

# Geophysical Research Letters

## RESEARCH LETTER

10.1029/2020GL090481

### Key Points:

- We investigate three atmospheric chemical mechanisms of varying complexity and process representation using graph-theoretical techniques
- Results are consistent with existing computational approaches, demonstrating key similarities and differences between the mechanisms
- These techniques provide chemical state independent, structural insights into existing mechanisms, and methods for new mechanism development

### Supporting Information:

- Supplementary Material S1

### Correspondence to:

S. J. Silva,  
samsilva@pnnl.gov

### Citation:

Silva, S. J., Burrows, S. M., Evans, M. J., & Halappanavar, M. (2021). A graph theoretical intercomparison of atmospheric chemical mechanisms. *Geophysical Research Letters*, 48, e2020GL090481. <https://doi.org/10.1029/2020GL090481>

Received 21 AUG 2020

Accepted 29 NOV 2020

© 2020. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

## A Graph Theoretical Intercomparison of Atmospheric Chemical Mechanisms

Sam J. Silva<sup>1</sup> , Susannah M. Burrows<sup>1</sup> , Mathew J. Evans<sup>2,3</sup> , and Mahantesh Halappanavar<sup>4</sup>

<sup>1</sup>Earth and Biological Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA, USA, <sup>2</sup>Wolfson Atmospheric Chemistry Laboratories, Department of Chemistry, University of York, York, UK, <sup>3</sup>National Centre for Atmospheric Science, University of York, York, UK, <sup>4</sup>Physical and Computational Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA, USA

**Abstract** Graph-theoretical methods have revolutionized the exploration of complex systems across scientific disciplines. Here, we demonstrate their applicability to the investigation and comparison of three widely used atmospheric chemical mechanisms of varying complexity: the Master Chemical Mechanism v3.3, GEOS-Chem v12.6, and the Super-Fast chemical mechanism. We investigate these mechanisms using a class of graphical models known as species-reaction graphs and find similarities between these chemical reaction systems and other systems arising in nature. Several graph theoretical properties are consistent across mechanisms, including strong dynamical system disequilibrium and clustering of chemically related species. This formalism also reveals key differences between the mechanisms, some of which have characteristics inconsistent with domain knowledge; e.g., isoprene and peroxy radical chemistry exhibit substantially different graph properties in each mechanism. Graph-theoretical methods provide a promising set of tools for investigating atmospheric chemical mechanisms, complementing existing computational approaches, and potentially opening new avenues for scientific discovery.

**Plain Language Summary** A type of math known as “graph theory” provides a set of tools for scientific discovery in a variety of contexts. These include the study of social networks, the power grid, and the human brain. In this paper, we apply these tools to the study of the chemical reactions that occur in the atmosphere. We investigate and intercompare several different descriptions of these reactions, and find that these new graph theory-based techniques provide information consistent with existing scientific techniques, and potentially allow for new avenues for scientific discovery.

### 1. Introduction

The system of chemical reactions in the atmosphere drives critically important processes relevant to climate change, air pollution, and global biogeochemical cycles. Chemical reactions in the atmosphere influence the formation and lifetime of many important greenhouse gases, serving as the dominant production and loss mechanism for many of the strongest radiative forcing agents (e.g., ozone and methane, IPCC, 2013). The abundances of the two most deadly atmospheric pollutants in terms of their impact on global human health, particulate matter and tropospheric ozone, are directly controlled by atmospheric chemical reactions (Cohen et al., 2017; Seinfeld & Pandis, 2016). Biogeochemical cycles are influenced through the fluxes of chemical species in and out of terrestrial and marine ecosystems, all of which is modulated by atmospheric chemistry (Committee on the Future of Atmospheric Chemistry Research, 2016).

These reactions are highly nonlinear in nature, and occur over a wide range of time scales and chemical regimes (Logan et al., 1981). Theoretical understanding of this chemical system thus necessitates the development of detailed representations of atmospheric chemical reactions. These representations, known as atmospheric chemical mechanisms, are used to interpret observations of chemical abundances and predict the chemical behavior of the atmosphere (Brasseur & Jacob, 2017).

In practice, chemical mechanisms are designed with varying levels of detail and process representation, depending on their intended application and the associated requisite design features (e.g., Bates & Jacob, 2019; Brown-Steiner et al., 2018; Chapman, 1930; Jenkin et al., 1997). Mechanisms can vary in size from fewer

than 25 species and reactions to >10,000. These larger mechanisms frequently include contributions from teams of researchers for a variety of purposes, and do not always follow consistent design principles. As the complexity of these mechanisms grows, understanding the behavior of these systems through manual inspection of the chemical reactions becomes intractable. Thus, the emergent properties and characteristics of the chemical system are typically investigated using computational approaches, e.g., where the system of differential equations for chemical abundances represented by the mechanism is solved for a given set of initial conditions (Brasseur & Jacob, 2017).

In this work, we present an alternative approach to investigating complex atmospheric chemical mechanisms. We show how discrete mathematical techniques from the field of graph theory can provide insights into the inherent structural properties of atmospheric chemical mechanisms that are independent of any specific chemical state, can be used to evaluate the consistency of a mechanism with existing domain knowledge, and can potentially provide new avenues for insight into atmospheric chemical systems. We begin by introducing chemical graph theory in the context of atmospheric chemistry, followed by several examples demonstrating how these graph-theoretical methods can allow for the quantitative exploration and intercomparison of gas-phase atmospheric chemical mechanisms across scales.

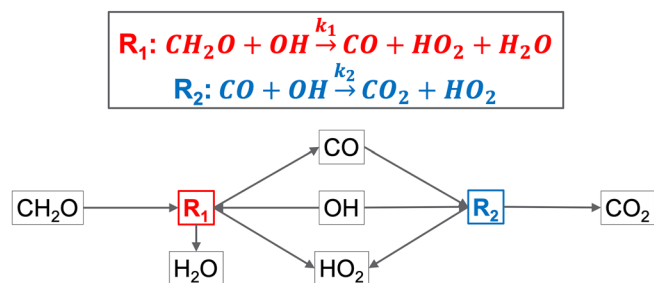
## 2. Chemical Graph Theory

Graph theory provides a set of tools for understanding the relationships within any type of complex system, such as a chemical mechanism, that can be represented as a “graph.” In general terms, a graph,  $G = (V, E)$ , is a defined mathematical construct that is composed of a set of vertices,  $V$ , interconnected by pairwise relationships expressed by a set of edges,  $E$ . Edges connect vertices whose interaction we are interested in understanding. Mathematical operations can be applied to graphs to provide insight into complex system behaviors (Estrada, 2016). As a set of data analytical tools, graph-theoretical methods are widely used in a variety of domains, including social network analysis, neuroscience, genetics, and ecology (Estrada, 2016). For a more detailed introduction to graph theory, see Wilson (2010) and Estrada (2016).

