# Reaction Networks Resemble Low-Dimensional Regular Lattices

Miko M. Stulajter[*,†,‡] and Dmitrij Rappoport[*,†]

†*Department of Chemistry, University of California Irvine, Irvine, CA 92697, USA*

‡*Computational Science Research Center, San Diego State University, San Diego, CA 92182, USA*

E-mail: mstulajt@uci.edu; dmitrij@rappoport.org

**Abstract**

The computational exploration, manipulation, and design of complex chemical reactions face fundamental challenges related to the high-dimensional nature of potential energy surfaces (PESs) that govern reactivity. Accurately modeling complex reactions is crucial for understanding the chemical processes involved in, for example, organocatalysis, autocatalytic cycles, and one-pot molecular assembly. Our prior research demonstrated that discretizing PESs using heuristics based on bond breaking and bond formation produces a reaction network representation with a low-dimensional structure (metric space). We now find that these stoichiometry-preserving reaction networks possess additional, though approximate, structure and resemble low-dimensional regular lattices with a small amount of random edge rewiring. The heuristics-based discretization thus generates a nonlinear dimensionality reduction by a factor of ten with an a posteriori error measure (probability of random rewiring). The structure becomes evident through a comparative analysis of CHNO reaction networks of varying stoichiometries against a panel of size-matched generative network models, taking into account their local, metric, and global properties. The generative models include

random networks (Erdős–Rényi and bipartite random networks), regular lattices (periodic and non-periodic), and network models with a tunable level of "randomness" (Watts–Strogatz graphs and regular lattices with random rewiring). The CHNO networks are simultaneously closely matched in all these properties by 3–4-dimensional regular lattices with 10% or less of edges randomly rewired. The effective dimensionality reduction is found to be independent of the system size, stoichiometry, and rule set, suggesting that search and sampling algorithms for PESs of complex chemical reactions can be effectively leveraged.

# 1 Introduction

Reaction networks (RNs) offer a discrete and relatively compact model of chemical reactivity in metabolism,[1–4] combustion,[5,6] protein folding,[7,8] and other types of complex chemical dynamics.[9–11] Network models with different definitions of their nodes and edges are used depending on the application. In kinetic modeling, it is convenient to assign network nodes to the individual compounds and connect all reactants and products of reactions by network edges, resulting in highly heterogeneous networks.[1,12] However, chemical dynamics and, in particular, potential energy surface (PES) exploration, is more naturally represented by a RN, in which network nodes correspond to the states of the chemical system and network edges are transitions between the system states.[13,14] As we showed previously, these stoichiometry-preserving transition networks, when equipped with shortest-path (chemical) distance as the distance metric, form discrete and low-dimensional metric spaces.[14] In comparison to the underlying PESs, metric-space models have about ten times smaller dimensionality, as measured by their fractal dimension.[15–17]

Even though lower dimensionality boosts search algorithm performance,[18,19] metric spaces tend to be more challenging data structures than Euclidean spaces like $\mathbb{R}^d$. The main drawback of metric spaces compared to Euclidean spaces is that they do not support a coordinate representation and cannot effectively locate their elements, having only a concept of a dis-

tance between pairs of elements. The low dimensionality of RNs compared to the underlying PESs thus comes with a caveat: the loss of the coordinate representation. However, in this work, we show that it is possible to lift the caveat if we allow for a certain (small) amount of error in the form of random edge rewiring. Specifically, we find that RNs can be closely matched by low-dimensional regular lattices with a small fraction of randomly reassigned (rewired) edges. Therefore, the coordinate representation can be restored by embedding RNs in Euclidean spaces $\mathbb{R}^d$, generating a nonlinear dimensionality reduction between the high-dimensional PES $\mathbb{R}^{3N-6}$ (where $N$ is the number of atoms) and the low-dimensional embedding space $\mathbb{R}^d$ with controllable error. The dimensionality reduction is about tenfold and holds for a range of RN sizes, stoichiometries, and rule sets. The RN embeddings can be leveraged by efficient search and sampling algorithms, taking advantage of their low-dimensional coordinate representation.

