R. Estrela, G. M. Roudier, C. Sotin, P. Rimmer, A. Valio, R. West, K. Pearson, N. Huber-Feely and R. T. Zellem, "Detection of an atmosphere on a rocky exoplanet", arXiv:2103.05657 (2021).

Temkin, O. N., A. V. Zeigarnik and D. Bonchev, *Chemical Reaction Networks: A Graph-Theoretical Approach* (CRC Press, 1996).

Thompson, M. A., J. Krissansen-Totton, N. Wogan, M. Telus and J. J. Fortney, "The case and context for atmospheric methane as an exoplanet biosignature", Proceedings of the National Academy of Sciences 119, 14, e2117933119 (2022).

Thorngren, D., P. Gao and J. J. Fortney, "The intrinsic temperature and radiative–convective boundary depth in the atmospheres of hot Jupiters", The Astrophysical Journal 884, 1, L6 (2019).

Tian, M. and K. Heng, "Atmospheric chemistry of secondary and hybrid atmospheres of super earths and sub-neptunes", arXiv preprint arXiv:2301.10217 (2023).

Tijhuis, L., M. C. Van Loosdrecht and J. Heijnen, "A thermodynamically based correlation for maintenance gibbs energy requirements in aerobic and anaerobic chemotrophic growth", Biotechnology and Bioengineering 42, 4, 509–519 (1993).

Tinetti, G., S. Rashby and Y. L. Yung, "Detectability of red-edge-shifted vegetation on terrestrial planets orbiting M stars", The Astrophysical Journal Letters 644, 2, L129 (2006).

Tremblay, L., M. R. Line, K. B. Stevenson, T. Kataria, R. T. Zellem, J. J. Fortney and C. V. Morley, "The detectability and constraints of biosignature gases in the near & mid-infrared from transit transmission spectroscopy", arXiv:1912.10939 [astro-ph] (2019).

Truong, N. and J. Lunine, "Volcanically extruded phosphides as an abiotic source of venusian phosphine", Proceedings of the National Academy of Sciences 118, 29, e2021689118 (2021).

Tsai, S.-M., J. R. Lyons, L. Grosheintz, P. B. Rimmer, D. Kitzmann and K. Heng, "VULCAN: an open-source, validated chemical kinetics Python code for exoplanetary atmospheres", The Astrophysical Journal Supplement Series 228, 2, 20 (2017).

Valentine, D. L., "Biogeochemistry and microbial ecology of methane oxidation in anoxic environments: a review", Antonie Van Leeuwenhoek 81, 271–282 (2002).

Venot, O., E. Hébrard, M. Agúndez, M. Dobrijevic, F. Selsis, F. Hersant, N. Iro and R. Bounaceur, "A chemical model for the atmosphere of hot Jupiters", Astron. Astrophys. Suppl. Ser. 546, A43 (2012).

Vidal, C., "Stellivore extraterrestrials? Binary stars as living systems", Acta Astronautica 128, 251–256 (2016).

Vidal, C., "Energy rate density as a technosignature: The case for stellivores", in "Life in the Universe: Big History, SETI and the Future of Humankind. IBHA & INAF-IASF MI Symposium 15-16 July 2019", (2020).

VijayaVenkataRaman, S., S. Iniyan and R. Goic, "A review of climate change, mitigation and adaptation", Renewable and Sustainable Energy Reviews 16, 1, 878–897 (2012).

Villanueva, G., M. Cordiner, P. Irwin, I. de Pater, B. Butler, M. Gurwell, S. Milam, C. Nixon, S. Luszcz-Cook, C. Wilson *et al.*, "No evidence of phosphine in the atmosphere of venus from independent analyses", Nature Astronomy 5, 7, 631–635 (2021).

Virtanen, P., R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt and SciPy 1.0 Contributors, "SciPy 1.0: fundamental algorithms for scientific computing in Python", Nature Methods 17, 261–272 (2020).

Visscher, H., C. V. Looy, M. E. Collinson, H. Brinkhuis, J. H. Van Konijnenburg-Van Cittert, W. M. Kürschner and M. A. Sephton, "Environmental mutagenesis during the end-permian ecological crisis", Proceedings of the National Academy of Sciences 101, 35, 12952–12956 (2004).

Vázquez, A., R. Pastor-Satorras and A. Vespignani, "Large-scale topological and dynamical properties of the Internet", Physical Review E 65, 6, 066130 (2002).

Walker, S. I., W. Bains, L. Cronin, S. DasSarma, S. Danielache, S. Domagal-Goldman, B. Kacar, N. Y. Kiang, A. Lenardic, C. T. Reinhard, W. Moore, E. W. Schwieterman, E. L. Shkolnik and H. B. Smith, "Exoplanet biosignatures: future directions", Astrobiology 18, 6, 779–824 (2018).

Walker, S. I., L. Cronin, A. Drew, S. Domagal-Goldman, T. Fisher, M. Line and C. Millsaps, "Probabilistic biosignature frameworks", in "Planetary Astrobiology", p. 477 (University of Arizona Press, 2020).

Walker, S. I., H. Kim and P. C. Davies, "The informational architecture of the cell", Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374, 2063, 20150057 (2016).

Wang, W.-X., B.-H. Wang, B. Hu, G. Yan and Q. Ou, "General dynamics of topology and traffic on weighted technological networks", Physical Review Letters 94, 18, 188702 (2005).

Welbanks, L., P. McGill, M. Line and N. Madhusudhan, "On the application of bayesian leave-one-out cross-validation to exoplanet atmospheric analysis", The Astronomical Journal 165, 3, 112 (2023).

West, G. B., J. H. Brown and B. J. Enquist, "A general model for the origin of allometric scaling laws in biology", Science 276, 5309, 122–126 (1997).

West, G. B., J. H. Brown and B. J. Enquist, "A general model for the structure and allometry of plant vascular systems", Nature 400, 6745, 664–667 (1999).

Willacy, K., M. Allen and Y. Yung, "A new astrobiological model of the atmosphere of titan", The Astrophysical Journal 829, 2, 79 (2016).

Wogan, N., J. Krissansen-Totton and D. C. Catling, "Abundant atmospheric methane from volcanism on terrestrial planets is unlikely and strengthens the case for methane as a biosignature", arXiv:2009.07761 (2020).

Wogan, N. F. and D. C. Catling, "When is chemical disequilibrium in Earth-like planetary atmospheres a biosignature versus an anti-biosignature? Disequilibria from dead to living worlds", The Astrophysical Journal 892, 2, 127 (2020).

Woitke, P., O. Herbort, C. Helling, E. Stüeken, M. Dominik, P. Barth and D. Samra, "Coexistence of CH4, CO2 and H2O in exoplanet atmospheres", arXiv:2010.12241 (2020).

Wong, M. L., A. Prabhu, J. Williams, S. M. Morrison and R. M. Hazen, "Toward network-based planetary biosignatures: atmospheric chemistry as unipartite, unweighted, undirected networks", Journal of Geophysical Research: Planets 128, 6, e2022JE007658 (2023).

Wordsworth, R. and R. Pierrehumbert, "Abiotic oxygen-dominated atmospheres on terrestrial habitable zone planets", The Astrophysical Journal Letters 785, 2, L20 (2014).

Wright, J. T., "Exoplanets and SETI", in "Handbook of Exoplanets", pp. 3405–3412 (Springer International Publishing, 2018).

Wright, K., D. Gleeson, C. Williamson, S. Grasby, J. Spear, R. Pappalardo and A. Templeton, "A sulfur-based glacial ecosystem as a model for the habitability of Europa and Mars", in "Astrobiology Science Conference 2010: Evolution and Life: Surviving Catastrophes and Extremes on Earth and Beyond", vol. 1538, p. 5526 (2010).

Wuchty, S., E. Ravasz and A.-L. Barabási, "The architecture of biological networks", in "Complex Systems Science in Biomedicine", pp. 165–181 (Springer, 2006).

Wunderlich, F., M. Scheucher, J. L. Grenfell, F. Schreier, C. Sousa-Silva, M. Godolt and H. Rauer, "Detectability of biosignatures on lhs 1140 b", Astronomy & Astrophysics 647, A48 (2021).

Zahnle, K., M. Claire and D. Catling, "The loss of mass-independent fractionation in sulfur due to a palaeoproterozoic collapse of atmospheric methane", Geobiology 4, 4, 271–283 (2006).

Zahnle, K., M. S. Marley, R. S. Freedman, K. Lodders and J. J. Fortney, "Atmospheric sulfur photochemistry on hot Jupiters", The Astrophysical Journal 701, 1, L20–L24 (2009).

APPENDIX A

PERMISSION

Permission was granted by all co-authors of the paper "Inferring Exoplanet Disequilibria with Multivariate Information in Atmospheric Reaction Networks", published in *The Astronomical Journal* 164 (2) in 2022, for inclusion in this work.

APPENDIX B

INFERRING EXOPLANET DISEQUILIBRIA WITH MULTIVARIATE INFORMATION IN ATMOSPHERIC REACTION NETWORKS: SUPPLEMENTARY INFORMATION

Methods

Atmospheric Modeling with VULCAN

The network analysis was a multi-step 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 (400K to 3000K, in *sim*100K increments) and, later, metallicities (0.1 to 300 times that of the solar system, evaluated on a log scale) and pressures (from 50mb to 150mb, 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.

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 upon 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 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 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 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_neighbor_degree methods.

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

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

Table B.1: K-S values for average clustering coefficient across all interpolated distributions, with respect to different K_{zz} values. Distributions become more distinguishable from each other as the difference in K_{zz} value between them increases

Mean Degree				
W.r.t Kzz =	$K_{zz} = 0$	$K_{zz} = 10^6$	$K_{zz} = 10^8$	$K_{zz} = 10^{10}$
$K_{zz} = 0$	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

Table B.2: K-S values for mean degree across all interpolated distributions, with respect to different K_{zz} values. Distributions become more distinguishable from each other as the difference in K_{zz} value between them increases.

an initial temperature and metallicity.

Results

Comparisons to Phi

The distribution of phi values also allowed distinguishing between different values of K_{zz}, though this was difficult to quantify due 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.

Average Shortest Path Length				
W.r.t Kzz =	$K_{zz} = 0$	$K_{zz} = 10^6$	$K_{zz} = 10^8$	$K_{zz} = 10^{10}$
$K_{zz} = 0$	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

Table B.3: K-S values for average shortest path length across all interpolated distributions, with respect to different K_{zz} values. Distributions become more distinguishable from each other as the difference in K_{zz} value between them increases

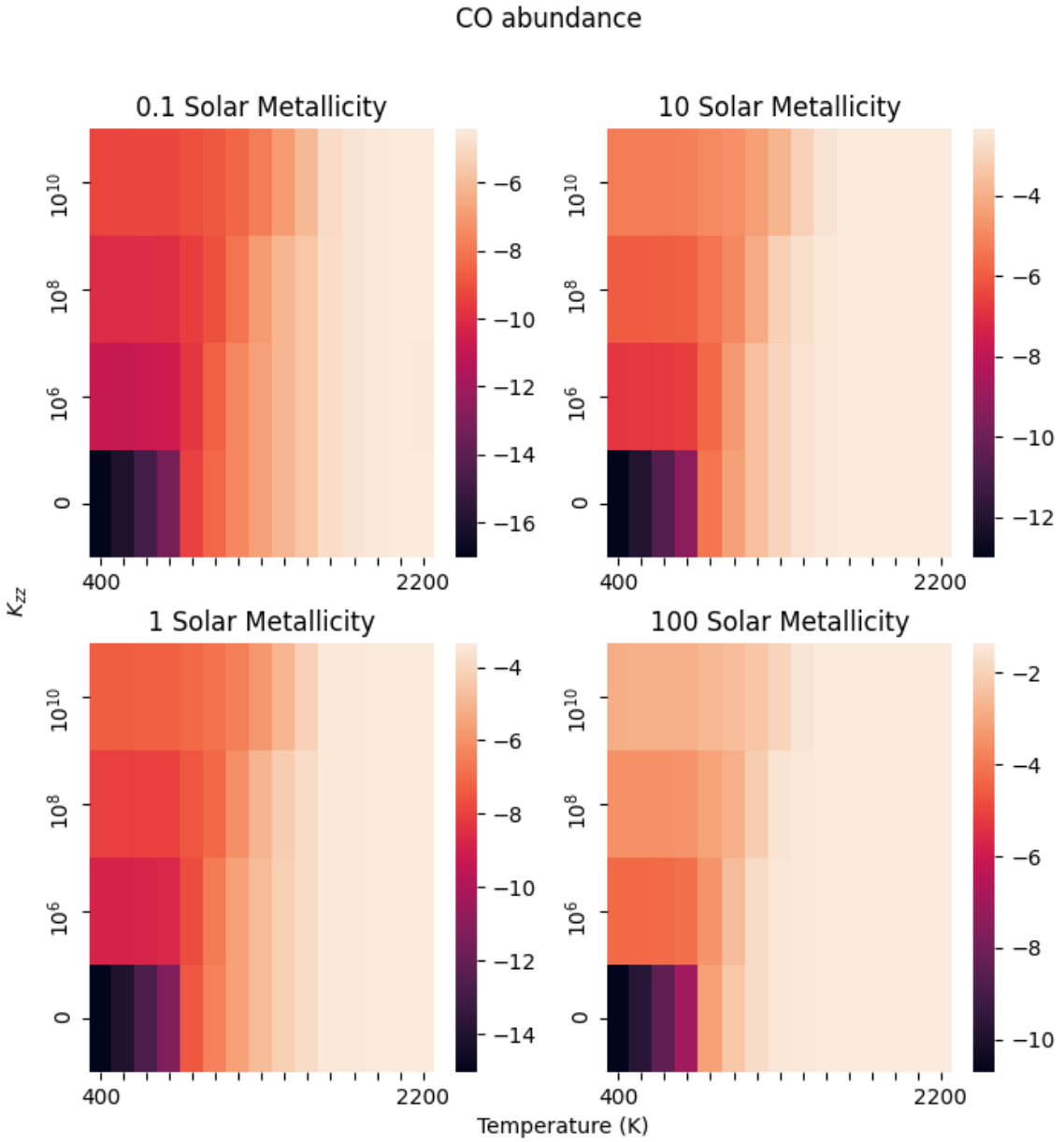


Figure B.1: The abundance of CO as a function of temperature and K_{zz}, at 0.1, 1, 10, and 100 times solar metallicity. The x-axis is temperature, y-axis is K_{zz}, and color corresponds to the abundance of CO (see color bar for each individual plot for scale).