Graph-theoretical methods have already been applied in a variety of chemical applications including the investigation of reaction mechanisms (e.g., Lu & Law, 2005, 2006), and particularly for organic chemical synthesis (Bajczyk et al., 2018; Fuller et al., 2012; Grzybowski et al., 2009). For example, Grzybowski et al. (2009) compiled a list of all known reactions in the synthesis of organic molecules into a graph, and illustrated the vast complexity of this chemical system and the potential utility of graph-theoretical methods for chemical synthesis. In an atmospheric context, Dobrijevic et al. (1995, 2010) used similar methods in studies aimed at exploring the important production pathways of organic species in the atmospheres of Neptune and Titan.

Chemical mechanisms can be represented by a specific type of graph known as a species-reaction graph (Feinberg, 2019; Sakamoto et al., 1988). In a species-reaction graph, each chemical species and each reaction equation is represented as a vertex in the graph, and edges connect each reaction with its chemical reactants and products. The edges are directional, with edges pointing from reactants to each of the reaction equations they participate in, and edges pointing from the reaction equations to each of their products. Species-reaction graphs belong to a class of graphs termed a “directed bipartite network,” where vertices belong to two distinct sets (i.e., chemical species, and reaction equations), and where every edge connects a vertex from one set (a chemical species) to a vertex from the other set (a reaction). The species-reaction graph is distinct from simply connecting reactants to products, as has been done in previous work (Grzybowski et al., 2009). Using the bipartite representation with intermediate reaction vertices is preferable, as it directly represents the interactions between reactants, and is thus a more realistic representation of the complex system (Estrada, 2016; Feinberg, 2019).

An example of a simple species-reaction graph for a two-reaction mechanism is shown in Figure 1. The reactants  $\text{CH}_2\text{O}$  and  $\text{OH}$  are connected to their reaction vertex  $R_1$ , which is connected to its products,  $\text{H}_2\text{O}$ ,  $\text{CO}$ , and  $\text{HO}_2$ . The second reaction, the oxidation of  $\text{CO}$ :  $\text{CO} + \text{OH}$ , is encoded analogously.



**Figure 1.** A two-reaction chemical mechanism and the corresponding species-reaction graph. Each chemical species, as well as each reaction equation, is represented by a single “vertex” (shown as a box), while each reaction is represented by a directed “edge” (shown as an arrow in the diagram). Red and blue colors highlight the reaction equation vertices and the edges belonging to the corresponding chemical reaction.

### 3. Chemical Mechanisms

In this work, we present a graph theoretical analysis of three gas-phase atmospheric chemical mechanisms from models of varying complexity. All graph theoretical analyses presented here were completed using two freely available software packages: gephi 0.9.2 (Bastian et al., 2009) and igraph 1.0.0 (Csardi & Nepusz, 2006).

The simplest chemical mechanism used here is the “Super-Fast” mechanism (Cameron-Smith et al., 2006). It was designed for computational efficiency and to be used in multidecadal or multicentury global climate model simulations, with a major aim of predicting the concentration of ozone and the chemical control on the atmospheric lifetime of methane. It is currently an option for use in two major U.S. climate models: the Community Earth System Model and the Energy Exascale Earth System Model. The mechanism has 18 chemical species and 20 unique chemical reactions; for more details, see Cameron-Smith et al. (2006). Although it is a simplified representation of atmospheric chemistry, the Super-Fast

mechanism produces reasonable estimates of the concentrations of pollutants and the chemical lifetimes of greenhouse gases (Brown-Steiner et al., 2018).

As an example of intermediate complexity chemical mechanisms, we use the mechanism from the GEOS-Chem model ([www.geos-chem.org](http://www.geos-chem.org); Bey et al., 2001). This mechanism is predominantly used in an offline chemical transport model (i.e., with prescribed atmospheric transport and physics) and typically is run over time periods ranging from a few days to a few years. These differences enable chemical transport models to afford the increased computational burden required to achieve greater chemical complexity and realism. More specifically, we use the “Tropchem” mechanism from GEOS-Chem v12.6, which was designed to simulate the chemical composition of the atmosphere to high fidelity in the global troposphere, with a detailed representation of HO<sub>x</sub>-NO<sub>x</sub>-BrO<sub>x</sub>-VOC-Halogen chemistry (Mao et al., 2013; Sherwen et al., 2016; Travis et al., 2016). This mechanism is an order of magnitude more complex than the Super-Fast mechanism, with ca. 200 species and 750 reactions.

The most detailed representation of atmospheric chemistry used in this work is the “Master Chemical Mechanism,” or MCM (<http://mcm.leeds.ac.uk/>; Jenkin et al., 1997). We use the MCMv3.3, which was designed as a zero-dimensional model for atmospheric chemistry and is too computationally expensive to be routinely used in 3-D chemistry and climate models. The MCM is one of the most comprehensive existing atmospheric chemical mechanisms, containing nearly 6,000 species and 17,000 reactions. It has a state-of-the-science representation of known atmospheric chemical reactions (Bloss et al., 2005; Jenkin et al., 1997, 2012, 2015; Saunders et al., 2003), and is frequently used as a reference case when developing reduced-complexity mechanisms (e.g., Chan Miller et al., 2017).

The species-reaction graphs for each of the three mechanisms described above are visualized in Figure 2. The large range in chemical complexity is visually apparent as the MCM is far more densely connected and less sparse than the Super-Fast mechanism.

### 4. Global Graph Properties

Global properties of entire graphs can provide useful summary statistics of the overall complex system, including the presence of highly clustered vertices, dynamical instabilities, and more. We begin by describing a set of scalar quantities known as modularity and reciprocity for each chemical mechanism and linking them to known design considerations in the development of chemical mechanisms, followed by a brief description of other graph theoretically interesting properties. The graph properties in this section are calculated directly from the structure of the graphs and are independent of any particular chemical state.