The additional (approximate) structure of RNs is a feature of the discretization of reaction paths into sequences of elementary transformations. In many instances, the individual bond breaks and bond formations can be composed in different orders while summing up to the same total reaction.[14,20] The permutations of the elementary transformations create (hyper)cubes within a RN, which correspond to fragments of regular lattices. The dimensionality of the hypercubes is related to the number of permuting RN edges and, thus, to the number of bonds involved in the reaction, which is typically much smaller than the total number of molecular degrees of freedom. The permutation symmetry is not exact since some edges are prohibited by valence rules and are further restricted if thermodynamic and kinetic feasibility are taken into account, for example, using edge weights.[14] However, the deviation from permutation symmetry can be modeled by random edge rewiring, which moreover provides a quantitative measure of the deviation from the symmetry through the rewiring probability $p_\mathrm{r}$. In this work, we explore the structures of RNs by comparing them to a panel of size-matched generated network models ranging from regular lattices to random networks. The generative models are useful surrogates for the real networks as they are cheap

to construct, scale to arbitrary size, and can be easily tuned using adjustable parameters, for example, dimensionality and degree of "randomness".

RNs of the transition network type can be defined constructively by the choice of the initial node and the transformation rule set. Given a set of stoichiometry-preserving and reversible rules, repeated application of the rules to the initial node and subsequent nodes must terminate after a finite number of steps by combinatorial arguments. RNs of transition network type are thus finite, connected, undirected graphs that, when combined with an appropriate graph distance metric, define metric spaces.[21] Chemical distance is defined as the shortest-path distance on the networks of bond breaks and bond formations.[22,23] The finite size and metric structure distinguish transition networks from other types of RNs, which expand indefinitely unless the construction is stopped. In the following, we will only consider RNs of the transition type and will drop the additional qualification.

We construct and analyze RNs of different stoichiometries containing carbon, hydrogen, nitrogen, and oxygen (CHNO) atoms using polar reactivity rules. Subsequently, we compare these RNs to a panel of generative network models ranging from regular to random. For regular structures, we examine lattices with non-periodic (LNP) and periodic (LP) boundary conditions, both based on a single repeating pattern. Our analysis of random networks includes Erdős–Rényi (ER)[24] and bipartite random (BR), which are formed by creating network edges randomly with a predefined probability. Additionally, Watts–Strogatz (WS) networks[25] and modified regular lattices with rewiring (LNP-R and LP-R for non-periodic and periodic boundary conditions, respectively) contain a tunable degree of randomness via a rewiring probability parameter.

The networks are evaluated based on their structural characteristics, which are categorized into local, metric, and global properties. Local properties take into account only the node adjacencies; examples of local properties are node degree and clustering coefficient. Metric properties involve measures such as average and maximum distances between nodes. Finally, global properties consider aspects like connectedness and network dimension.[26–28]

Different network properties are relevant in different circumstances, indicating no simple universal approach for network comparison exists. For example, scale-free (SF) networks are defined by their degree distribution,[29,30] while small-world networks, a network class that includes WS networks, are characterized by their diameter (or average shortest path length) and clustering coefficient.[26–28,31,32] Specialized search algorithms have been developed for SF networks, utilizing their heterogeneous structure.[33,34] In contrast, for other types of networks, the primary predictor of the search time is the network dimensionality,[18,19,35,36] typically defined by the box-covering dimension,[15–17,37] as specialized search algorithms are not usually available.

Exploration of PESs is currently a very active area of research. In addition to rules-based methods by us[13,14,38] and others,[9,10,39–42] accelerated molecular dynamics methods are important tools for exploring PES, for example, metadynamics,[43–45] artificial force-based,[46–48] manual steering,[49] and ab initio nanoreactor[50–53] approaches. Moreover, reaction template-based methods[54–57] and machine learning approaches[58–61] have been applied with success. Quantum chemistry-based exploration generates stoichiometry-preserving networks, in which the nodes correspond to individual molecular conformations.[46–48,62] Each node of the RNs considered here can be understood as a collection of all conformations for a given constitution (bonding pattern) and thus a coarse-grained representation of the quantum chemistry-based networks. These and other methods can benefit from a better understanding of the underlying data structure.

This paper is organized as follows. The details of constructing the discrete RN representations and the generative network models, along with the analysis of the networks, are presented in Section 2. Section 3 contains the analysis of the local, metric, and global network properties of RNs and the comparison between RNs and sized-matched generative models networks. We discuss the results in Section 4 and give our conclusions in Section 5.