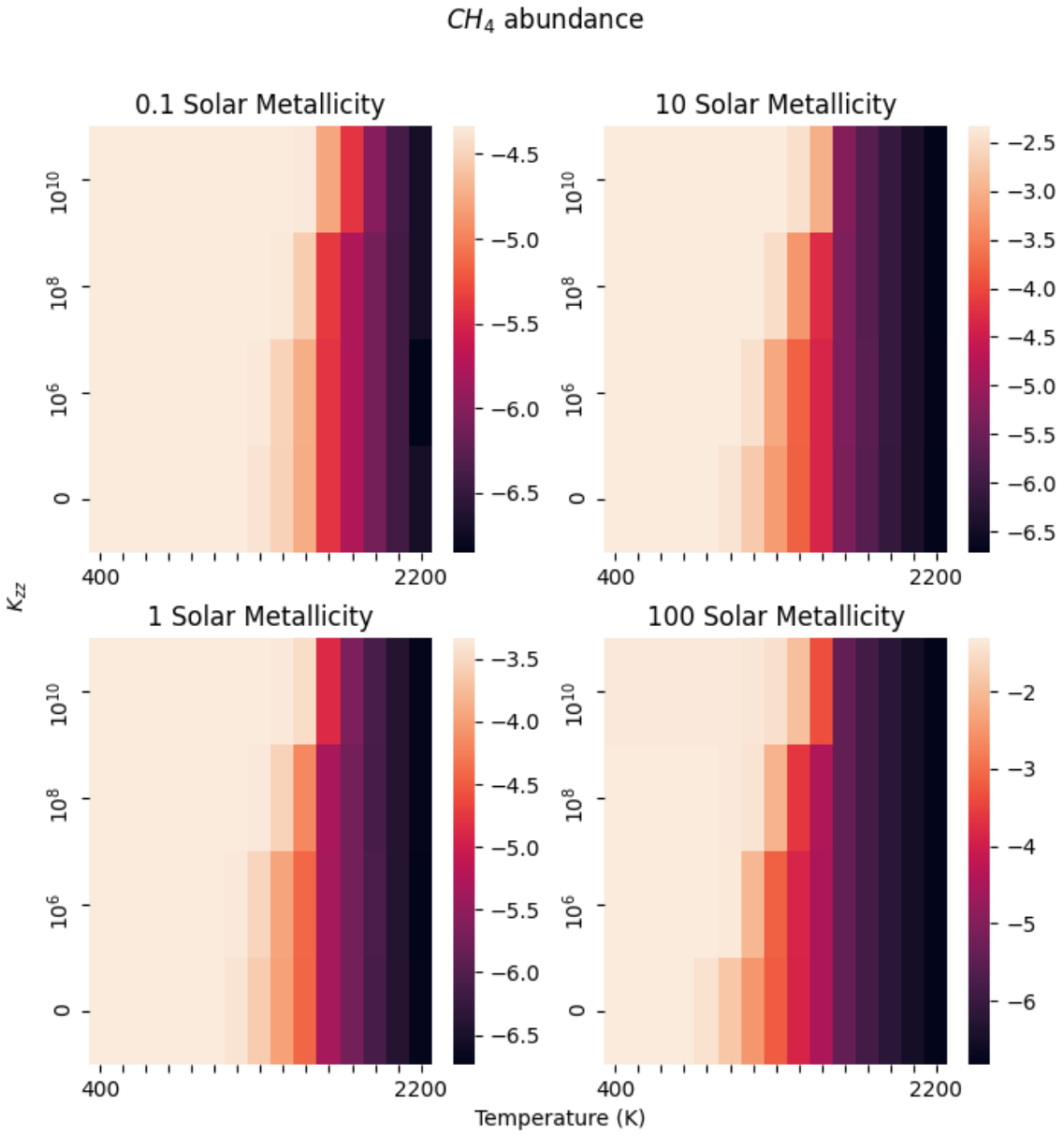


Figure B.2: The abundance of CH_4 as a function of temperature and K_{zz}, at 0.1, 1, 10, and 100 times solar metallicity. The x-axis is temperature, y-axis is Kzz, and color corresponds to the abundance of CH_4 (see color bar for each individual plot for scale).

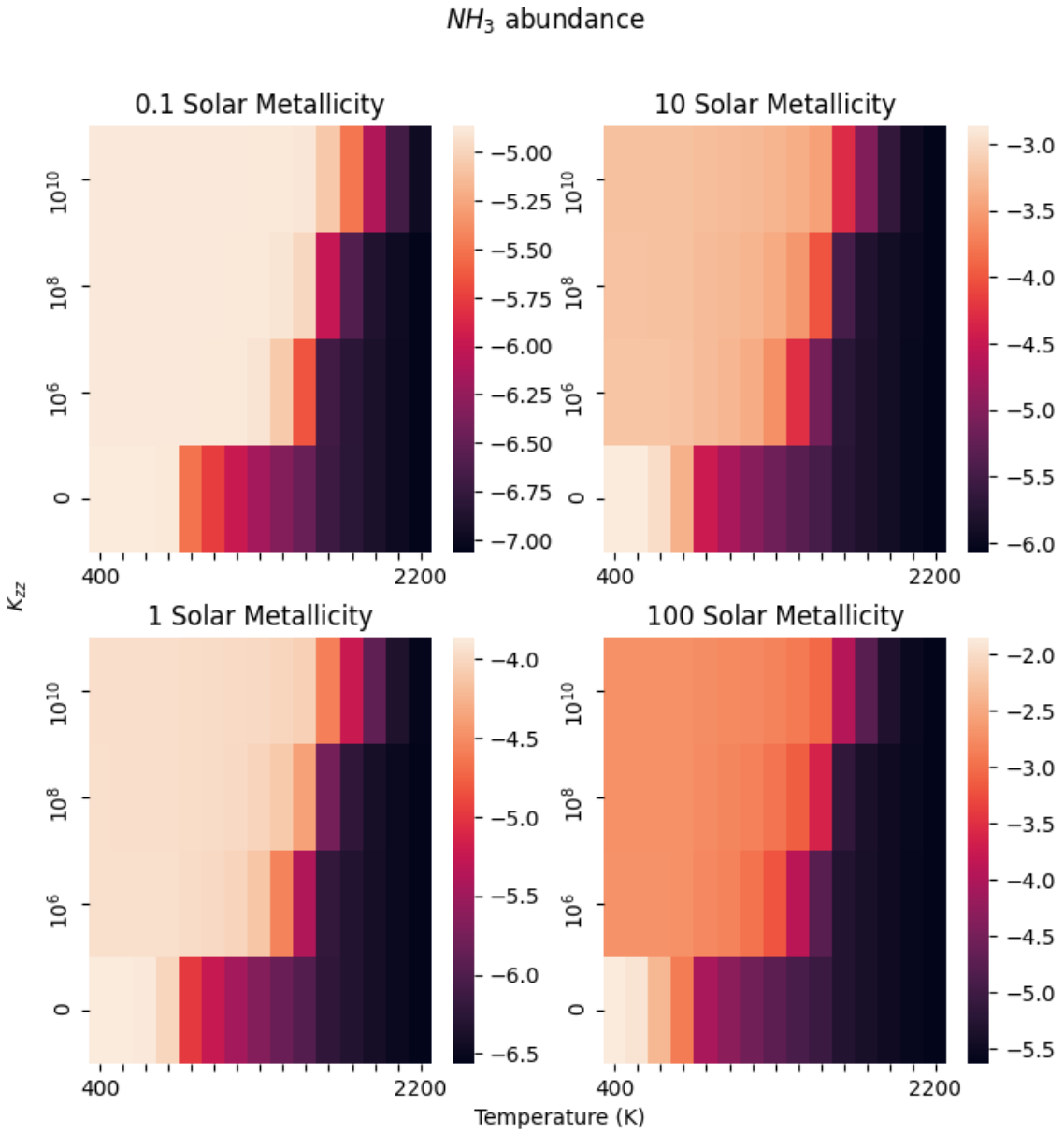


Figure B.3: The abundance of NH_3 as a function of temperature and K_{zz}, at 0.1, 1, 10, and 100 times solar metallicity. The x-axis is temperature, y-axis is K_{zz}, and color corresponds to the abundance of NH_3 (see color bar for each individual plot for scale).

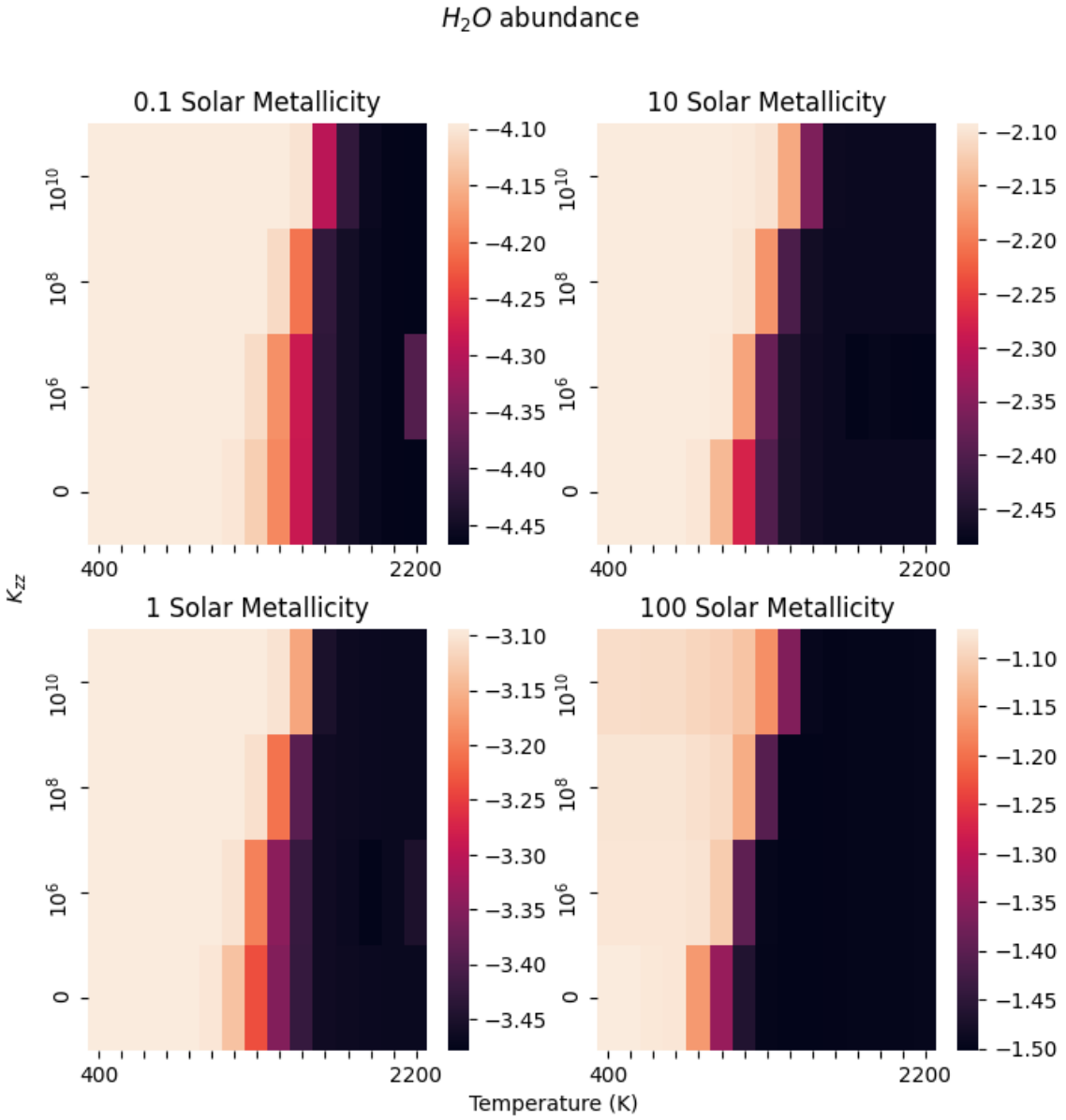


Figure B.4: The abundance of H_2O as a function of temperature and K_{zz}, at 0.1, 1, 10, and 100 times solar metallicity. The x-axis is temperature, y-axis is K_{zz}, and color corresponds to the abundance of H_2O (see color bar for each individual plot for scale).

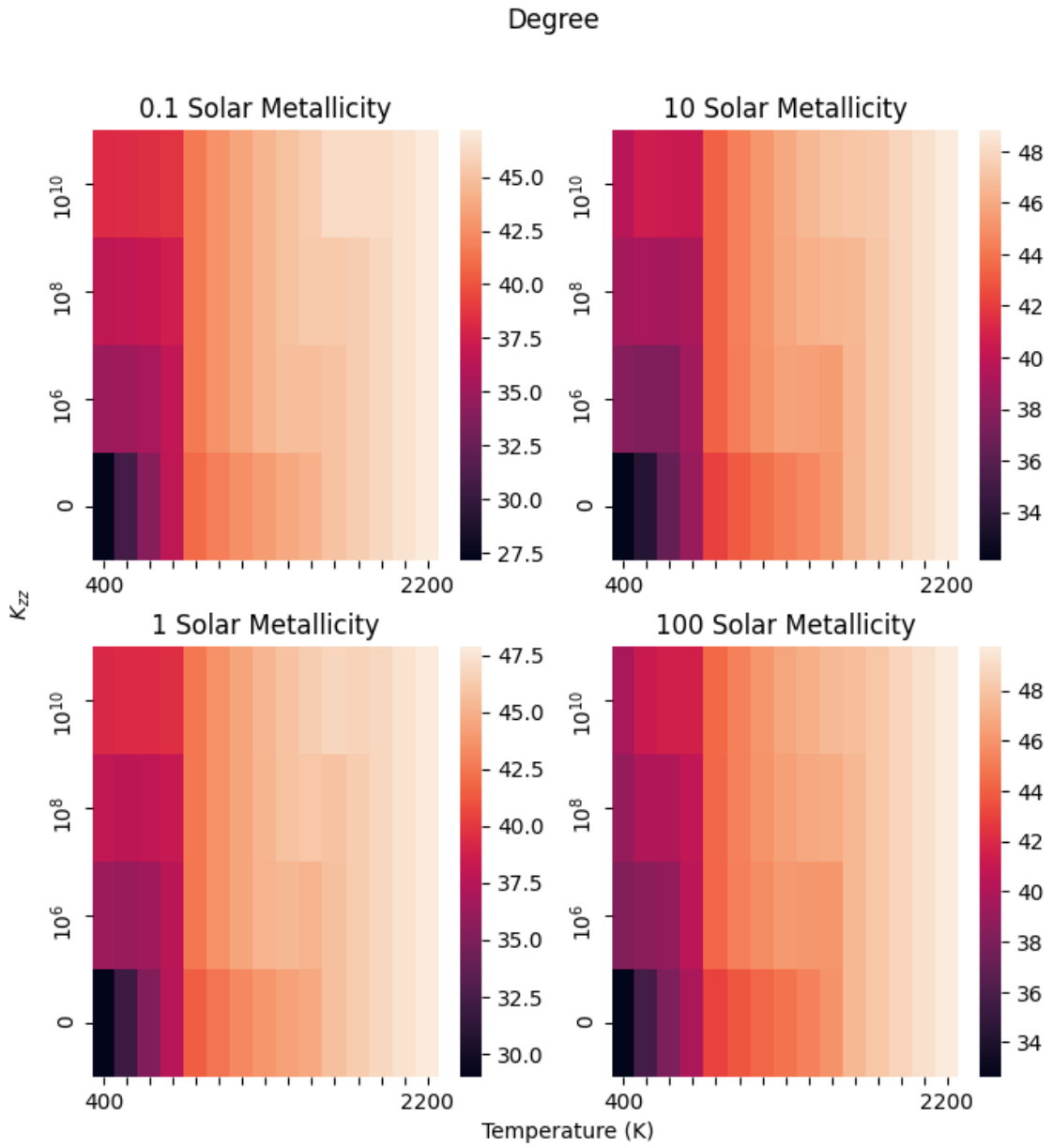


Figure B.5: Mean degree as a function of temperature and K_{zz}, at 0.1, 1, 10, and 100 times solar metallicity. The x-axis is temperature, y-axis is K_{zz}, and color corresponds to the weighted mean degree (see color bar for each individual plot for scale).