**Super-Fast**

Climate Model

M = 0.37, R = 0.09,  $\gamma = -1.5$

**GEOS-Chem**

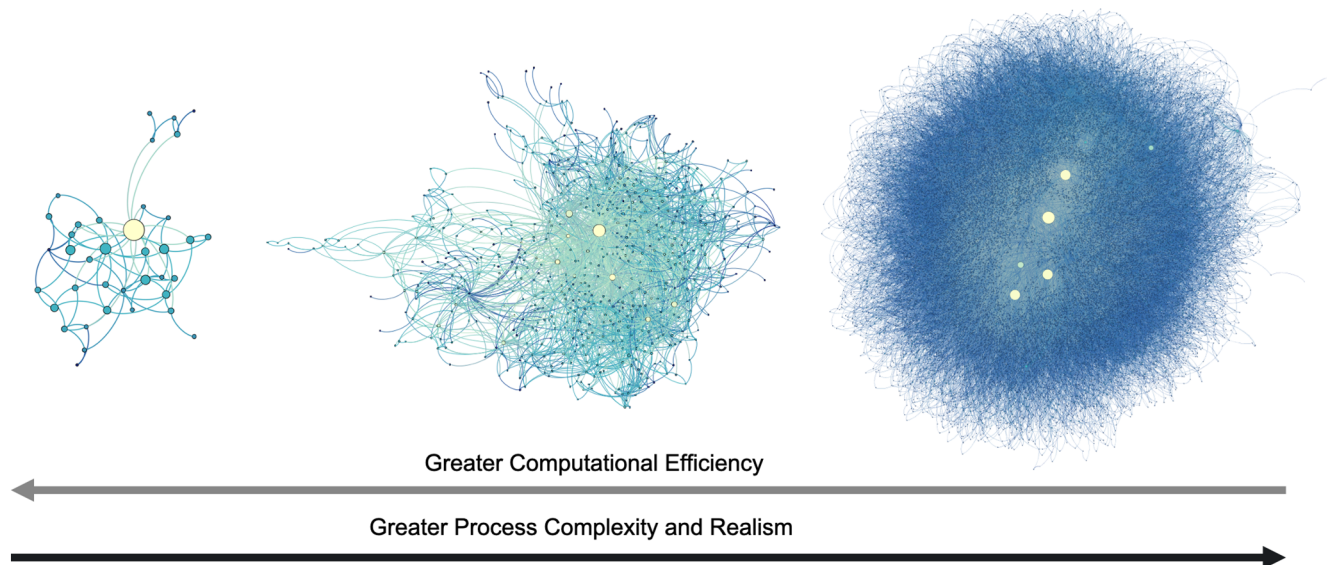
Chemical Transport Model

M = 0.42, R = 0.04,  $\gamma = -2.1$

**MCMv3.3**

Reference Box Model

M = 0.57, R = 0.02,  $\gamma = -3.2$



**Figure 2.** A visual representation of the species-reaction graphs used in this work, analogous to Figure 1. Vertex color (dark-light) and size (small-large) scales on a per-graph basis with the out-degree centrality (the number of edges leaving a node, see Section 5). Use cases and global graph properties (M = modularity, R = Reciprocity,  $\gamma$  = Gamma fitting parameter, see Section 4) are summarized below graph labels.

#### 4.1. Modularity

Modularity is a measure of clustering (or community structure) in graphs, aimed at characterizing graphs with densely connected groups or clusters of vertices (Newman, 2006). Modularity scores range from 0 to 1, with higher values indicative of a more “modular” graph. A graph with completely random connections between vertices has a modularity score of zero. For more details on the calculation of modularity see Estrada (2016) and Newman and Girvan (2004). The modularity of each chemical mechanism investigated in this work is given in Figure 2. As chemical mechanism complexity increases and more classes of reactions are considered, there is an associated increase in modularity, a result consistent with a previous analysis of simplified chemical mechanisms of Earth and other planetary atmospheres (Estrada, 2016). This indicates that the more complex representation (i.e., MCM) is both less random in its organization and contains more clusters of interrelated vertices than the simple mechanisms (Super-Fast). This reflects in part the fact that many closely interrelated chemical species are grouped together (known as “lumping”) in simpler chemical mechanisms (e.g., Emmerson & Evans, 2009). This lumping effectively replaces clusters of highly interconnected vertices or chemically similar species with a single vertex, which reduces the mechanism’s modularity.

#### 4.2. Reciprocity

Reciprocity is a measure of what fraction of the edges from one vertex to another point in both directions. In the context of directed graphs of complex systems such as the species-reaction graph, low reciprocity typically indicates that the system is highly unstable (Estrada, 2016). Chemically, reciprocity represents the fraction of single step reactions in a given mechanism where at least one product is also a reactant. In atmospheric chemical mechanisms, reciprocal reactions exist for two main reasons: they either are reactions that require the presence of chemically inert third bodies (e.g., H<sub>2</sub>O or N<sub>2</sub>), or are approximations of fast chemistry in reduced-complexity mechanisms (e.g., rapid catalytic reaction cycles or radical recycling approximated by a single reaction step). Reciprocity values shown in Figure 2 indicate that while all three graphs have low reciprocity, reciprocity decreases with graph complexity consistent with a simplified representation of fast chemical processes that are more directly resolved in the detailed mechanisms (e.g., OH recycling and isoprene chemistry).

### 4.3. Other Properties of Interest

There are other properties of chemical mechanisms that are interesting from a graph theoretical perspective, and aid in relating these graphs to wider classes of well-studied graphs. These include the approximate power-law scaling of the degree distributions, and the small-world classification of the graphs.