# 2 Methods

## 2.1 Reaction Networks

The procedure for generating RNs is described elsewhere;[13,14] thus, we only give a brief overview. The RNs are collections of molecules subject to the constraint that the total stoichiometry (total number of atoms) within a node across all molecules in that node must be fixed. The systems explored here were initialized from a combination of formaldehyde ($CH_2O$), hydrogen cyanide (HCN), and water ($H_2O$) molecules (CHNO networks). These molecules were chosen for their simplicity and the diversity of atoms they contain, within the constraint of only using CHNO atoms. The initial choice of starting molecules is not critical as only the number of each atom type is important, because the network generation process is exhaustive. CHNO networks with up to 4 carbon atoms were generated. For each combination of HCN and $CH_2O$ molecules, up to 15 $H_2O$ molecules were added, creating a homologous series of RNs. The pure CHO networks (which are initialized with $CH_2O$ and $H_2O$ molecules) have an average carbon oxidation state of 0, while in the CHNO networks, the average carbon oxidation number is between 0 and +2. We use Hill notation[63] to denote the RN stoichiometries. The edges of the RNs are stoichiometry-preserving reactive transformations. The transformation rules are chosen to be reversible and complete with respect to normal polarity rules and are presented in Table 1. The decomposition into polar transformation covers more than just polar reactions; in fact, all reactions that can be formally written as a combination of electron pair transfers ("arrow pushing")[64,65] are representable in this way.[38]

All RNs were built using the open-source `colibri2` software package.[66] The Python code was parallelized to decrease the network construction times and to enable the generation and analysis of larger networks. With the parallelization and optimization performed, RNs composed of up to 57 atoms could be built. The construction of the largest RN ($C_3N_1O_{10}H_{21}$) required three days on 48 CPU cores. The reaction rules were encoded as SMARTS strings[67]

Table 1: Polar CHNO transformation rules used in this work.

| Bond Dissociation | Bond Association | Bond Breaking | Bond Formation |
|---|---|---|---|
| $H{-}C \rightarrow H^{+}\cdots C^{-}$ | $H^{+}\cdots C^{-} \rightarrow H{-}C$ | $C{=}C \rightarrow C^{+}{-}C^{-}$ | $C^{+}{-}C^{-} \rightarrow C{=}C$ |
| $C{-}C \rightarrow C^{+}\cdots C^{-}$ | $C^{+}\cdots C^{-} \rightarrow C{-}C$ | $C{\equiv}C \rightarrow C^{+}{=}C^{-}$ | $C^{+}{=}C^{-} \rightarrow C{\equiv}C$ |
| $H{-}O \rightarrow H^{+}\cdots O^{-}$ | $H^{+}\cdots O^{-} \rightarrow H{-}O$ | $C{=}O \rightarrow C^{+}{-}O^{-}$ | $C^{+}{-}O^{-} \rightarrow C{=}O$ |
| $H{-}O^{+} \rightarrow H^{+}\cdots O$ | $H^{+}\cdots O \rightarrow H{-}O^{+}$ | $C{\equiv}O \rightarrow C^{+}{=}O^{-}$ | $C^{+}{=}O^{-} \rightarrow C{\equiv}O$ |
| $C{-}O \rightarrow C^{+}\cdots O^{-}$ | $C^{+}\cdots O^{-} \rightarrow C{-}O$ | $C{=}O^{+} \rightarrow C^{+}{-}O$ | $C^{+}{-}O \rightarrow C{=}O^{+}$ |
| $C{-}O^{+} \rightarrow C^{+}\cdots O$ | $C^{+}\cdots O \rightarrow C{-}O^{+}$ | $C{\equiv}O^{+} \rightarrow C^{+}{=}O$ | $C^{+}{=}O \rightarrow C{\equiv}O^{+}$ |
| $H{-}N \rightarrow H^{+}\cdots N^{-}$ | $H^{+}\cdots N^{-} \rightarrow H{-}N$ | $C{=}N \rightarrow C^{+}{-}N^{-}$ | $C^{+}{-}N^{-} \rightarrow C{=}N$ |
| $H{-}N^{+} \rightarrow H^{+}\cdots N$ | $H^{+}\cdots N \rightarrow H{-}N^{+}$ | $C{\equiv}N \rightarrow C^{+}{=}N^{-}$ | $C^{+}{=}N^{-} \rightarrow C{\equiv}N$ |
| $C{-}N \rightarrow C^{+}\cdots N^{-}$ | $C^{+}\cdots N^{-} \rightarrow C{-}N$ | $C{=}N^{+} \rightarrow C^{+}{-}N$ | $C^{+}{-}N \rightarrow C{=}N^{+}$ |
| $C{-}N^{+} \rightarrow C^{+}\cdots N$ | $C^{+}\cdots N \rightarrow C{-}N^{+}$ | $C{\equiv}N^{+} \rightarrow C^{+}{=}N$ | $C^{+}{=}N \rightarrow C{\equiv}N^{+}$ |