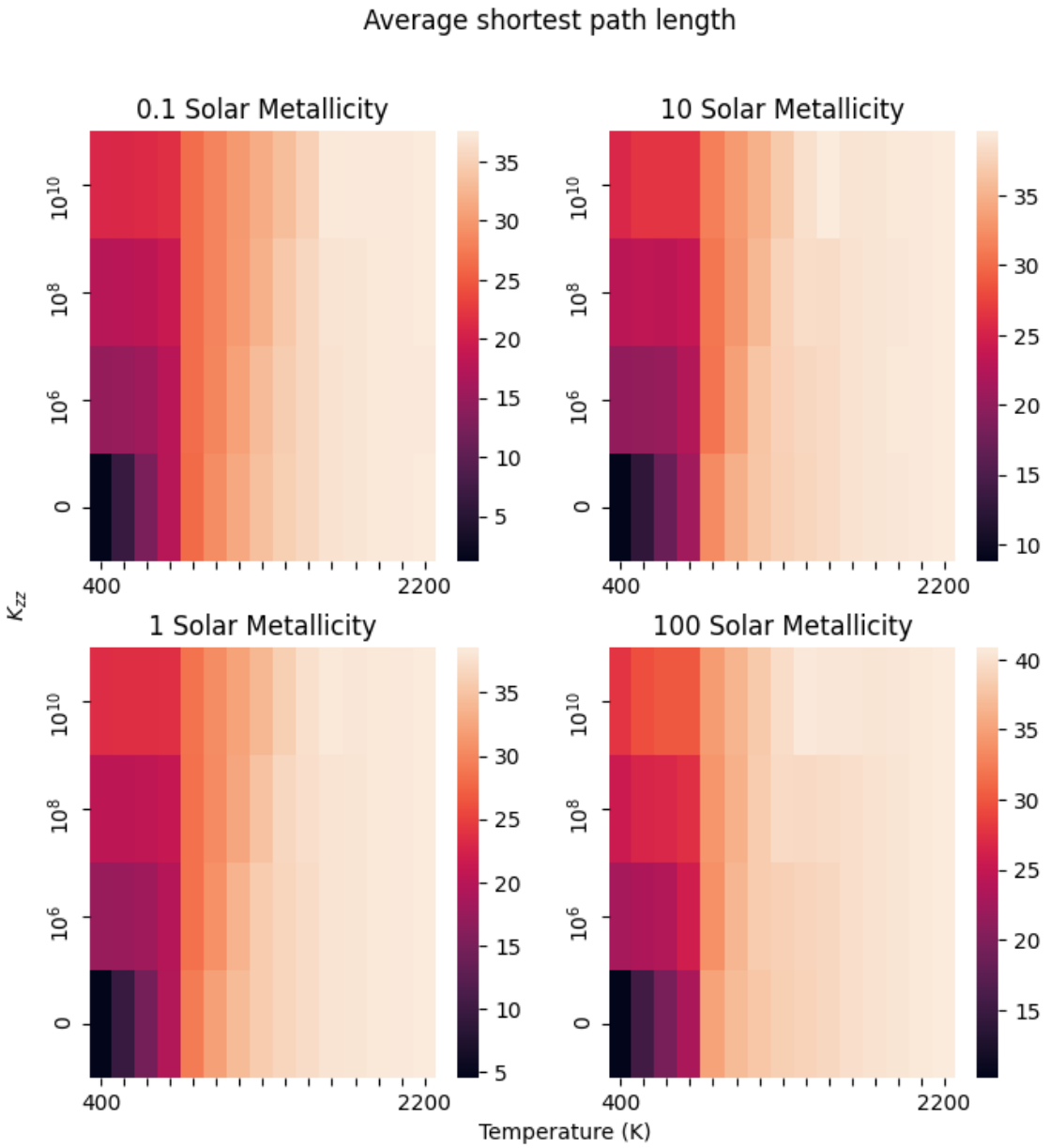


Figure B.6: Average shortest path length as a function of temperature and K_{zz}, at 0.1, 1, 10, and 100 times solar metallicity. The x-axis is temperature, y-axis is K_{zz}, and color corresponds to the weighted average shortest path length (see color bar for each individual plot for scale).

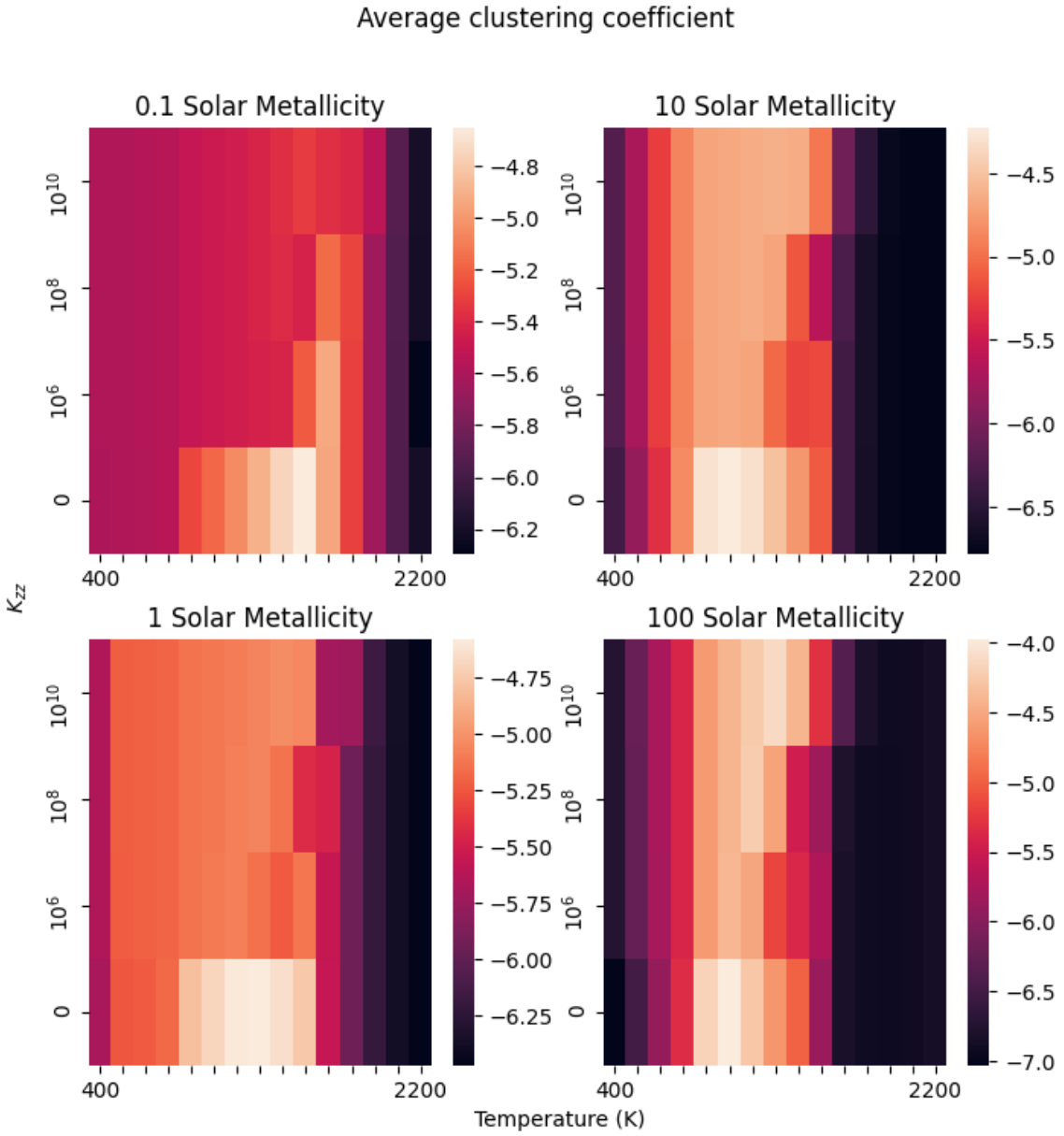


Figure B.7: Average clustering coefficient as a function of temperature and K_{zz}, at 0.1, 1, 10, and 100 times solar metallicity. The x-axis is temperature, y-axis is K_{zz}, and color corresponds to the weighted average clustering coefficient (see color bar for each individual plot for scale).

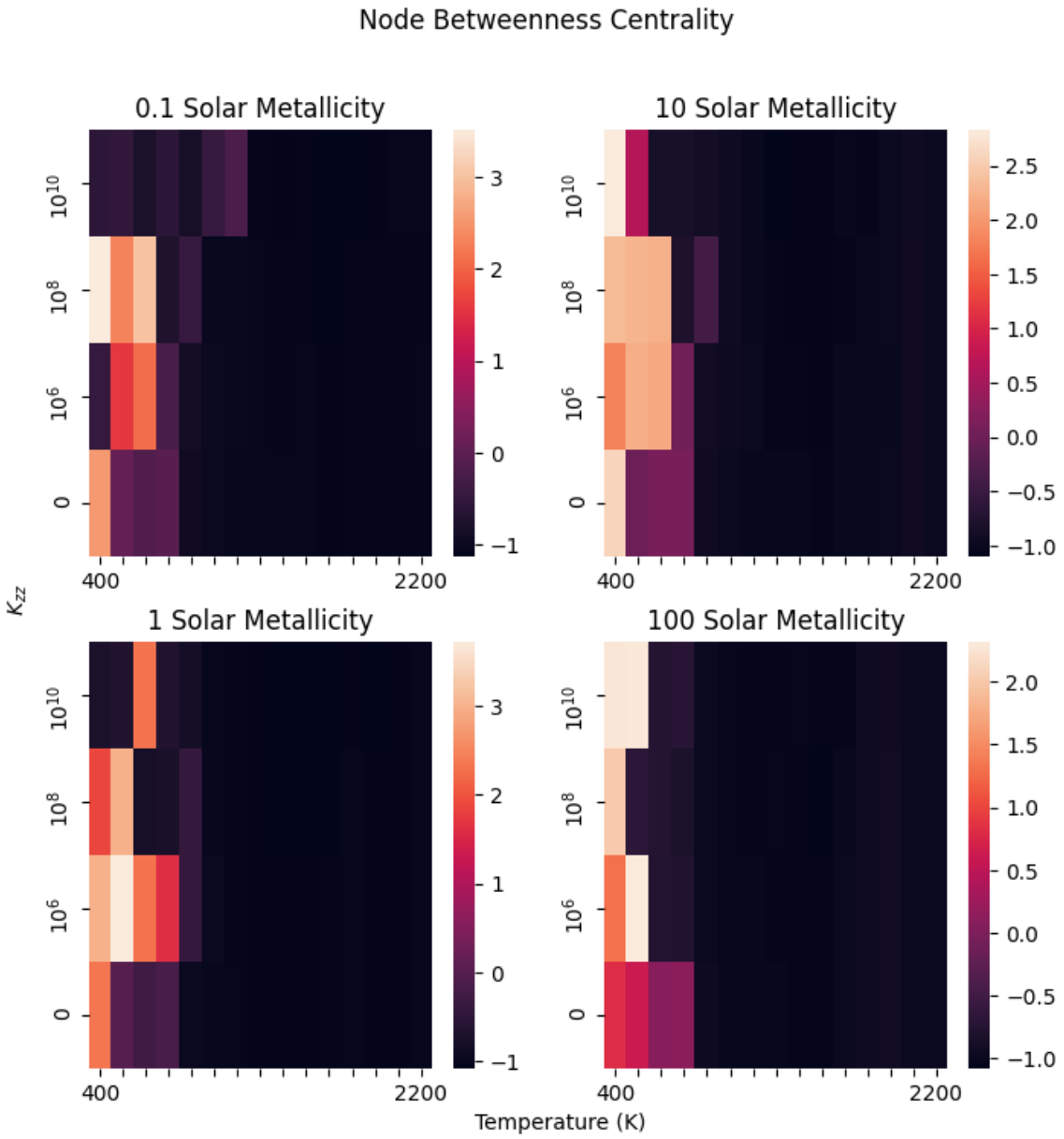


Figure B.8: Node betweenness centrality as a function of temperature and K_{zz}, at 0.1, 1, 10, and 100 times solar metallicity. The x-axis is temperature, y-axis is K_{zz}, and color corresponds to the weighted node betweenness centrality (see color bar for each individual plot for scale).

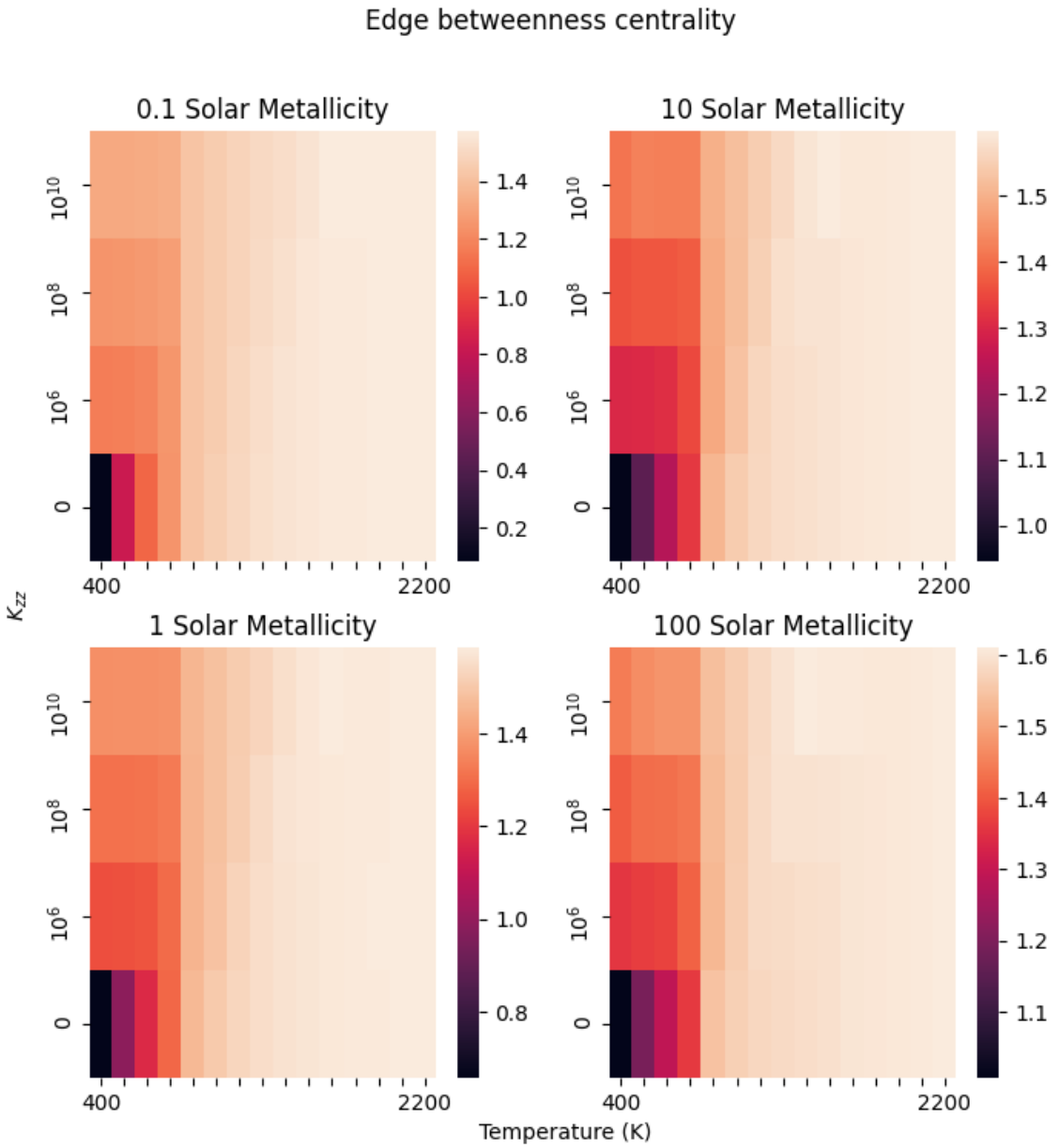


Figure B.9: Edge betweenness centrality as a function of temperature and K_{zz}, at 0.1, 1, 10, and 100 times solar metallicity. The x-axis is temperature, y-axis is K_{zz}, and color corresponds to the weighted edge betweenness centrality (see color bar for each individual plot for scale).