The number of edges connected to a given vertex is known as the vertex degree. The distribution of the degrees of all vertices in a graph is known as the “degree distribution.” For all three graphs studied here, much of the degree distribution (plotted for all vertices in each bipartite graph in Figure S1) roughly follows a power-law scaling, where the majority of vertices have a low degree, meaning they participate in few reactions, and a small subset of vertices has a very high degree, meaning they participate in many reactions. This decreasing power-law degree distribution (or “fat tailed” degree distribution) is commonly found throughout graphs in the natural world, including planetary chemical reaction networks (Broido & Clauset, 2019; Solé & Munteanu, 2004). To characterize these degree distributions, we fit the decreasing power-law portion of the distributions following  $D = n^\gamma$ , where  $D$  is the in-degree or out-degree,  $n$  is the number of vertices, and  $\gamma$  is the exponential free parameter. The  $\gamma$  for all three mechanisms is summarized in Figure 2. We find that as the graphs grow in complexity, their exponential constant  $\gamma$  becomes more negative. This indicates that larger graphs are more interconnected and less heterogeneous (e.g., Estrada, 2010), consistent with higher chemical process resolution.

In addition to approximate power-law degree distributions, these chemical mechanisms have properties consistent with a class of graphs known as “small-world” graphs (Watts & Strogatz, 1998), which are commonly observed in natural systems (Telesford et al., 2011). Mathematically, small-world graphs are graphs where the average path length between vertices scales proportionally to the natural log of the number of vertices in that graph (Watts & Strogatz, 1998), this relationship is visualized for the mechanisms in this work in Figure S2. Small-world graphs have vertices that largely only directly interact with a few vertices, but all other vertices in the graph are quite close (i.e., a small number of edges away). This is consistent with the highly coupled nature of atmospheric chemical reactions, where despite generally having only a few direct chemical interactions, a perturbation to a given chemical species can quickly have a cascading effect across many species in the reaction network.

## 5. Vertex Characterization

In addition to global properties, graph-theoretical methods can be used to characterize the relationships between individual vertices and the entire system. In this section, we use these methods to compare the treatment of various chemical species within the three chemical mechanisms. As with the global graph properties, these vertex characterization metrics are calculated independent of any particular chemical state. This is in contrast to other common methods of analyzing chemical mechanisms (e.g., integrating the systems of differential equations), which ultimately require some information regarding the chemical state.

One of the most commonly used vertex characterization methods is known as “vertex centrality.” Centrality is a metric that is calculated for each vertex in a graph, which quantifies its importance within the structure of the system (Estrada, 2016; Grzybowski et al., 2009). While many centrality metrics exist, we focus here on the “out-degree centrality.” Degree centrality is simply the number of edges connected to a given vertex (i.e., the degree of that vertex). For species vertices, the degree centrality corresponds to the number of times a species is represented in a reaction mechanism in any way. Out-degree centrality is the number of edges leaving a given vertex. For species vertices, the out-degree centrality is the number of times a species is represented in a reaction mechanism as a reactant. Vertices with higher out-degree centrality are commonly identified as important vertices within a graph (Estrada, 2016), and we use this metric here to quantify the important species within the structure of the chemical reaction mechanisms. We focus on out-degree centrality over the degree centrality to maintain a focus on reactively important species. Species with a high degree centrality, by contrast, include those that are present in many reactions as products but are not particularly relevant to the gas-phase chemistry (e.g., CO<sub>2</sub> in the GEOS-Chem mechanism, which is included in certain reactions to balance carbon). We acknowledge that this centrality score structural importance metric is quite simple, and likely misses some important indirect interactions in the system, particularly as

**Table 1**  
*Species Vertices Ranked by Out-Degree Centrality for Each of the Chemical Mechanisms Investigated in This Work*

Rank	Super-Fast	GEOS-Chem	MCMv3.3
1	OH	OH	OH
2	HO <sub>2</sub>	NO	HO <sub>2</sub>
3	O <sub>3</sub>	HO <sub>2</sub>	NO <sub>3</sub>
4	CH <sub>3</sub> O <sub>2</sub>	MCO <sub>3</sub>	NO
5	NO	MO <sub>2</sub>	NO <sub>2</sub>

they relate to less reactive species that are known to strongly influence the atmospheric chemical state (e.g., CH<sub>4</sub>).

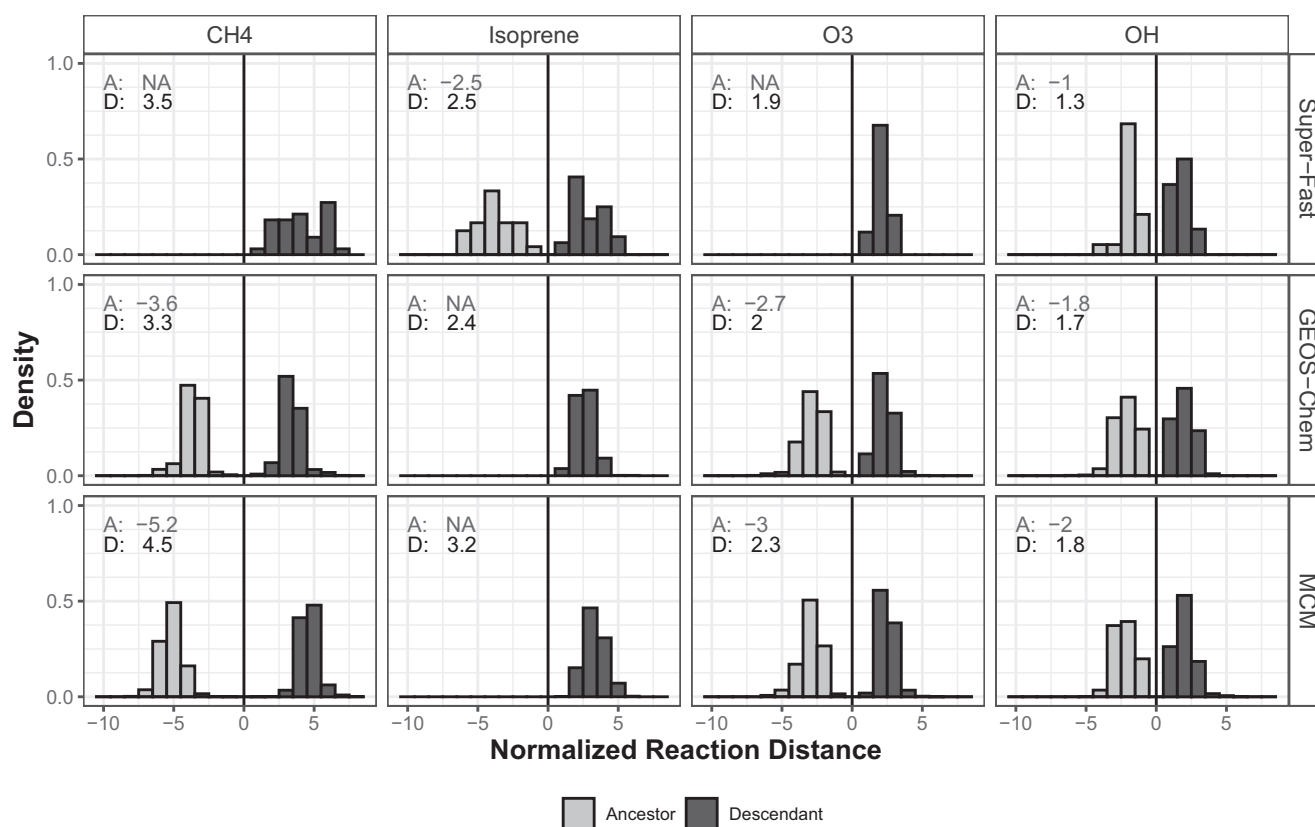
The five species with the highest out-degree centrality for each of the three chemical mechanisms are shown in Table 1. In each mechanism the chemical species OH has the highest centrality score. This structural importance is consistent with domain expertise in atmospheric chemistry, since OH is well-known to be the atmosphere's most important oxidant (Jacob, 1999; Seinfeld & Pandis, 2016). After the highest-ranked vertex of OH, the mechanisms differ in their ranking of species by out-degree centrality. All three include HO<sub>2</sub> and NO among the top five species, which reflects the importance of the HO<sub>x</sub> family (OH + HO<sub>2</sub>) and reactive nitrogen chemistry in the structure of all three models.