and applied to the molecular structures using the RDKit library.[68] The RN construction terminated when no new nodes could be generated. The networks were stored as compressed GraphML files.[69] The NetworKit library[70] was used for all RN manipulations and analyses.

## 2.2 Network Properties

We analyzed the constructed CHNO networks with respect to their basic properties (number of nodes $n$, number of edges $e$, and density $\rho$), local properties (average node degree $\overline{k}$ and average square clustering coefficient $\overline{C_4}$), metric properties (diameter $D$ and average shortest path length $\bar{l}$), and global properties (fractal dimension $d_\mathrm{f}$ and average growth exponent $\overline{\xi}$). The analysis treated the RNs as undirected simple graphs. To establish notation, we briefly describe the network properties and their computation below.

The network density $\rho = 2e/(n\,(n-1))$ is defined as the proportion of edges $e$ in the network relative to the complete graph with $n$ nodes. The degree $k_u$ of the node $u$ describes the number of edges incident to $u$, with the average node degree $\overline{k}$ defined as the mean over all network nodes. The clustering in the neighborhood of the node $u$ is typically described by its (triangle) clustering coefficient $C_u$.[26] However, as RNs are bipartite by construction because the number of bonds in a node is either even or odd when the multiplicity of bonds is included,[14] the conventional clustering coefficient is always zero and thus not a useful

measure. Instead, we use the square clustering coefficient $C_{4,u}$, which is defined for the node $u$ as the proportion of the pairs of its neighbors $v, w$ that share a neighboring node different from $u$,[71]

$$C_{4,u} = \sum_{\substack{v<w \\ v,\,w=1,\,\ldots,\,k_u}} q_{uvw} \Big/ \sum_{\substack{v<w \\ v,\,w=1,\,\ldots,\,k_u}} \left[(k_v - \eta_{uvw})(k_w - \eta_{uvw}) + q_{uvw}\right] . \tag{1}$$

Here, $q_{uvw}$ is the number of squares which involve $u$, $v$, and $w$, and $\eta_{uvw} = 1 + q_{uvw} + \theta_{vw}$ with $\theta_{vw} = 1$ if $v$ and $w$ are adjacent and 0 otherwise.[72] The mean of $C_{4,u}$ over all network nodes yields the average square clustering coefficient $\overline{C_4}$.

For an undirected graph with the shortest path distance between nodes $u, v \in 1, \ldots, n$ denoted as $l_{uv}$, the diameter $D$ and the average shortest path length $\bar{l}$ are defined by the maximum and the mean over all pairs of nodes, respectively, $D = \max_{u<v} l_{uv}$ and $\bar{l} = 2/(n(n-1)) \sum_{u<v} l_{uv}$. Small-world networks are characterized by a logarithmic dependence of the metric properties on the network size, $\bar{l} \sim \ln n$.[26–28,31,32]

The calculation of the fractal dimension, $d_\mathrm{f}$, was performed using the compact box burning (CBB) algorithm,[16,17,73,74] which approximates the box-covering dimension.[15,17,37,75] In preliminary studies, the CBB algorithm proved to be more efficient than the alternative maximum excluded mass burning and greedy clustering algorithms.[15] As the CBB algorithm is stochastic, we determined the standard deviation of $d_\mathrm{f}$ values obtained in independent CBB calculations to be less than 1% in all cases. Therefore, only one calculated CBB fractal dimension is reported per RN. The CBB algorithm was parallelized over box sizes, which allowed the calculation of the CBB fractal dimension of the $C_2H_{20}N_2O_9$ network with $n$ = 1,226,147 nodes in seven days on 40 CPU cores.