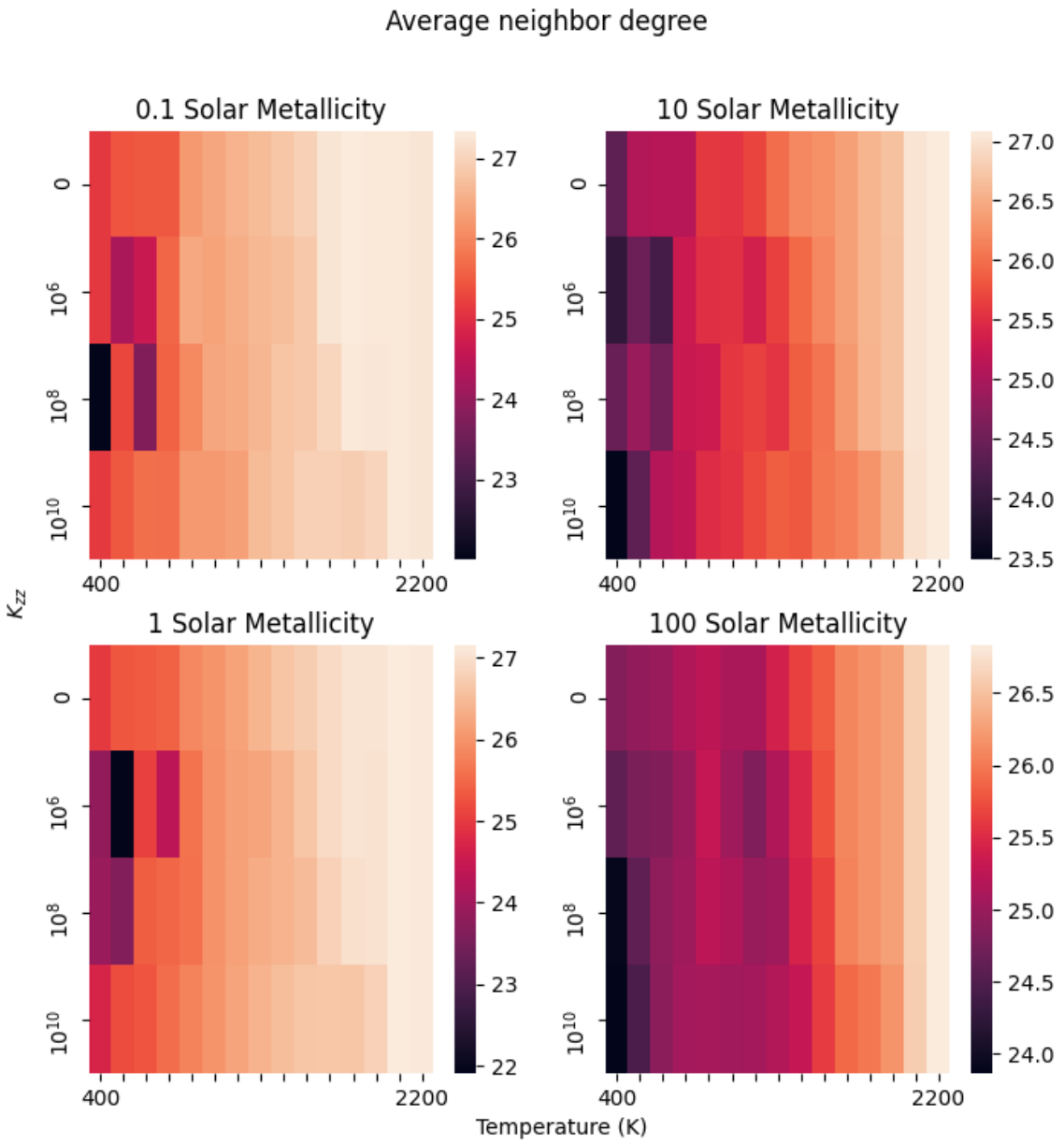


Figure B.10: Average neighbor degree as a function of temperature and K_{zz}, at 0.1, 1, 10, and 100 times solar metallicity. The x-axis is temperature, y-axis is K_{zz}, and color corresponds to the weighted average neighbor degree (see color bar for each individual plot for scale).

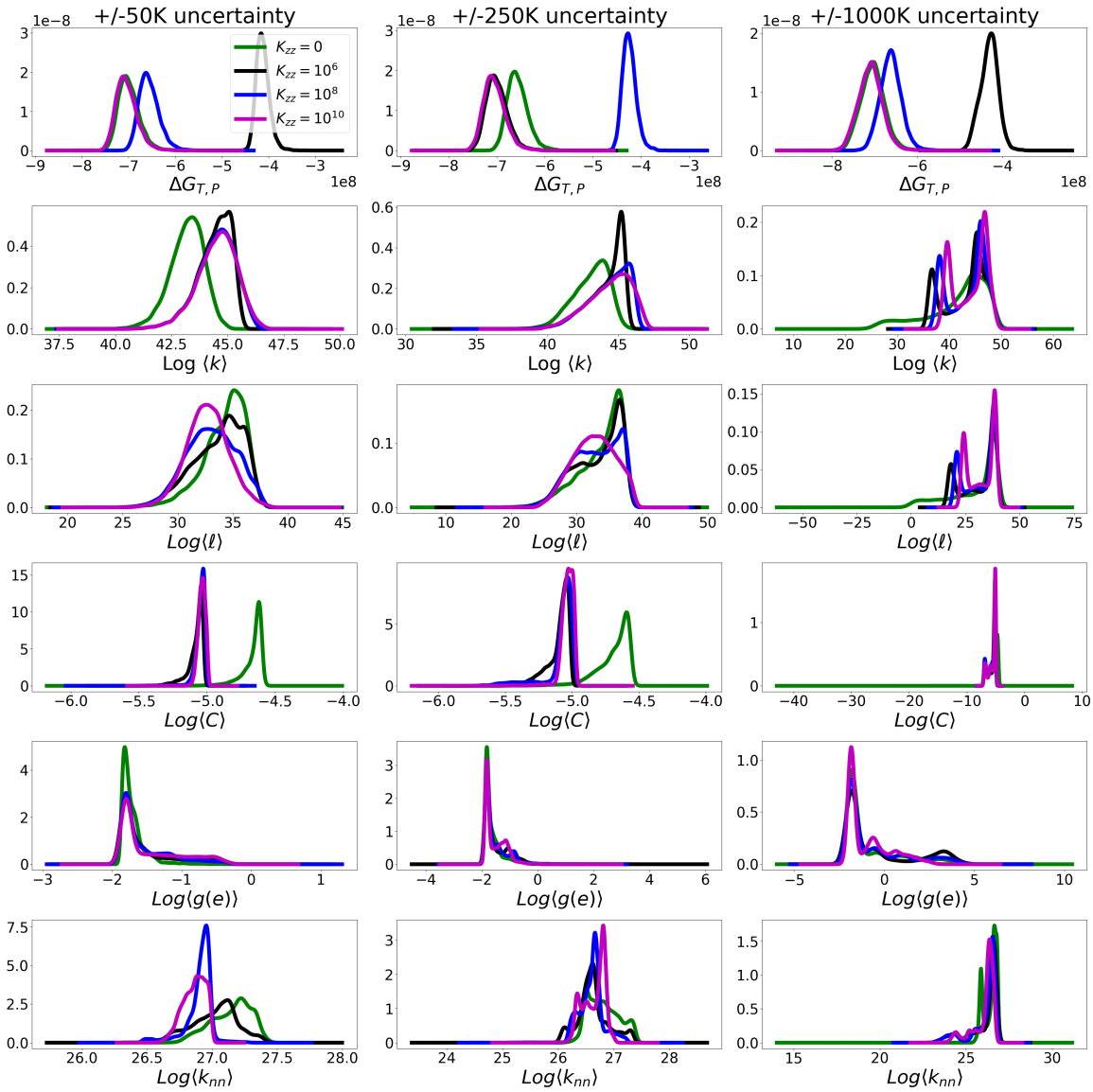


Figure B.11: Distributions of network and thermodynamic measurements, interpolated from a 10,000 point normal distribution of initial conditions centered on a temperature of 1200K, at K_{zz} values of 0, 10^6, 10^8, and 10^{10}, and uncertainties in temperature of 50K, 250K, and 1000K. The bimodal distributions seen in the ±1000K distributions are an artifact of discarding members of the distribution which fell outside the bounds of the original model grid, and thus could not be interpolated.

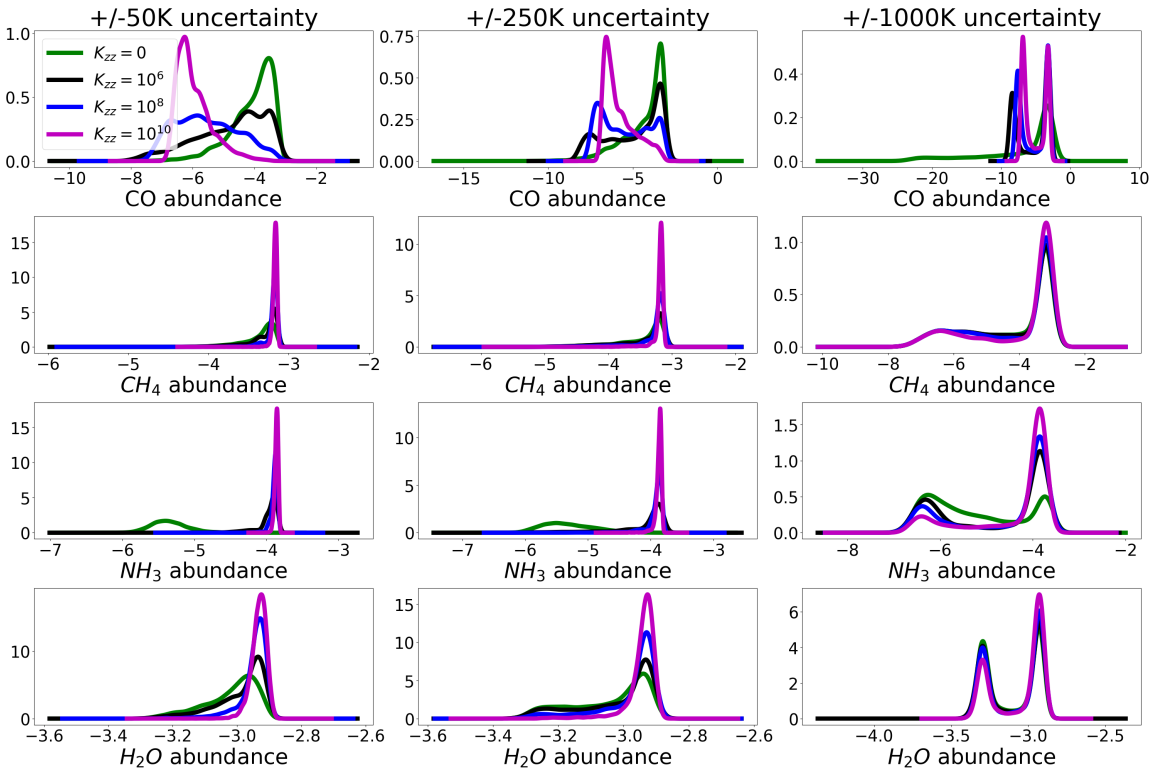


Figure B.12: Distributions of molecular abundances, interpolated from a 10,000 point normal distribution of initial conditions centered on a temperature of 1200K, at K_{zz} values of 0, 10^6, 10^8, and 10^{10}, and uncertainties in temperature of 50K, 250K, and 1000K. The bimodal distributions seen in 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.

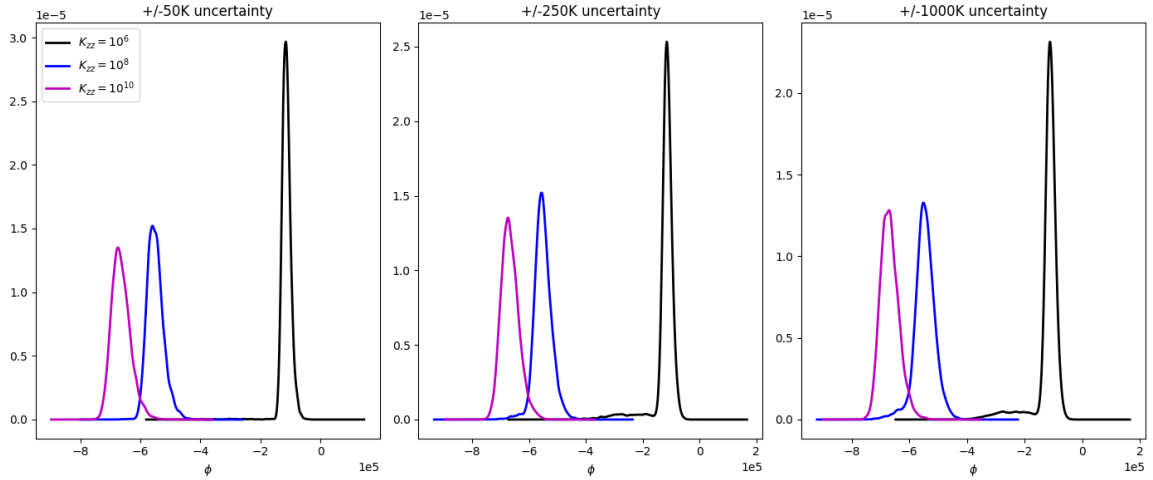


Figure B.13: Phi distributions over a range of uncertainties and K_{zz} values, drawn from a 10,000 point normal distribution of initial conditions centered at 1200K.

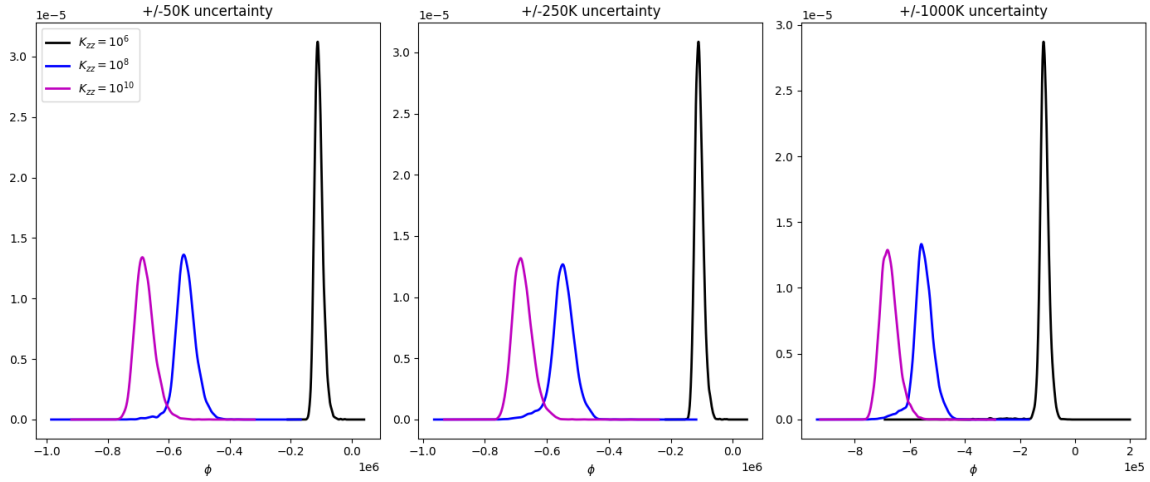


Figure B.14: Phi distributions over a range of uncertainties and K_{zz} values, drawn from a 10,000 point normal distribution of initial conditions centered at 2000K.