In the MCM, HO<sub>x</sub> and reactive nitrogen species (NO, NO<sub>2</sub>, and NO<sub>3</sub>) comprise the entirety of the five highest-ranked species. In the GEOS-Chem and Super-Fast mechanisms, peroxy radicals are among the highest-ranking species: methylperoxy radical in the Super-Fast, and methylperoxy (MO<sub>2</sub>) and peroxyacetyl radicals (MCO<sub>3</sub>) in GEOS-Chem. The MCM includes a highly detailed representation of peroxy radical chemistry, with >1,200 total peroxy radical species. GEOS-Chem includes >20 peroxy radical species, and Super-Fast includes only one (methylperoxy radical). This reduction of species increases the relative importance of peroxy radicals in the simpler schemes. If peroxy radicals were similarly grouped together as a single vertex in the graph of the MCM mechanism (i.e., as "RO<sub>2</sub>"), it would similarly appear in the top five as in the Super-Fast and GEOS-Chem. Much of the difference between the GEOS-Chem and MCM mechanisms lies in this treatment of peroxy radicals. Consistent with atmospheric chemistry domain knowledge, in the MCM all organic peroxy radicals systematically react with NO, NO<sub>2</sub>, NO<sub>3</sub>, HO<sub>2</sub>, or an organic peroxy radical. To reduce complexity, not all of these organic peroxy radical pathways are included in GEOS-Chem or Super-Fast.

These graph theoretical importance metrics can yield insights into the structure of these mechanisms independent of any particular chemical conditions. By solely characterizing structural features of the graph, these methods can be more robust to compensating errors in mechanism construction, e.g., elevated production and loss terms canceling out during model integration, which would be obscured by other computational analysis methods. It should be noted that while the characterization metrics are independent of chemical state, the mechanism structure is the result of design choices that are implicitly driven by the atmospheric chemical state. These choices are largely influenced by current understanding of the reactive chemistry of the atmosphere, and by the relative importance assigned by experts to various aspects of atmospheric chemistry in the mechanism design. For example, the higher rank of NO for GEOS-Chem is indicative of design choices to emphasize high NO<sub>x</sub> urban chemistry over low NO<sub>x</sub> remote chemistry.

These structural importance metrics ultimately have implications for the computational efficiency of a given mechanism. In the case of GEOS-Chem, the large NO out-degree centrality importance ranking means that for all chemical conditions, the mechanism calculates many reactions including NO. Even in regions or use cases where the ultimate prediction of the atmospheric chemical state is not sensitive to NO concentrations, the GEOS-Chem mechanism still represents a relatively large number of these reactions, leading to reduced computational efficiency for a given level of target mechanism performance in certain chemical regimes.

Species centrality rankings like those in Table 1 can be used to confront the process representation within the mechanisms with a priori assumptions and knowledge of atmospheric chemical reactions. This is not specifically limited to out-degree centrality. For example, in the GEOS-Chem mechanism CO<sub>2</sub> has an in-degree centrality of 31, meaning that in GEOS-Chem there are 31 reactions that produce CO<sub>2</sub>. This conflicts with domain knowledge in atmospheric chemistry where only one reaction is the dominant chemical source of CO<sub>2</sub>, namely the oxidation of CO by OH, and reflects choices made to balance carbon in the mechanism. Similar conclusions can be reached investigating the Super-Fast mechanism where the compound isoprene has an in-degree centrality greater than zero, despite that there are no major chemical sources of isoprene thought to exist in the atmosphere. Both of these inconsistencies with domain knowledge exist to



**Figure 3.** Lineage histograms for CH<sub>4</sub>, isoprene, ozone, and OH in all three chemical mechanisms. Ancestor distances are shown in gray, and descendant distances in black. Reaction distances are scaled by the natural log of the average path length in each reaction for standard comparison. Normalized average path lengths for ancestor distances “A” and descendant distances “D” are shown in each facet.

achieve a given design goal of the mechanisms, specifically carbon closure in the GEOS-Chem mechanism (Safieddine et al., 2017) and the maintenance of reactive carbon in the Super-Fast mechanism (Cameron-Smith et al., 2006).

To further investigate the differences between the three chemical mechanisms, we explore the graph lineages of several species that are well-known to be important in atmospheric chemistry. In the context of graph theory, lineage corresponds to the minimum number of edges between a given a vertex and all other vertices in the graph. Lineage distances in the backwards direction are known as ancestor distances, and those in the forward direction are known as descendant distances. The ancestor distance indicates how many reaction steps a species is from all possible precursors, and the descendant distance indicates how many reaction steps a species is from all possible products. To facilitate comparison between mechanisms of varying complexity, ancestor and lineage distances are normalized by the natural log of the average path length in each graph. This accounts for the fact that larger graphs can have larger lineage distances simply due to their increased size. These are most readily visualized through histograms, shown in Figure 3 for the following species commonly studied in atmospheric chemistry: CH<sub>4</sub>, isoprene (C<sub>5</sub>H<sub>8</sub>), ozone (O<sub>3</sub>), and OH.