The average growth factor $\overline{\xi}$ describes the increase of the number of nodes $n_u^i$ in the cluster or radius $i$ centered at the node $u$, that is, $n_u^i = |\{v|l_{uv} \leq i\}|$. The logistic fit of $n_u^i$ versus $i$ yields the exponent $\gamma_u$. The average growth factor is then defined as $\overline{\xi} = \exp(\overline{\gamma})$, where the mean is taken over all network nodes. The calculation of metric and global properties

requires the matrix of shortest path lengths between all pairs of nodes, which was determined using parallel breadth-first search (BFS).[70]

## 2.3 Generative Models

Size-matched networks were constructed using a variety of generative models, ranging from random to regular. Erdős–Rényi (ER) networks[24] and bipartite random (BR) networks[76] are obtained by a random assignment of edges with a fixed probability $p_c$ for a given number of nodes $n$. BR networks differ from ER networks in that they contain random edges between two disjoint sets of nodes with the fraction of nodes in the first node set being $b_s$. Both ER and BR networks may contain disconnected components. Regular lattices with non-periodic boundary conditions (LNP) are constructed as $d$-dimensional Cartesian products of the one-dimensional path graph with itself. Connecting the nodes of the outer faces of the LNP networks with their opposite faces generates regular lattices with periodic boundary conditions (LP), which are free of edge effects.

In addition to fully random (ER and BR) and fully regular (LNP and LP) networks, we considered several classes of networks with tunable "randomness", which is achieved by random rewiring of a regular network. Watts–Strogatz (WS) networks[25] start from a ring graph with $n$ nodes, in which each node is additionally joined by edges with $K/2$ preceding and $K/2$ following nodes in the ring order. Shortcuts are then created by replacing some edges $uv$ with probability $p_r$ by $uw$, where the node $w$ is chosen with a uniform probability from the remaining $(n-2)$ nodes. Modified regular lattices (LNP-R and LP-R) generalize the WS construction to higher dimensions by applying random rewiring with probability $p_r$ to the LNP and LP networks, respectively. Random rewiring in the LNP-R and LP-R networks break the bipartite nature of 2D and 3D lattices. As a result of the random rewiring, WS, LNP-R, and LP-R networks may be disconnected.

Since the networks with randomness (ER, BP, WS, LNP-R, and LP-R networks) have variability in network structures, an average of five independent trials is reported for each

generative model. In the presence of disconnected components, the weighted average of the network properties by the number of nodes in each component is given. All generative models were constructed using the NetworkX library[77] software package and custom code (available on GitHub[78]). All networks were saved as compressed GraphML files,[69] and the NetworKit library[70] was used for all analyses.

# 3 Results

## 3.1 Reaction Network Properties

In this work, we consider CHNO networks, which are constructed from combinations of $CH_2O$, HCN, and $H_2O$ molecules using polar reactivity rules (Table 1). Each RN is uniquely identified by its stoichiometry, and the RNs differing only by the number of $H_2O$ molecules form a homologous series. The changes in key RN properties as functions of the number of atoms, $N$, are shown in Figure 1, with homologous series indicated by lines. The smallest RNs ($CH_2O$ and CHN) have path structures with $n = 4$ nodes and $e = 3$ edges; the largest RN has the $C_3H_{21}NO_{10}$ stoichiometry and consists of $n = 15{,}244{,}289$ nodes and $e = 101{,}328{,}842$ edges. In total, we investigate 114 RNs belonging to 11 homologous series.