Node Betweenness Centrality				
W.r.t Kzz =	$K_{zz} = 0$	$K_{zz} = 10^6$	$K_{zz} = 10^8$	$K_{zz} = 10^{10}$
$K_{zz} = 0$	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 B.4: K-S values for node betweenness centrality across all interpolated distributions, with respect to different K_{zz} values.

Edge Betweenness Centrality				
W.r.t Kzz =	$K_{zz} = 0$	$K_{zz} = 10^6$	$K_{zz} = 10^8$	$K_{zz} = 10^{10}$
$K_{zz} = 0$	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 B.5: K-S values for edge betweenness centrality across all interpolated distributions, with respect to different K_{zz} values.

Average Neighbor Degree				
W.r.t Kzz =	$K_{zz} = 0$	$K_{zz} = 10^6$	$K_{zz} = 10^8$	$K_{zz} = 10^{10}$
$K_{zz} = 0$	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

Table B.6: K-S values for average neighbor degree across all interpolated distributions, with respect to different K_{zz} values.

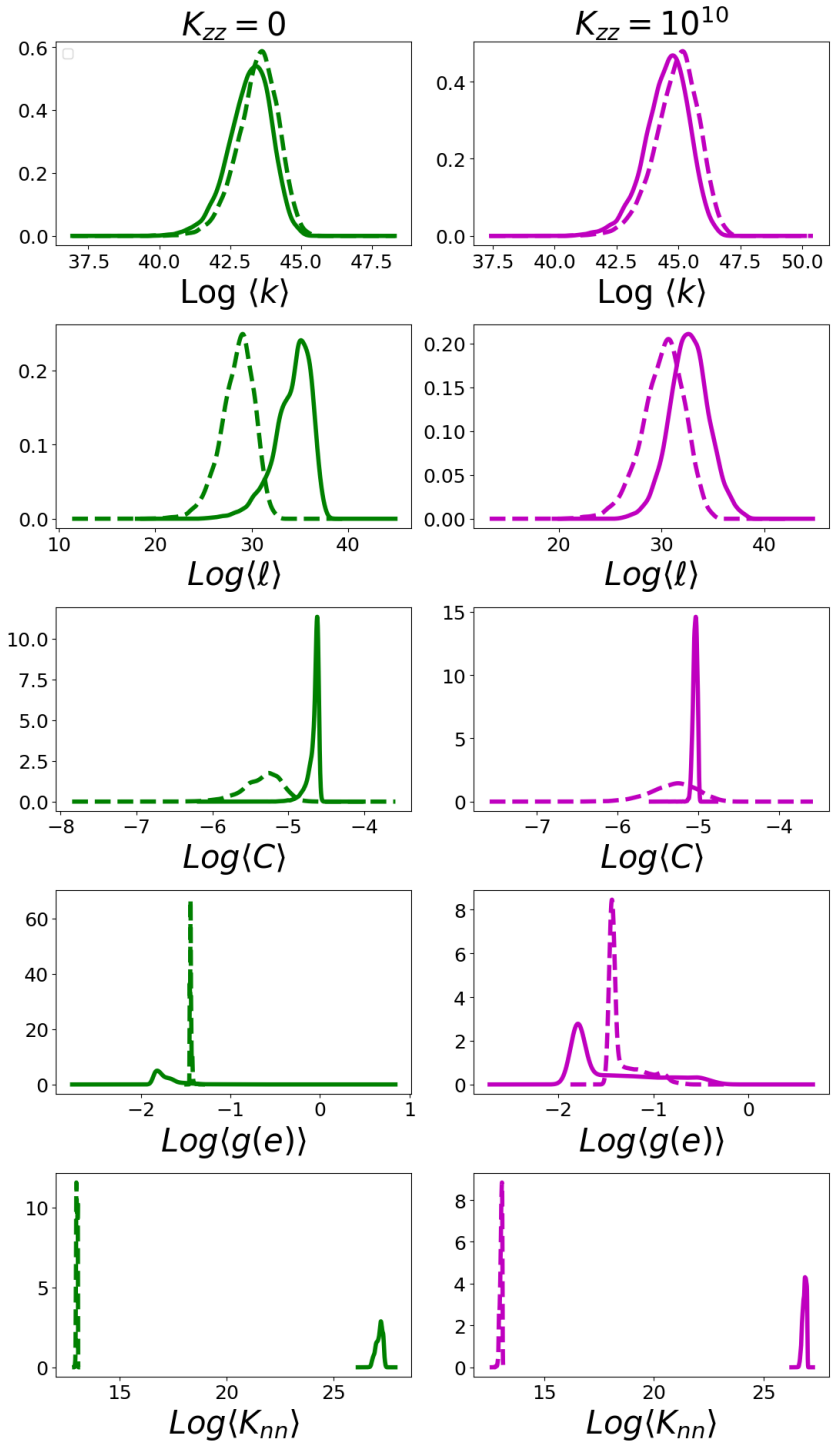


Figure B.15: T=1200K with perturbations with all C-bearing compounds removed. Distributions of network parameters when species containing carbon are removed from the network (dashed line) vs complete networks (solid line), at 1200K and 1 solar metallicity. Interestingly, while the absolute values of the parameters change, the overall shape of them remains the same, despite the significant reduction in the network.

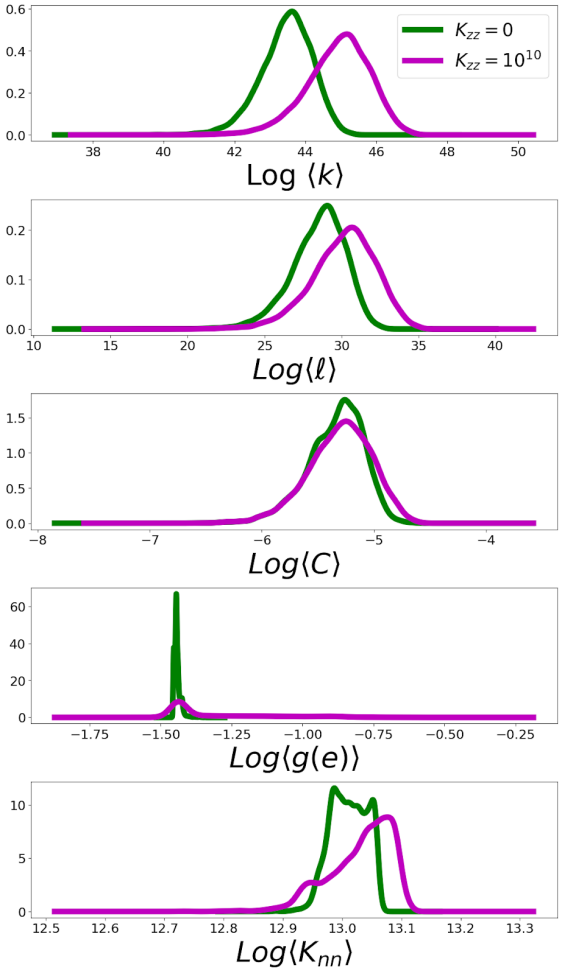


Figure B.16: Distributions of network parameters when species containing carbon are removed from the network, at 1200K and 1 solar metallicity. While the absolute values of the parameters changed compared to the complete networks, the distributions are still distinguishable from each other as a function of K_{zz}.

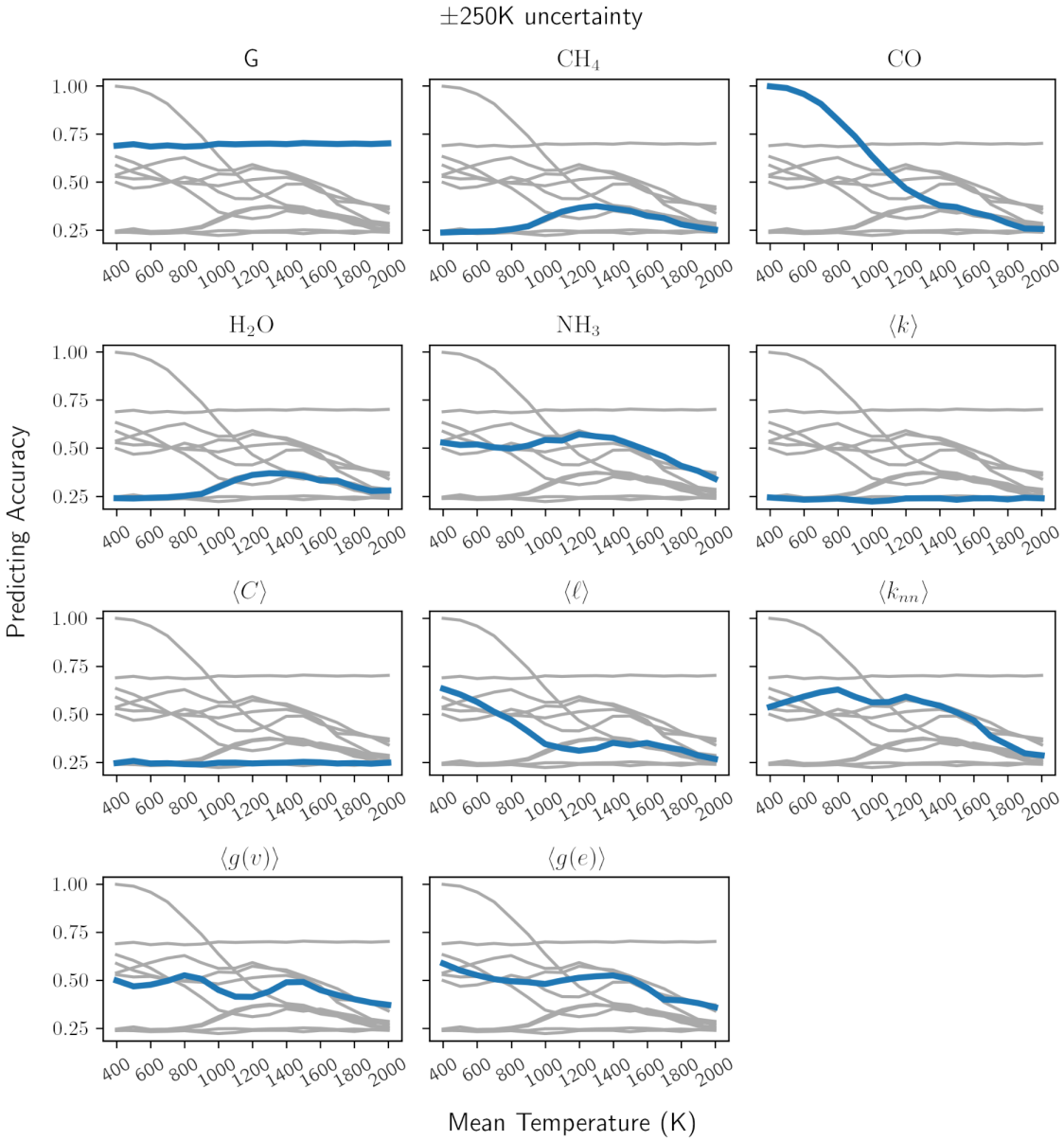


Figure B.17: Predictive power of variables for the disequilibrium state of the atmospheric chemical reaction networks. Each panel of the plot shows the predicting accuracy of an individual variable for atmospheric networks associated with K_{zz}= 0, 10^6, 10^8, and 10^{10} as a function of the mean temperature with uncertainty ± 250K (where grey lines show the data on all plots for ease of comparison). The variables include G (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.), the abundances of chemical species NH_3, CH_4, CO, and H_2O, and the average topological properties (see Table 3.1 and Section 3.2.2).

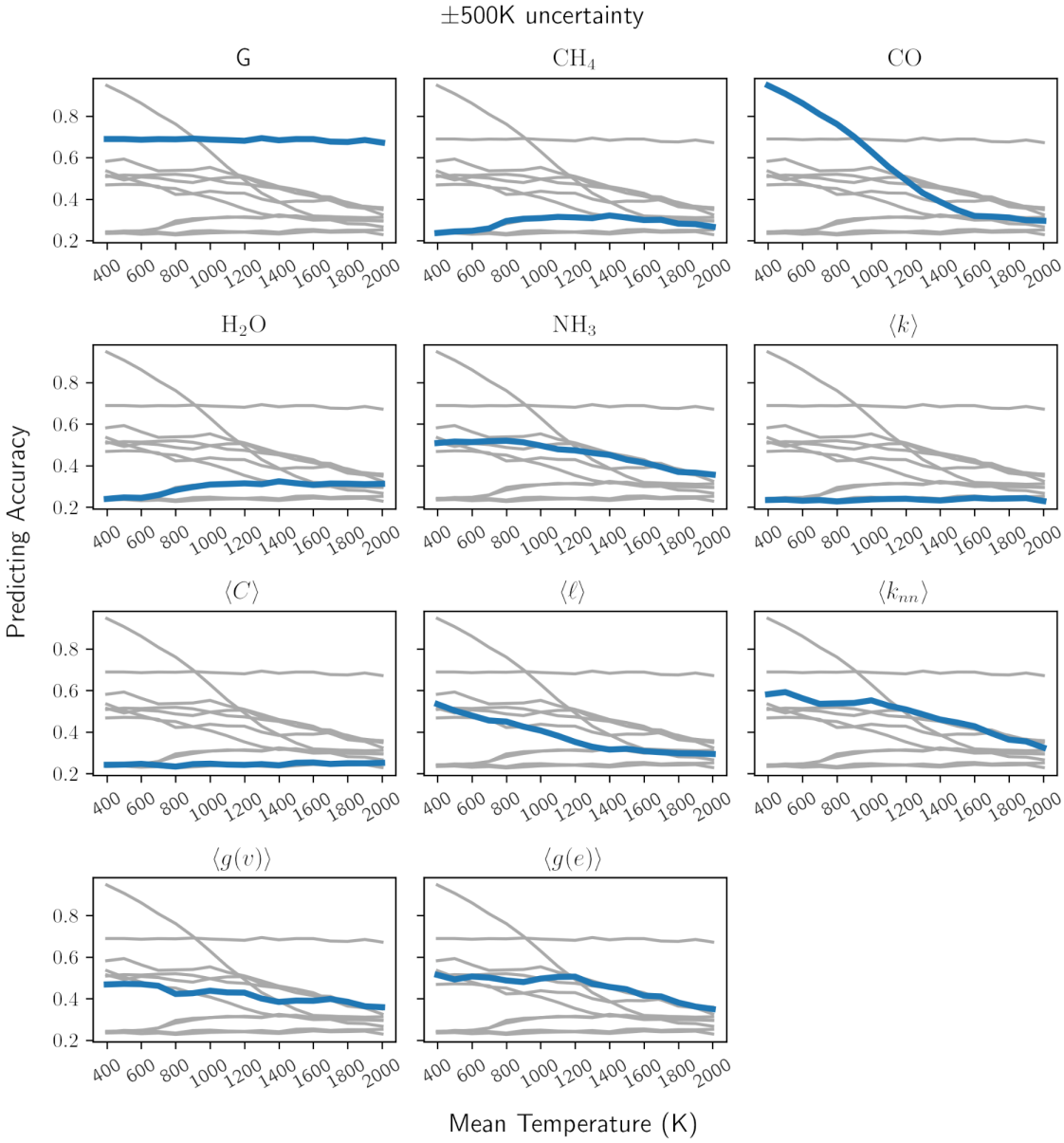


Figure B.18: Predictive power of variables for the disequilibrium state of the atmospheric chemical reaction networks. Each panel of the plot shows the predicting accuracy of an individual variable for atmospheric networks associated with $K_{zz} = 0$, 10^6, 10^8, and 10^{10} as a function of the mean temperature with uncertainty ±500K (where grey 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_2O, and the average topological properties (see Table 3.1 and Section 3.2.2)