The differences in these chemical mechanisms are visually apparent in Figure 3. In general, the largest outlier is the Super-Fast mechanism, with GEOS-Chem and the MCM representing similar relative lineages of the four species.

In both GEOS-Chem and the MCM, methane has the largest average ancestor distance, or number of reaction steps, reflecting the fact that the small gas-phase chemical source of methane (oxidation of large molecules such as acetaldehyde and vinylhydroperoxide) is not tightly coupled to most of the reactions that occur in the mechanisms. In Super-Fast, the chemical source of methane is not represented at all, a simplification related to the overwhelming fraction of the methane source from direct emissions. From

the ancestor distances, it is apparent that the majority of species in all three mechanisms are in some way chemical products related to methane chemistry. This is largely achieved through the influence of methane on atmospheric oxidation. The Super-Fast chemistry shows a much larger spread in the relative number of reaction steps, and as the mechanisms increase in size, the fraction of chemical species influenced by methane oxidation grows.

The lineage analysis also immediately reveals some other key ways in which the simplified Super-Fast mechanism differs from the more complex GEOS-Chem and MCM mechanisms, particularly in its treatment of the formation both isoprene and ozone. Isoprene has no chemical sources in the real atmosphere, nor in the two more complex mechanisms, and therefore has no ancestors (as seen in Figure 3). However, in the Super-Fast mechanism, isoprene has chemical precursors driven in some way by  $\sim 65\%$  of the mechanism. This nonphysical relationship was implemented in the Super-Fast mechanism to meet the design goal of maintaining the abundance of reactive carbon in the atmosphere in order to reproduce ozone concentrations and the chemical loss of methane at low computational cost (Cameron-Smith et al., 2006). For ozone, the dominant production pathway in all mechanisms is through direct photolysis of  $\text{NO}_2$ . This is an assumption made for computational simplicity, as the reaction of  $\text{O} + \text{O}_2$  is exceedingly fast in the atmosphere and the atmospheric reservoir of  $\text{O}_2$  is large. This is the only formation mechanism for ozone in the Super-Fast mechanism, and as such there are no gas-phase reaction precursors to ozone production. There are precursors in the GEOS-Chem and MCM mechanisms; however, as the chemistry is sufficiently detailed to represent other possible formation pathways (e.g.,  $\text{OH} + \text{OH}$ , etc.).

Of the four species described here, the representation of OH within all three mechanisms is the most similar from a lineage perspective. Consistent with domain knowledge in atmospheric chemistry, OH is tightly coupled to most species in the mechanism, with low average ancestor and descendant distances. As the mechanisms increase in complexity, the fraction of the mechanism related to the formation of OH increases from  $\sim 0.5$  in the Super-Fast to 0.95 in the MCM. The same conclusion holds for the species OH influences, which increases from  $\sim 0.8$  in the Super-Fast to 1.0 in the MCM.

## 6. Summary and Future Outlook

Graph-theoretical methods provide a quantitative set of tools for investigating chemical mechanisms relevant to atmospheric chemistry. We demonstrate their applicability through comparison of three chemical mechanisms that vary across a wide spectrum of complexity using a class of graphical models known as species-reaction graphs. The mechanisms are similar at a high level but have substantial differences apparent upon investigation of graph theoretical properties. These differences are structural within the mechanisms and are independent of chemical state. Interrogation of chemical mechanisms with graph-theoretical techniques can provide objective metrics for assessing the properties of the mechanisms against domain knowledge and prior beliefs. Beyond the possibilities for additional insights and perspectives in atmospheric chemistry, a key advantage of these graph-theoretical methods is their modest computational requirements: any part of the chemical analysis we presented in this paper can be completed in  $<5$  min of total compute time using only a laptop computer with freely available open source software. Ultimately, graph-theoretical techniques enable a new set of tools that can be used in conjunction with traditional computational approaches for the investigation and interpretation of atmospheric chemical mechanisms.

In our future work, we aim to generalize the methods proposed in this paper across a variety of atmospheric conditions, while enhancing an additional set of promising graph-theoretical tools relevant to atmospheric chemistry. We plan to directly incorporate chemical information about reaction rates and net chemical fluxes into the analysis presented in this paper, including the impacts of processes external to the gas-phase kinetic mechanism such as photolysis, heterogeneous chemistry, and chemical sources and sinks. A particularly interesting direction for future work is the application of graph clustering (e.g., Barber, 2007; Newman & Girvan, 2004) to atmospheric chemistry. Graph clustering on species-reaction graphs may enable the development of reduced-complexity chemical mechanisms, with new reproducible and quantitative methods of model simplification, building on previous literature from other fields of chemistry (Feinberg, 2019; Lu & Law, 2006; Sander et al., 2019). Additional future work will explore how graph-theoretical techniques can be applied to chemical mechanisms to rapidly diagnose errors in chemical mechanism construction



(e.g., missing or incorrect edges or reactions, Chapelle et al., 2010), the model's sensitivity to process uncertainties (e.g., reaction rates, fluxes, etc., Weir et al., 2017), and the influence of new reactions on the complex system (e.g., through influence maximization techniques, Minutoli et al., 2019). As the modern understanding of the chemistry of the atmosphere grows more complex, there is an increasingly pressing need for novel approaches to gain insight into, and simplify, this complexity. Graph-theoretical techniques thus represent a valuable complement to existing tools in computational atmospheric chemistry research.

## Data Availability Statement

All the data used in this work, specifically the mechanism graphs files, are available at 10.5281/zenodo.3970944. The gephi software is available at <https://gephi.org/users/download/> and the igraph software is available at <https://igraph.org/r/>.