The network size, $n$, increases steeply with the number of carbon atoms; for the same number of carbon atoms, the RNs with higher HCN content are larger than those with higher $CH_2O$ content for the same number of $H_2O$ molecules, see Figure 1(a). Within each homologous series, $n$ initially increases rapidly as $H_2O$ molecules are added to the RN stoichiometry but eventually reaches a crossover point, after which the increase in $n$ is much slower and levels off as seen through the asymptotic behavior of the growth. Below the crossover point, the network nodes contain unsaturated molecules; the addition of $H_2O$ molecules introduces additional reactions, which contribute to the network growth. However, above the crossover point, any further increase in network size is due to bond breaking and formation in water, with no new reactivity added. The RNs are relatively sparse ($\rho < 10^{-2}$ in

all except the smallest RNs) and become even sparser as they increase in size, see Figure 1(b). The asymptotic behavior observed in $n$ is also seen in $\rho$ as $H_2O$ molecules are added. The reason for increasing sparsity with larger RNs is that the ratio of new nodes to new edges created gets larger with increasing numbers of atoms. Since the proportion of new nodes to new edges created gets larger for larger RNs, increasingly larger RNs will become sparser and sparser.

Insights into the distances within a RN are given by the metric properties, $D$ and $\bar{l}$. Figure 1(c) shows that even for large networks, $\bar{l}$ remains relatively small. The growth of $\bar{l}$ shows a polynmial nature with $N$ as $H_2O$ molecules are added, meaning, on average, it takes only a few steps to get from one node to another. For example, even in the largest network we considered, $C_3H_{21}NO_{10}$, the average shortest path length is only $\bar{l} = 15.41$. The smaller (water-poor) members of each homologous series follow quite accurately the logarithmic relationship between the number of nodes, $n$, and $\bar{l}$, which is characteristic of small-world networks. However, after the crossover to water-rich RNs, $\bar{l}$ increases faster than logarithmically with $n$. For the same number of $H_2O$ molecules and carbon atoms, the average distances increase with higher HCN content. The network diameter, $D$, shows qualitatively the same behavior as $\bar{l}$.

The local properties of average node degree, $\overline{k}$, and average square clustering coefficient, $\overline{C_4}$, provide insight into the connectivity among nearby nodes. As Figure 1(d) shows, $\overline{k}$ increases rapidly with the number of atoms in water-poor RNs before leveling off in water-rich members of each homologous series as seen through the asymptotic growth. The small $\overline{k}$ means that overall, most nodes have relatively few connections. RNs rich in HCN have higher average node degrees, reflecting a larger diversity of reactivity patterns compared to RNs with higher $CH_2O$ content. The values of $\overline{C_4}$, follow the opposite trend, as Figure 1(e) shows. With the exception of the two RNs with linear structures (CHN and $CH_2O$), which have no square clusters, $\overline{C_4}$ values decrease with increasing network size. Within a homologous series, $\overline{C_4}$ initially decreases rapidly when $H_2O$ molecules are added but appears to approach

a constant value for water-rich RNs as seen through the asymptotic growth. Less squares of nodes means fewer nodes are interconnected locally with each other. Similarly, the increase in the average node degree, $\overline{k}$, with higher HCN content corresponds to a decrease in the average square clustering coefficient, $\overline{C_4}$.

We considered two global RN properties, the average growth factor, $\overline{\xi}$, and the fractal dimension, $d_\mathrm{f}$, which offer an understanding of the broader structures within the networks. The average growth factor, $\overline{\xi}$, which indicates how many nodes can be reached within a given number of steps, shows only small differences between different RN stoichiometries and sizes. All RNs considered in this work have an average growth factor close to 2. Moreover, $\overline{\xi}$ decreases slightly with the number of atoms, consistent with lower network connectedness in larger RNs observed in smaller average square clustering coefficients, $\overline{C_4}$. A notable exception is the homologous series derived from $CH_2O$. We note that the growth factor does not seem to be correlated with the HCN content, unlike all other network properties. The origin of this convergence is not clear.

We have investigated in detail the fractal dimensions, $d_\mathrm{f}$, of CHO RNs approximated by the CBB and two other approximate methods, maximum excluded mass burning (MEMB) and greedy coloring, in previous work.[14] These results and our preliminary studies of CHNO networks showed that the CBB method was sufficiently reliable and computationally efficient for our purposes. Moreover, we estimated the statistical errors of the CBB method due to its stochastic algorithm using the standard deviations over five independent calculations in $C_2H_8N_2O$, $C_3H_8O_4$, $C_3H_5NO_2$, and $C_3H_4N_2O$ RNs, which yielded $3.63 \pm 0.01$, $3.54 \pm 0.01$, $3.540 \pm 0.004$, and $3.55 \pm 0.02$, respectively. Since the statistical errors were on the order of 1%, only one CBB estimation of the $d_\mathrm{f}$ value was performed for all other RNs. For all networks, for which the calculation of $d_\mathrm{f}$ was feasible, $d_\mathrm{f}$ was found to be less than 5, as Figure 1(f) shows. The value of $d_\mathrm{f}$ stays below 5 because of the logarithmic growth seen. In the base RNs of each homologous series (containing zero $H_2O$ molecules), the fractal dimension is $d_\mathrm{f} \approx 1$ for one-carbon RNs, $d_\mathrm{f} \approx 2$ for two-carbon RNs, $d_\mathrm{f} \approx 3$ for three-carbon