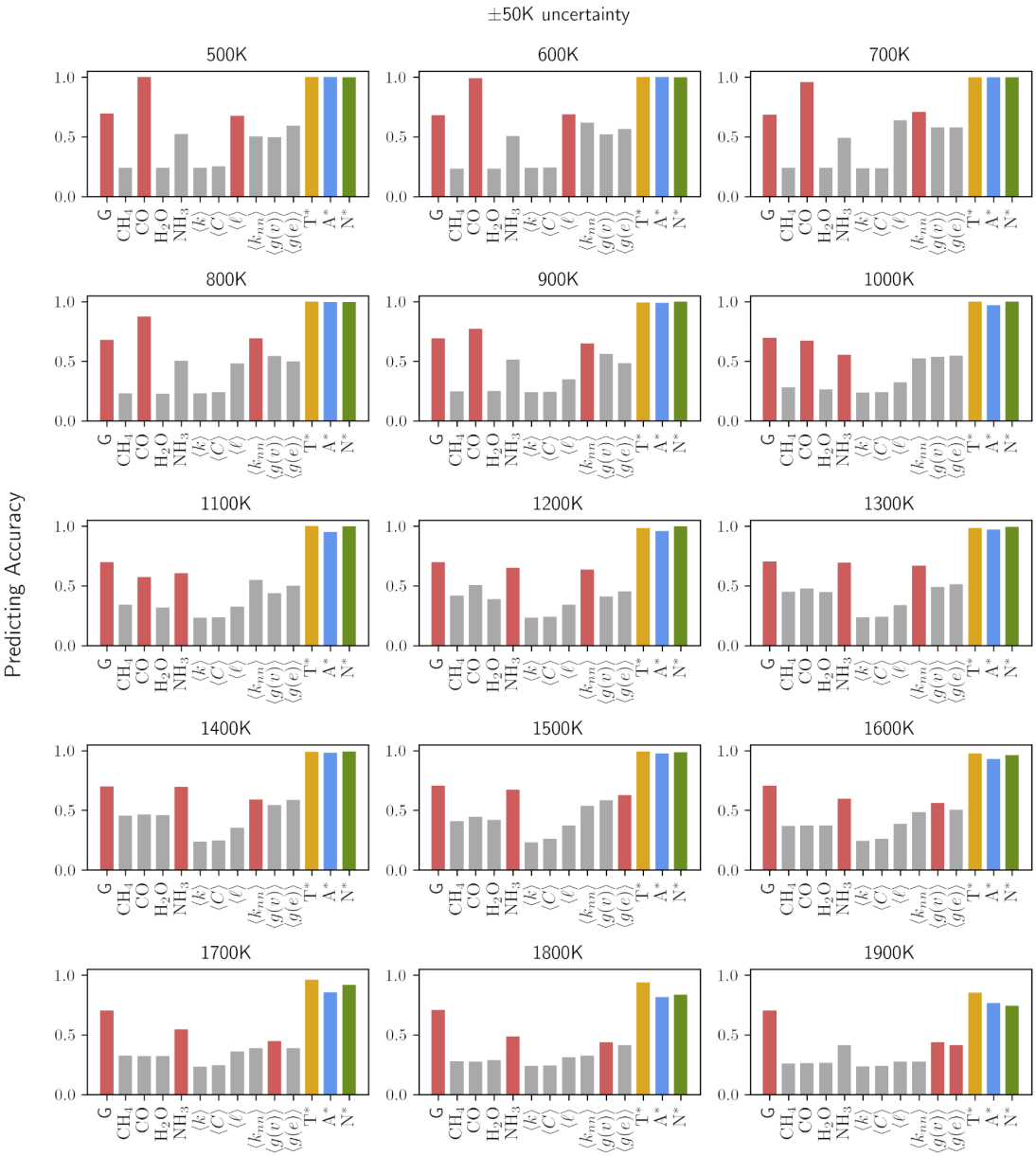


Figure B.19: Boosting up the predicting accuracy for the disequilibrium state of the atmospheric chemical reaction networks with multivariate information. Each panel of the plot shows the predicting accuracy of variables for atmospheric networks associated with $K_{zz} = 0$, 10^6, 10^8, and 10^{10} at a given mean temperature between 500 and 1900 K with uncertainty ±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 3.1 and Section 3.2.2). 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 3.9, respectively. The three different multivariable sets demonstrated higher predicting accuracy than individual variables even in the cases where the highest predictors are not included.

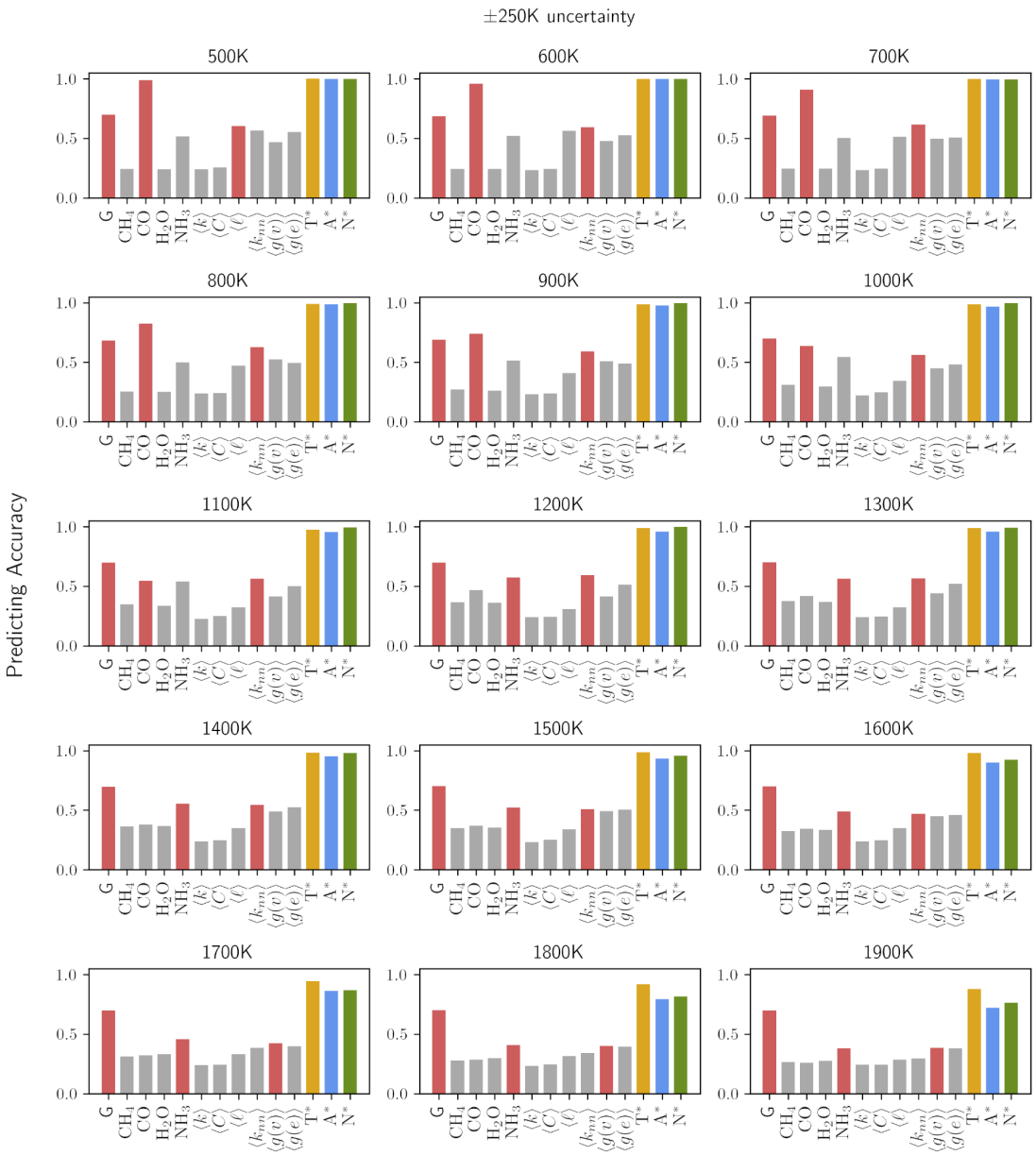


Figure B.20: Boosting up the predicting accuracy for the disequilibrium state of the atmospheric chemical reaction networks with multivariate information. Each panel of the plot shows the predicting accuracy of variables for atmospheric networks associated with $K_{zz} = 0$, 10^6, 10^8, and 10^{10} at a given mean temperature between 500K and 1900K with uncertainty ±250K. The three variables with the highest predictive power at each mean temperature compared to other individual variables (in grey) are colored red (for detailed information about topological variables used in this plot, see Table 3.1 and Section 3.2.2). 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.

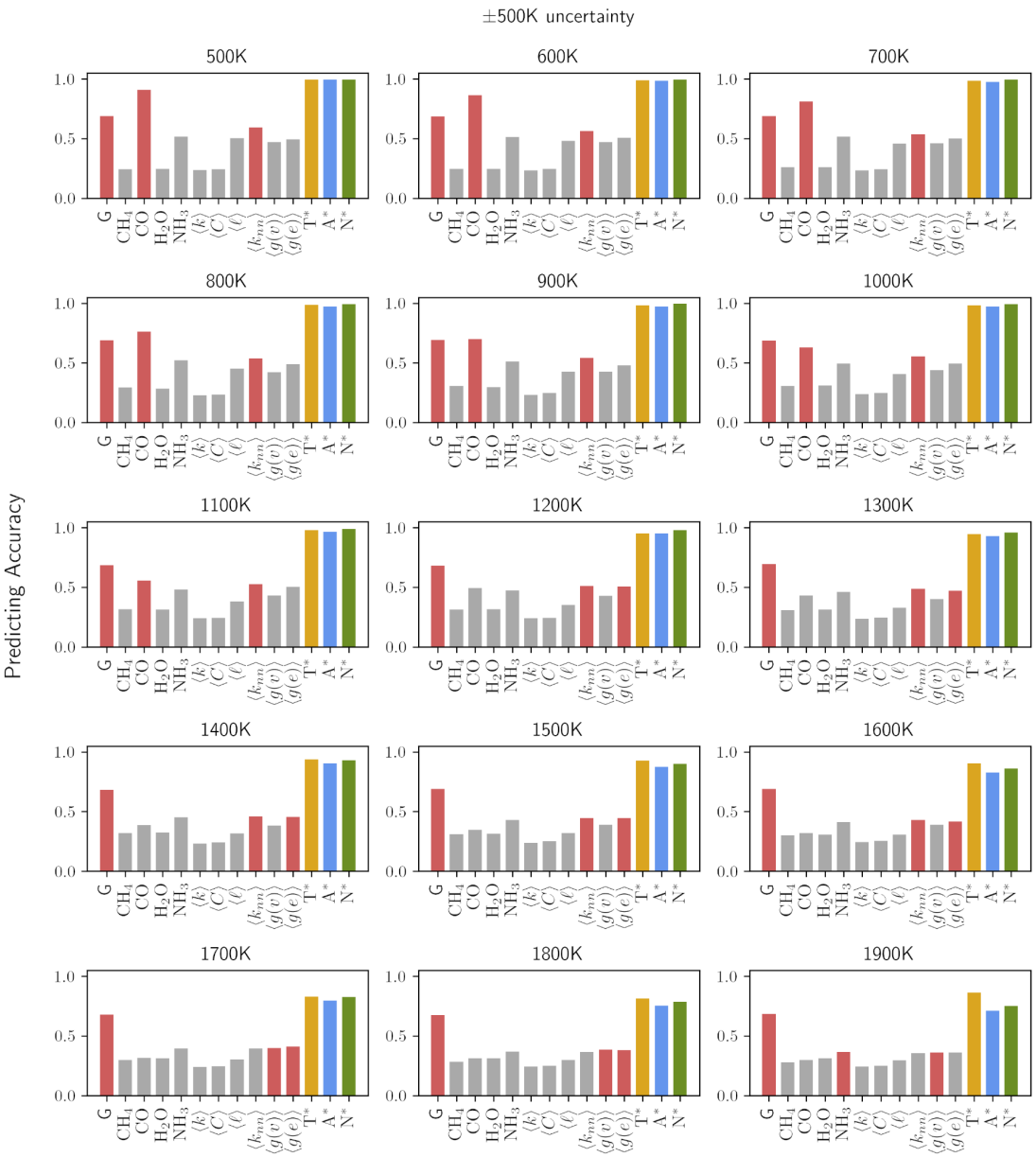


Figure B.21: Boosting up the predicting accuracy for the disequilibrium state of the atmospheric chemical reaction networks with multivariate information. Each panel of the plot shows the predicting accuracy of variables for atmospheric networks associated with $K_{zz} = 0$, 10^6, 10^8, and 10^{10} at a given mean temperature between 500K and 1900K with uncertainty ±500K. 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 3.1 and Section 3.2.2). 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.

APPENDIX C

ASSESSING BIOSIGNATURES AND TECHNOSIGNATURES IN THE ATMOSPHERES OF TERRESTRIAL EXOPLANETS USING NETWORK THEORY: SUPPLEMENTARY INFORMATION

Methods

Modeling with atmos

Atmos is a combined photochemical and climate model, though this work only made use of its photochemical aspect; see Section 2.3.2. Atmospheric photochemistry is modeled via a stack of 200 plane-parallel layers from the surface to an altitude of 100 km, with 0.5 km layer spacing.

The species in the model are divided up into long-lived and short-lived; short-lived species are not included in the Jacobian that is solved self-consistently at each time-step, as they are considered to already be at equilibrium. The mixing ratio of species found by solving flux and mass continuity equations in each layer simultaneously using a reverse-Euler method, providing exact solutions at steady state. Vertical transport is represented via molecular and eddy diffusion, and can be modified by user-set boundary conditions. Finally, the model includes a sophisticated method for modeling aerosol and haze formation.

The photochemical model is considered converged when redox is conserved and a re-run of the model using last run's output as initial conditions occurs quickly (i.e., $<$ 50 time steps). The models were allowed to run for a maximum of 5×10^{10} simulated years; this extreme length of time was sometimes required in order for the models to reach equilibrium due to limits on available computational power.

Table C.1: The initial conditions for all long-lived species present in the atmos template for the atmospheres used in the Archean Earth analogue populations of models.