## Acknowledgments

A portion of the research described in this manuscript was conducted under the Laboratory Directed Research and Development Program at Pacific Northwest National Laboratory (PNNL), a multiprogram national laboratory operated by Battelle for the U.S. Department of Energy. S. J. Silva is grateful for the support of the Linus Pauling Distinguished Postdoctoral Fellowship program. Contributions from M. Halappanavar were supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration. We thank the editor and two anonymous reviewers for their helpful comments.

## References

- Bajczyk, M. D., Dittwald, P., Wolos, A., Szymkuć, S., & Grzybowski, B. A. (2018). Discovery and enumeration of organic-chemical and biomimetic reaction cycles within the network of chemistry. *Angewandte Chemie International Edition*, 57(9), 2367–2371. <https://doi.org/10.1002/anie.201712052>
- Barber, M. J. (2007). Modularity and community detection in bipartite networks. *Physical Review E*, 76(6), 066102. <https://doi.org/10.1103/PhysRevE.76.066102>
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media*, (Vol. 2). <https://gephi.org/users/publications/>
- Bates, K. H., & Jacob, D. J. (2019). A new model mechanism for atmospheric oxidation of isoprene: Global effects on oxidants, nitrogen oxides, organic products, and secondary organic aerosol. *Atmospheric Chemistry and Physics*, 19(14), 9613–9640. <https://doi.org/10.5194/acp-19-9613-2019>
- Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., et al. (2001). Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. *Journal of Geophysical Research*, 106(D19), 23073–23095. <https://doi.org/10.1029/2001JD000807>
- Bloss, C., Wagner, V., Jenkin, M. E., Volkamer, R., Bloss, W. J., Lee, J. D., et al. (2005). Development of a detailed chemical mechanism (MCMv3.1) for the atmospheric oxidation of aromatic hydrocarbons. *Atmospheric Chemistry and Physics*, 5(3), 641–664. <https://doi.org/10.5194/acp-5-641-2005>
- Brasseur, G. P., & Jacob, D. J. (2017). *Modeling of atmospheric chemistry*. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/9781316544754>
- Broido, A. D., & Clauset, A. (2019). Scale-free networks are rare. *Nature Communications*, 10(1), 1017. <https://doi.org/10.1038/s41467-019-08746-5>
- Brown-Steiner, B., Selin, N. E., Prinn, R., Tilmes, S., Emmons, L., Lamarque, J.-F., & (2018). Evaluating simplified chemical mechanisms within present-day simulations of the community Earth system model version 1.2 with CAM4 (CESM1.2 CAM-chem): MOZART-4 vs. reduced hydrocarbon vs. Super-Fast chemistry. *Geoscientific Model Development*, 11(10), 4155–4174. <https://doi.org/10.5194/gmd-11-4155-2018>
- Cameron-Smith, P., Lamarque, J.-F., Connell, P., Chuang, C., & Vitt, F. (2006). Toward an Earth system model: Atmospheric chemistry, coupling, and petascale computing. *Journal of Physics: Conference Series*, 46, 343–350. <https://doi.org/10.1088/1742-6596/46/1/048>
- Chapelle, O., Schkopf, B., & Zien, A. (2010). *Semi-supervised learning*, 1st ed., The MIT Press.
- Chan Miller, C., Jacob, D. J., Marais, E. A., Yu, K., Travis, K. R., Kim, P. S., et al. (2017). Glyoxal yield from isoprene oxidation and relation to formaldehyde: Chemical mechanism, constraints from SENEX aircraft observations, and interpretation of OMI satellite data. *Atmospheric Chemistry and Physics*, 17(14), 8725–8738. <https://doi.org/10.5194/acp-17-8725-2017>
- Chapman, S. (1930). A theory of upper atmospheric ozone. *Memories of Royal Meteorological Society*, 3, 103–125.
- Cohen, A. J., Brauer, M., Burnett, R., Anderson, H. R., Frostad, J., Estep, K., et al. (2017). Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: An analysis of data from the global burden of diseases study 2015. *The Lancet*, 389(10082), 1907–1918. [https://doi.org/10.1016/S0140-6736\(17\)30505-6](https://doi.org/10.1016/S0140-6736(17)30505-6)
- Committee on the Future of Atmospheric Chemistry Research, Board on Atmospheric Sciences and Climate, Division on Earth and Life Studies, & National Academies of Sciences, Engineering, and Medicine. (2016). *The future of atmospheric chemistry research: Remembering yesterday, understanding today, anticipating tomorrow*. Washington, DC: National Academies Press. <https://doi.org/10.17226/23573>
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695(5), 1–9.
- Dobrijevic, M., Hébrard, E., Plessis, S., Carrasco, N., Pernot, P., & Bruno-Clayes, M. (2010). Comparison of methods for the determination of key reactions in chemical systems: Application to Titan's atmosphere. *Advances in Space Research*, 45(1), 77–91. <https://doi.org/10.1016/j.asr.2009.06.005>
- Dobrijevic, M., Parisot, J. P., & Dutour, I. (1995). A study of chemical systems using signal flow graph theory: Application to Neptune. *Planetary and Space Science*, 43(1), 15–24. [https://doi.org/10.1016/0032-0633\(94\)00147-J](https://doi.org/10.1016/0032-0633(94)00147-J)
- Emmerson, K. M., & Evans, M. J. (2009). Comparison of tropospheric gas-phase chemistry schemes for use within global models. *Atmospheric Chemistry and Physics*, 9(5), 1831–1845. <https://doi.org/10.5194/acp-9-1831-2009>
- Estrada, E. (2010). Quantifying network heterogeneity. *Physical Review E*, 82(6), 066102. <https://doi.org/10.1103/PhysRevE.82.066102>
- Estrada, E. (2016). *The structure of complex networks: Theory and applications*. Oxford, UK: Oxford University Press.
- Feinberg, M. (2019). The species-reaction graph. In M. Feinberg (Ed.), *Foundations of chemical reaction network theory* (Vol. 202, pp. 205–240). Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-03858-8\\_11](https://doi.org/10.1007/978-3-030-03858-8_11)
- Fuller, P. E., Gothard, C. M., Gothard, N. A., Weckiewicz, A., & Grzybowski, B. A. (2012). Chemical network algorithms for the risk assessment and management of chemical threats. *Angewandte Chemie International Edition*, 51(32), 7933–7937. <https://doi.org/10.1002/anie.201202210>