RNs, and $d_f = 4.55$ for $C_4H_8O_4$, the only four-carbon RN, for which the calculation of $d_f$ was feasible. However, the values of $d_f$ of RNs with the same number of carbon atoms diverge as $H_2O$ molecules are added, and the networks with higher HCN content show a faster increase in the value of $d_f$ compared to the RNs with higher $CH_2O$ content. The initial rapid increase levels off with the crossover from water-poor to water-rich RNs. The complete data on the CHNO networks is given in Section S4 of the Supporting Information (SI).

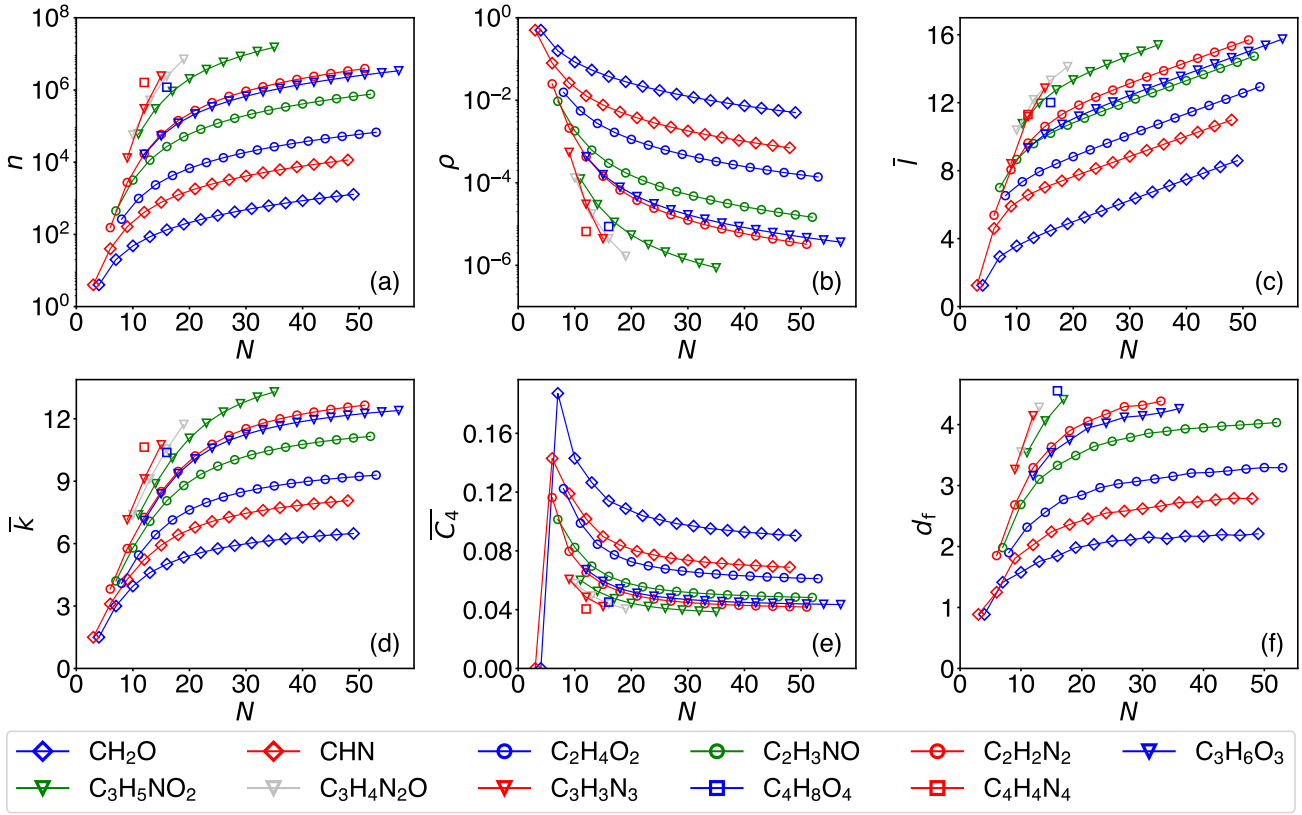



Figure 1: Variation of RN properties with the number of atoms, $N$: (a) number of nodes, $n$, (b) density, $\rho$, (c) average shortest path length, $\bar{l}$, (d) average node degree, $\overline{k}$, (e) average square clustering coefficient, $\overline{C_4}$, (f) fractal dimension, $d_f$. Homologous series are indicated by lines and denoted by their base RNs.

In addition to the averages of node degrees, $\overline{k}$, and square clustering coefficients, $\overline{C_4}$, we considered their distributions across a network in order to evaluate the heterogeneity of the RNs with respect to local properties. The shape of the degree distribution depends on the network size and composition, with the smaller water-poor networks having a tail-heavy (positively skewed) distribution and the larger, in particular, water-rich networks