Species	Deposition velocity (molecules/cm^2/s)	Constant mixing ratio (fractional)	Constant upward flux (molecules/cm^2/s)
O	1.0	1.00×10^{-8}	n/a
O_2	0	n/a	n/a
H_2O	0	n/a	n/a
H	1.0	n/a	n/a
OH	1.0	n/a	n/a
HO_2	1.0	n/a	n/a
H_2O_2	2.0×10^{-1}	n/a	n/a
H_2	2.4×10^{-4}	n/a	1.0×10^{10}
CO	1.2×10^{-4}	n/a	n/a
HCO	1.0	n/a	n/a
H_2CO	2.0×10^{-1}	n/a	n/a
CH_4 (abiotic surface flux)	0	n/a	10^{-2}-10^{8}
CH_4 (biotic surface flux)	0	n/a	10^{3}-10^{13}
CH_4 (anomalous high flux)	0	n/a	10^{3}-10^{13}

CH_4 (abiotic steady state)	0	10^{-10}-10^{-2}	n/a
CH_3	1.0	n/a	n/a
C_2H_6	0	n/a	n/a
NO	3.0×10^{-4}	n/a	n/a
NO_2	3.0×10^{-3}	n/a	n/a
HNO	1.0	n/a	n/a
O_3	7.0×10^{-2}	n/a	n/a
HNO_3	2.0×10^{-1}	n/a	n/a
N	0	n/a	n/a
C_3H_2	0	n/a	n/a
C_3H_3	0	n/a	n/a
CH_3C_2H	0	n/a	n/a
CH_2CCH_2	0	n/a	n/a
C_3H_5	0	n/a	n/a
C_2H_5CHO	0	n/a	n/a
C_3H_6	0	n/a	n/a
C_3H_7	0	n/a	n/a
C_3H_8	0	n/a	n/a
C_2H_4OH	0	n/a	n/a
C_2H_2OH	0	n/a	n/a
C_2H_5	0	n/a	n/a
C_2H_4	0	n/a	n/a
CH	0	n/a	n/a
CH_3O_2	0	n/a	n/a
CH_3O	0	n/a	n/a
CH_2CO	0	n/a	n/a
CH_3CO	0	n/a	n/a
CH_3CHO	0	n/a	n/a
CH_23	0	n/a	n/a
C_2H	0	n/a	n/a
C_2	0	n/a	n/a
C_2H_3	0	n/a	n/a
HCS	0	n/a	n/a
CS_2	0	n/a	n/a
CS	0	n/a	n/a
OCS	0	n/a	n/a
S	0	n/a	n/a
HS	0	n/a	n/a
H_2S	2.0×10^{-2}	n/a	3.5×10^8
SO_3	0	n/a	n/a
S_2	0	n/a	n/a
HSO	1.0	n/a	n/a
H_2SO_4	1.0	n/a	n/a
SO_2	1.0	n/a	3.5×10^9

SO	0	n/a	n/a
CO_2	0	2.00×10^{-2}	n/a
SO_4 aerosol	1.0×10^{-2}	n/a	n/a
S_8 aerosol	1.0×10^{-2}	n/a	n/a
hydrocarbon aerosol 1	1.0×10^{-2}	n/a	n/a
hydrocarbon aerosol 2	1.0×10^{-2}	n/a	n/a

Table C.2: The initial conditions for all long-lived species present in the atmos template for the atmospheres used in the modern Earth analogue populations of models.

Species	Deposition velocity (molecules/cm^2/s)	Constant mixing ratio (fractional)	Constant upward flux (molecules/cm^2/s)
O	1.0	n/a	n/a
O_2	n/a	0.02-0.5	n/a
H_2O	0	n/a	n/a
H	1.0	n/a	n/a
OH	1.0	n/a	n/a
HO_2	1.0	n/a	n/a
H_2O_2	0.5	n/a	n/a
H_2	n/a	6.1×10^{-8}-0.19	3.00×10^{10}
CO	n/a	n/a	3.70×10^{11}
HCO	1.0	n/a	n/a
H_2CO	1.0×10^{-1}	n/a	n/a
CH_3O	0	n/a	n/a
CH_2OH	0	n/a	n/a
CH_3O_2	0	n/a	n/a
CH_3OOH	0	n/a	n/a
CH_4	n/a	n/a	3.00×10^{8}
CH_3	1.0	n/a	n/a
C_2H_6	n/a	n/a	9.00×10^{8}
NO	n/a	n/a	1.00×10^{9}
NO_2	3.0×10^{-3}	n/a	n/a
HNO	1.0	n/a	n/a
H_2S	n/a	n/a	3.00×10^{8}
HS	3.0×10^{-3}	n/a	n/a
S	1.0	n/a	n/a
SO	3.0×10^{-4}	n/a	n/a
SO_2	1.0	n/a	9.00×10^{9}
H_2SO_4	n/a	n/a	7.00×10^{8}

HSO	1.0	n/a	n/a
S_2	0	n/a	n/a
S_4	0	n/a	n/a
S_8	0	n/a	n/a
SO_3	0	n/a	n/a
OCS	n/a	n/a	1.50×10^{7}
S_3	0	n/a	n/a
O_3	0	n/a	n/a
HNO_3	2.0×10^{-1}	n/a	n/a
N	0	n/a	n/a
NO_3	0	n/a	n/a
N_2O	n/a	n/a	1.53×10^{9}
HO_2NO_2	2.0×10^{-1}	n/a	n/a
N_2O_5	0	n/a	n/a
NH	0	n/a	n/a
NH_2	0	n/a	n/a
NH_3	0	n/a	n/a
CO_2	5.0×10^{-5}	8×10^{-8}-0.4	6.875×10^{8}
CS	0	n/a	n/a
CS_2	0	n/a	n/a
HCS	0	n/a	n/a
C_2H_6S	n/a	n/a	3.3×10^{9}
C_2H_5S	0	n/a	n/a
C_2H_5SO	0	n/a	n/a
C_2H_5SO2	0	n/a	n/a
CH_3S	0	n/a	n/a
CH_3SO	0	n/a	n/a
CH_3SO_2	0	n/a	n/a
CH_3SO_3	0	n/a	n/a
CH_4SO_2	0	n/a	n/a
C_2H_6SO	0	n/a	n/a
C_2H_7SO	0	n/a	n/a
CCl_2F_2	n/a	4.2×10^{-11}-2.3×10^{-6}	n/a
$CClF_2$	6.3×10^{-10}	n/a	n/a
ClO	6.3×10^{-10}	n/a	n/a
ClOO	6.3×10^{-10}	n/a	n/a
Cl	6.3×10^{-10}	n/a	n/a
SO_4 aersol	1.0×10^{-2}	n/a	n/a
S_8 aersol	6.3×10^{-10}	n/a	n/a

Results

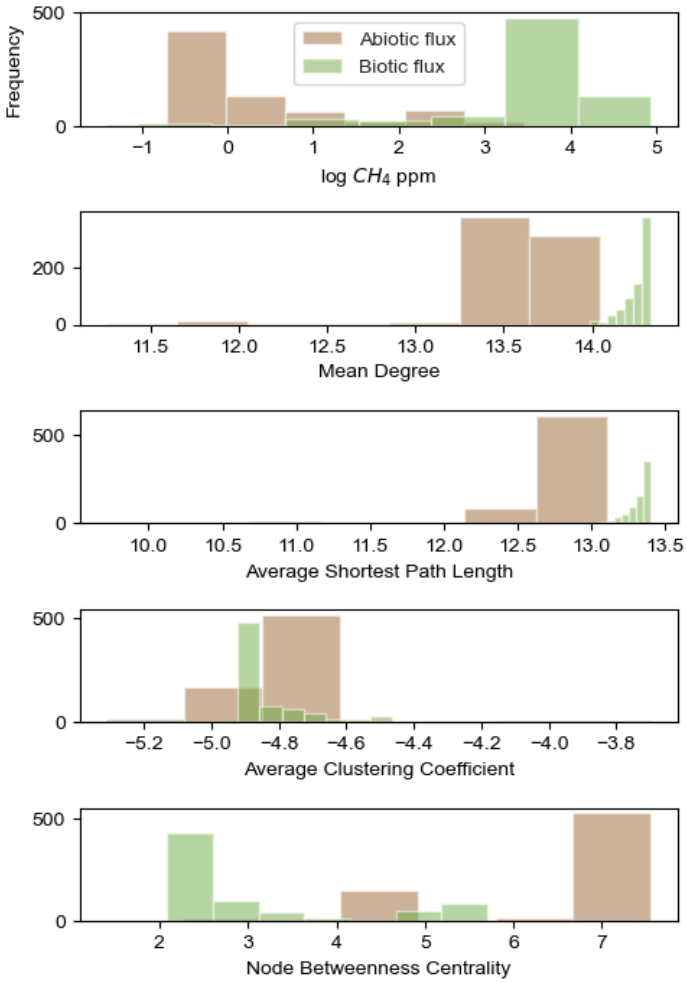


Figure C.1: Distributions of the frequency (y-axis) of the given network metric value and CH_4 abundance in log ppm (x-axis). Brown represents models where methane is modeled as an abiotic surface flux. Green represents models where methane is modeled as a biotic surface flux, with the methanogenesis pathway explicitly included in the CRN.

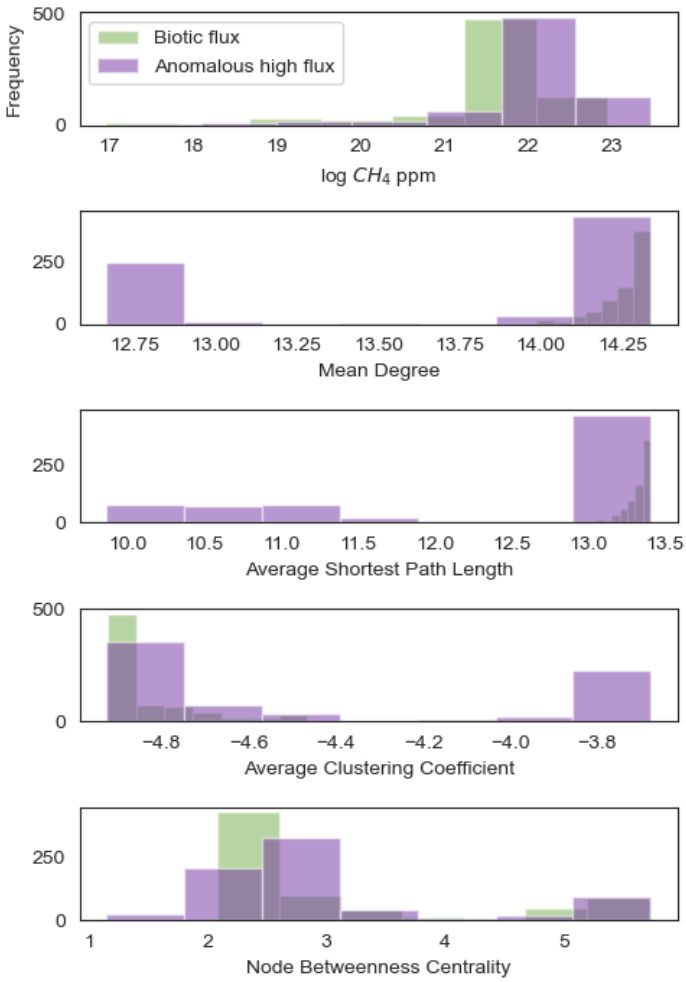


Figure C.2: Distributions of the frequency (y-axis) of the given network metric value and CH_4 abundance in log ppm (x-axis). Purple represents models where methane was modeled as originating from a surface flux of undetermined origin ("anomalous high flux"): while the flux was of the same magnitude as the biotic models, the methanogenesis pathway was not included in the model. Green represents models where methane is modeled as a biotic surface flux, with the methanogenesis pathway explicitly included in the CRN.

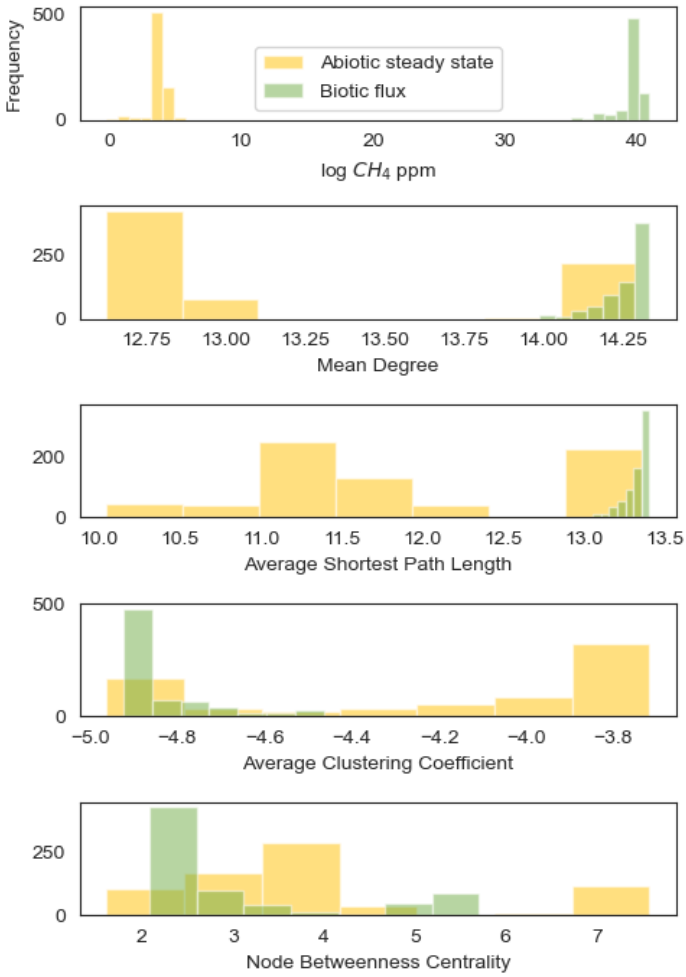


Figure C.3: Distributions of the frequency (y-axis) of the given network metric value and CH_4 abundance in log ppm (x-axis). Yellow represents models where methane is modeled as an abiotic steady-state component of the planet's atmosphere, while green represents the scenario with a biotic surface flux, with the methanogenesis pathway explicitly included in the CRN.

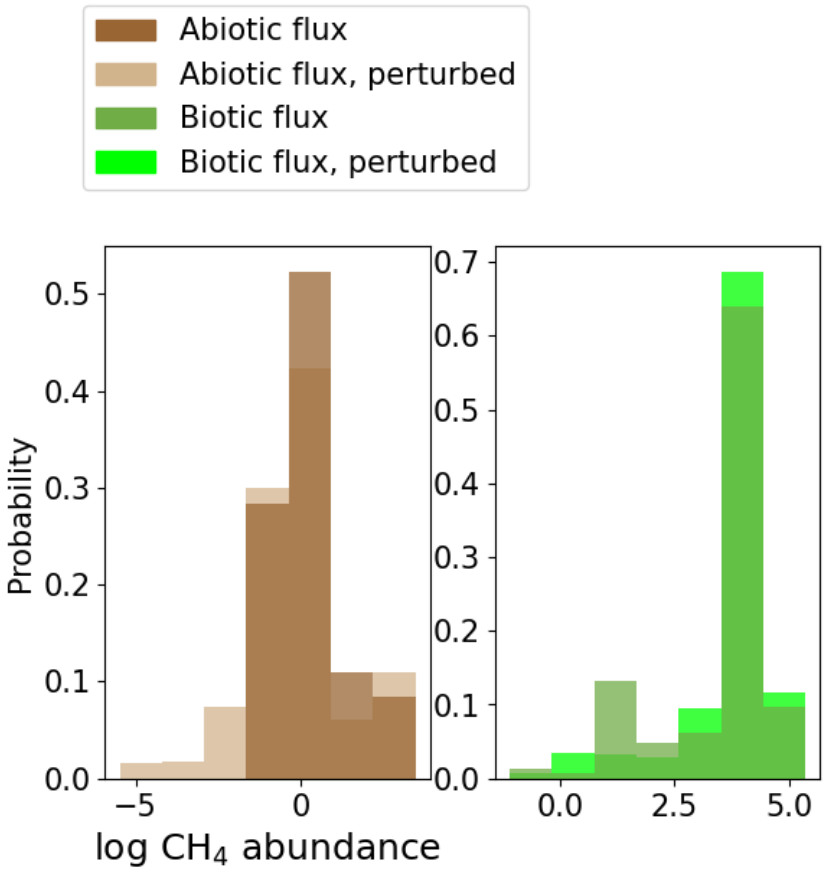


Figure C.4: The distributions of log CH_4 abundance of the perturbed and non-perturbed abiotic flux populations (left) and biotic flux populations (right).