- Grzybowski, B. A., Bishop, K. J. M., Kowalczyk, B., & Wilmer, C. E. (2009). The “wired” universe of organic chemistry. *Nature Chemistry*, 1(1), 31–36. <https://doi.org/10.1038/nchem.136>
- IPCC. (2013). *Climate change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change*. Cambridge, UK and New York, NY: Cambridge University Press. <https://doi.org/10.1017/CBO9781107415324>
- Jacob, D. J. (1999). *Introduction to atmospheric chemistry*. Princeton, NJ: Princeton University Press.
- Jenkin, M. E., Saunders, S. M., & Pilling, M. J. (1997). The tropospheric degradation of volatile organic compounds: A protocol for mechanism development. *Atmospheric Environment*, 31(1), 81–104. [https://doi.org/10.1016/S1352-2310\(96\)00105-7](https://doi.org/10.1016/S1352-2310(96)00105-7)
- Jenkin, M. E., Wyche, K. P., Evans, C. J., Carr, T., Monks, P. S., Alfarra, M. R., et al. (2012). Development and chamber evaluation of the MCM v3.2 degradation scheme for  $\beta$ -caryophyllene. *Atmospheric Chemistry and Physics*, 12(11), 5275–5308. <https://doi.org/10.5194/acp-12-5275-2012>
- Jenkin, M. E., Young, J. C., & Rickard, A. R. (2015). The MCM v3.3.1 degradation scheme for isoprene. *Atmospheric Chemistry and Physics*, 15(20), 11433–11459. <https://doi.org/10.5194/acp-15-11433-2015>
- Logan, J. A., Prather, M. J., Wofsy, S. C., & McElroy, M. B. (1981). Tropospheric chemistry: A global perspective. *Journal of Geophysical Research*, 86(C8), 7210–7254. <https://doi.org/10.1029/JC086iC08p07210>
- Lu, T., & Law, C. K. (2005). A directed relation graph method for mechanism reduction. *Proceedings of the Combustion Institute*, 30(1), 1333–1341. <https://doi.org/10.1016/j.proci.2004.08.145>
- Lu, T., & Law, C. K. (2006). On the applicability of directed relation graphs to the reduction of reaction mechanisms. *Combustion and Flame*, 146(3), 472–483. <https://doi.org/10.1016/j.combustflame.2006.04.017>
- Mao, J., Paulot, F., Jacob, D. J., Cohen, R. C., Crounse, J. D., Wennberg, P. O., et al. (2013). Ozone and organic nitrates over the eastern United States: Sensitivity to isoprene chemistry. *Journal of Geophysical Research: Atmospheres*, 118, 11256–11268. <https://doi.org/10.1002/jgrd.50817>
- Minutoli, M., Halappanavar, M., Kalyanaraman, A., Sathanur, A., McClure, R., & McDermott, J. (2019). *Fast and Scalable Implementations of Influence Maximization Algorithms*. In 2019 IEEE International Conference on Cluster Computing (CLUSTER), (1–12). <https://doi.org/10.1109/CLUSTER.2019.8890991>
- Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23), 8577–8582. <https://doi.org/10.1073/pnas.0601602103>
- Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2), 026113. <https://doi.org/10.1103/PhysRevE.69.026113>
- Safieddine, S. A., Heald, C. L., & Henderson, B. H. (2017). The global non-methane reactive organic carbon budget: A modeling perspective. *Geophysical Research Letters*, 44(8), 3897–3906. <https://doi.org/10.1002/2017GL072602>
- Sakamoto, A., Kawakami, H., & Yoshikawa, K. (1988). A graph theoretical approach to complex reaction networks. *Chemical Physics Letters*, 146(5), 444–448. [https://doi.org/10.1016/0009-2614\(88\)87475-X](https://doi.org/10.1016/0009-2614(88)87475-X)
- Sander, R., Baumgaertner, A., Cabrera-Perez, D., Frank, F., Gromov, S., Groß, J.-U., et al. (2019). The community atmospheric chemistry box model CAABA/MECCA-4.0. *Geoscientific Model Development*, 12(4), 1365–1385. <https://doi.org/10.5194/gmd-12-1365-2019>
- Saunders, S. M., Jenkin, M. E., Derwent, R. G., & Pilling, M. J. (2003). Protocol for the development of the master chemical mechanism, MCM v3 (Part A): Tropospheric degradation of non-aromatic volatile organic compounds. *Atmospheric Chemistry and Physics*, 3(1), 161–180. <https://doi.org/10.5194/acp-3-161-2003>
- Seinfeld, J. H., & Pandis, S. N. (2016). *Atmospheric chemistry and physics: From air pollution to climate change* (3rd ed.). Hoboken, NJ: Wiley.
- Sherwen, T., Schmidt, J. A., Evans, M. J., Carpenter, L. J., Großmann, K., Eastham, S. D., et al. (2016). Global impacts of tropospheric halogens (Cl, Br, I) on oxidants and composition in GEOS-Chem. *Atmospheric Chemistry and Physics*, 16(18), 12239–12271. <https://doi.org/10.5194/acp-16-12239-2016>
- Solé, R. V., & Munteanu, A. (2004). The large-scale organization of chemical reaction networks in astrophysics. *Europhysics Letters*, 68(2), 170–176. <https://doi.org/10.1209/epl/i2004-10241-3>
- Telesford, Q. K., Joyce, K. E., Hayasaka, S., Burdette, J. H., & Laurienti, P. J. (2011). The ubiquity of small-world networks. *Brain Connectivity*, 1(5), 367–375. <https://doi.org/10.1089/brain.2011.0038>
- Travis, K. R., Jacob, D. J., Fisher, J. A., Kim, P. S., Marais, E. A., Zhu, L., et al. (2016). Why do models overestimate surface ozone in the Southeast United States? *Atmospheric Chemistry and Physics*, 16(21), 13561–13577. <https://doi.org/10.5194/acp-16-13561-2016>
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684), 440–442. <https://doi.org/10.1038/30918>
- Weir, W. H., Emmons, S., Gibson, R., Taylor, D., & Mucha, P. J. (2017). Post-processing partitions to identify domains of modularity optimization. *Algorithms*, 10(3), 93. <https://doi.org/10.3390/a10030093>
- Wilson, R. J. (2010). *Introduction to graph theory* (5th ed.). Harlow: Pearson/Prentice Hall.