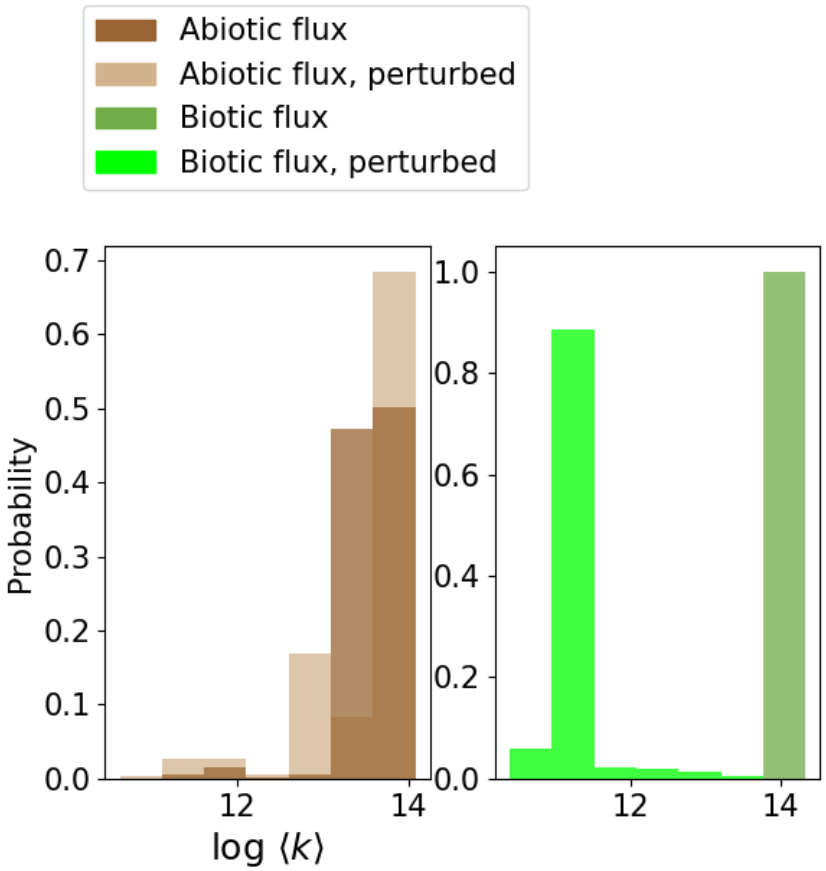


Figure C.5: The distributions of log mean degree of the perturbed and non-perturbed abiotic flux populations (left) and biotic flux populations (right).

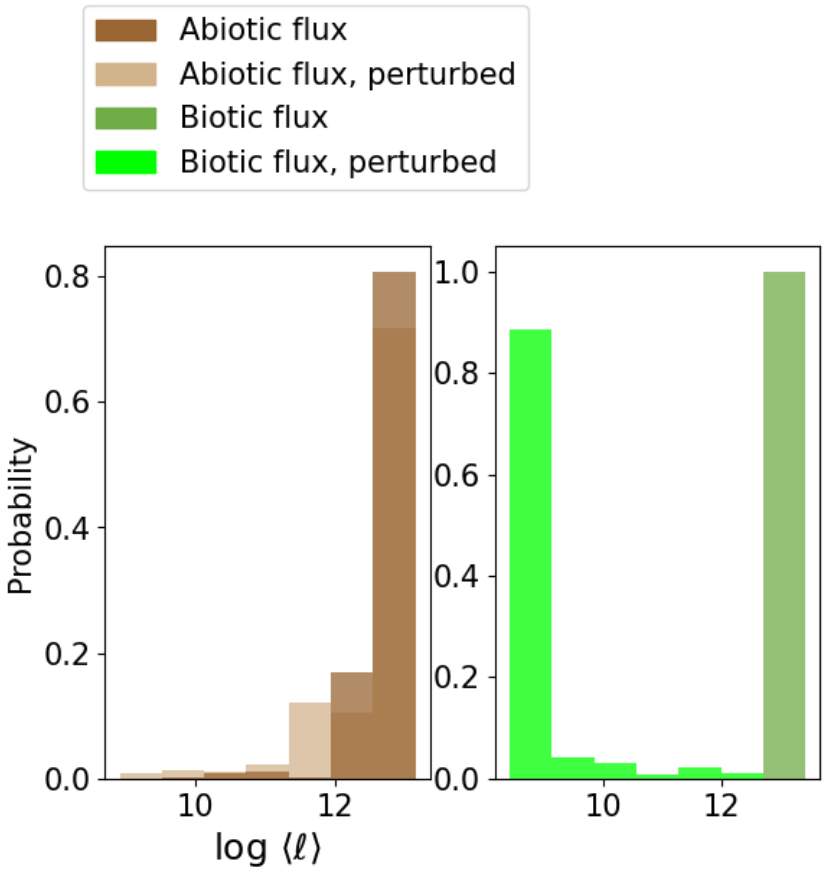


Figure C.6: The distributions of log average shortest path length of the perturbed and non-perturbed abiotic flux populations (left) and biotic flux populations (right).

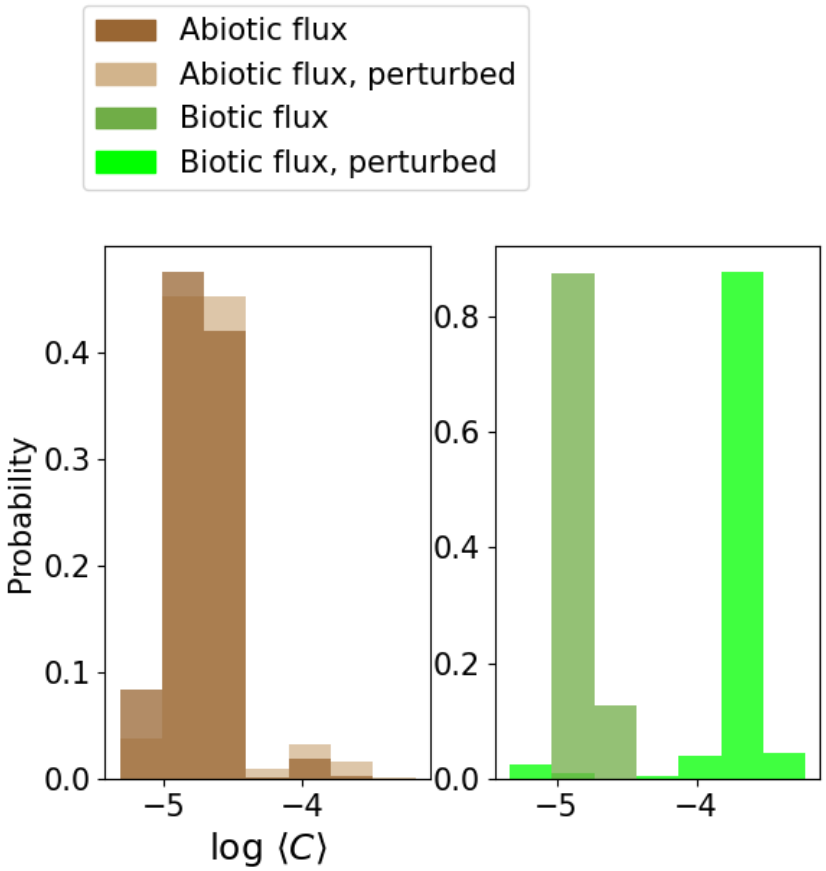


Figure C.7: The distributions of log average clustering coefficient of the perturbed and non-perturbed abiotic flux populations (left) and biotic flux populations (right).

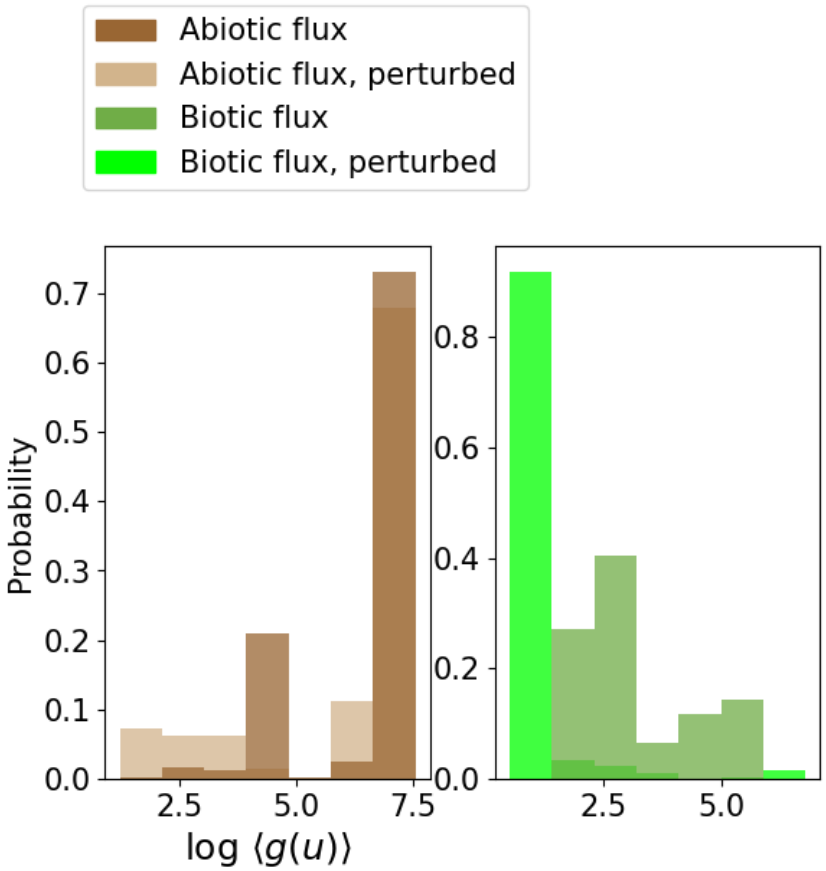


Figure C.8: The distributions of log node betweenness centrality of the perturbed and non-perturbed abiotic flux populations (left) and biotic flux populations (right).

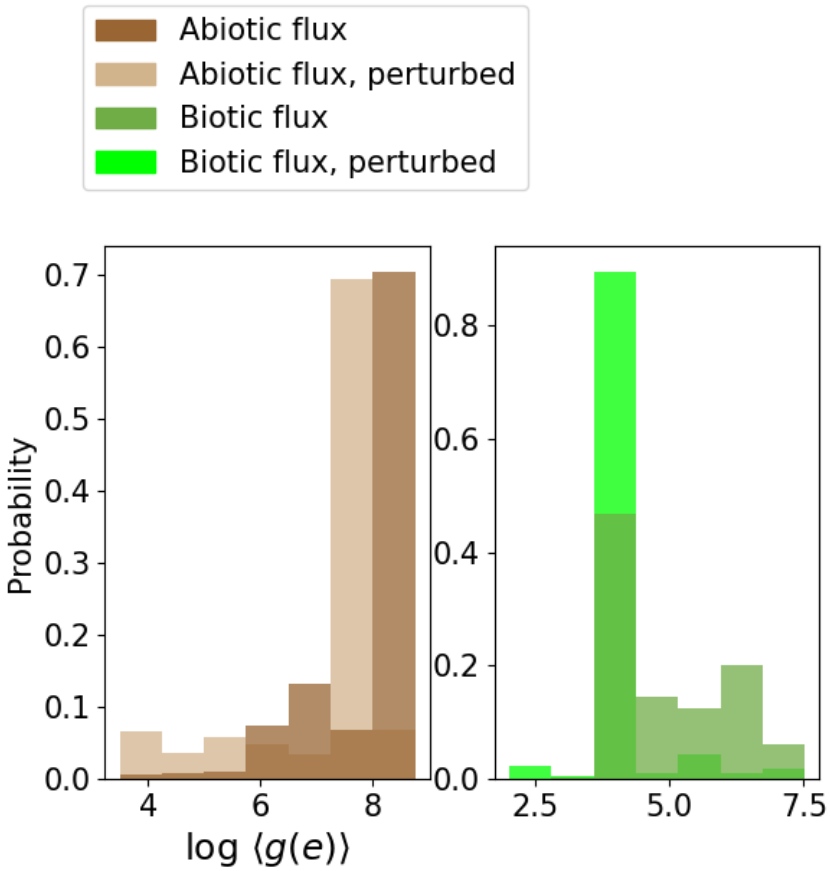


Figure C.9: The distributions of log edge betweenness centrality of the perturbed and non-perturbed abiotic flux populations (left) and biotic flux populations (right).

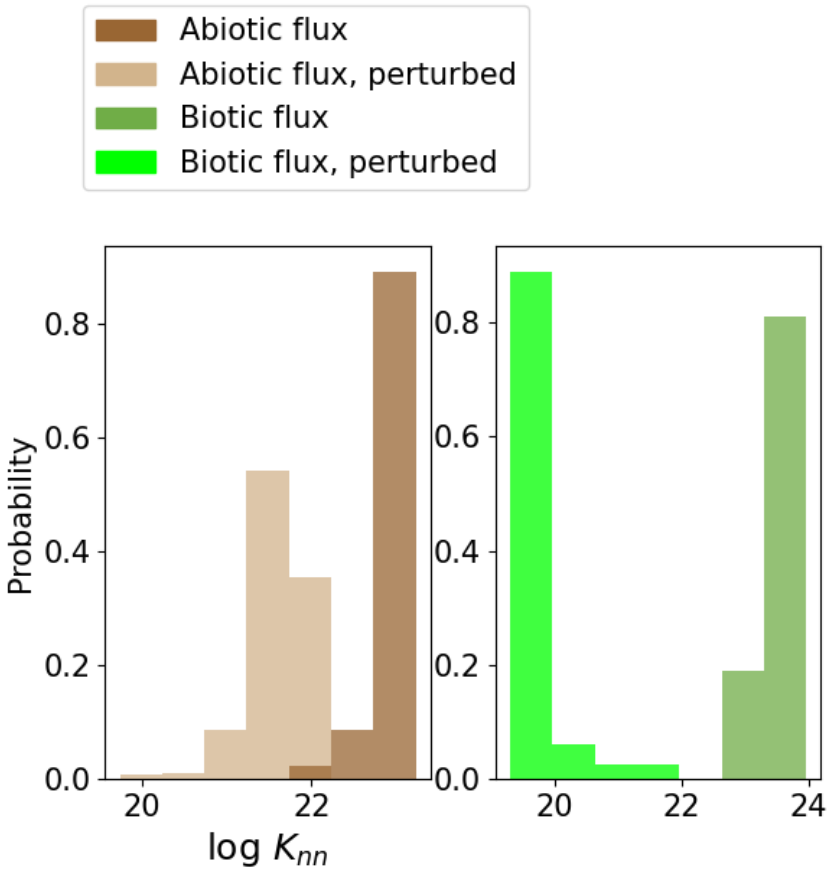


Figure C.10: The distributions of log average neighbor degree of the perturbed and non-perturbed abiotic flux populations (left) and biotic flux populations (right).