showing a normal distribution. Figure 2(a) shows the distribution of node degrees, $k$, in the water-rich $C_3H_{19}NO_9$ network with a $\overline{k} = 13.04$ and standard deviation $\Delta k = 3.49$. The water-poor networks thus show some amount of heterogeneity, but the addition of $H_2O$ molecules makes them more homogeneous, decreasing the relative spread of the node degrees compared to the mean. The Jarque–Bera tests[79] of node degree histograms confirm that the deviation from normality decreases with increasing $H_2O$ content and that water-rich RNs are accepted as normal at $p = 0.05$ significance level. In previous work, we noted that the node degree distributions in CHO RNs could be described by heavy-tail lognormal distributions.[14] However, this finding seems to hold only for the smaller, water-poor members of each homologous series. The node degree distributions of the other RNs are given in Section S6 of the SI.

The distribution of square clustering coefficients, $C_4$, is right-skewed, as shown in Figure 2(b) for the $C_3H_{19}NO_9$ network. It fits a lognormal distribution, with $\overline{\ln(C_4)} = -3.26$ and $\Delta \ln(C_4) = 0.20$. Nonetheless, the distribution of $C_4$ values is relatively narrow. The $C_4$ distributions of the other RNs are compiled in Section S7 of the SI.

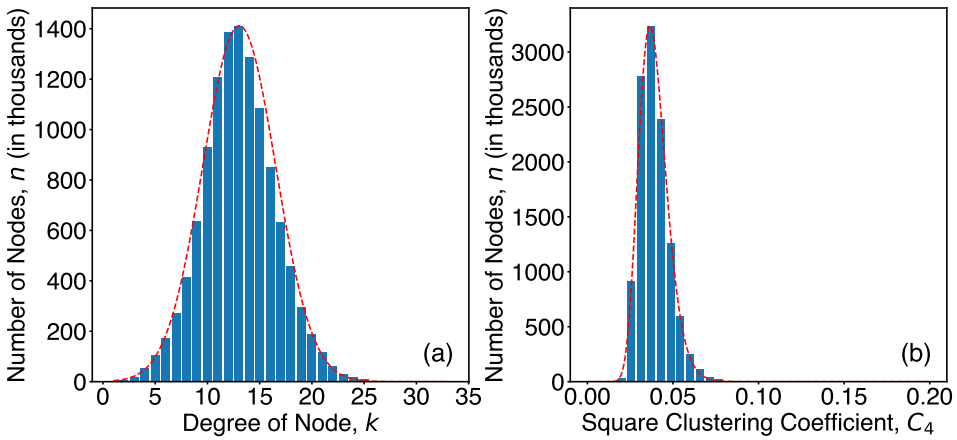



Figure 2: Histograms of (a) node degrees, $k$ (bin width $w_k = 1$), and (b) square clustering coefficients, $C_4$ (bin width $w_C = 5.7\times10^{-3}$ ) in the $C_3H_{19}NO_9$ network. A normal distribution is fitted for $k$, and a lognormal distribution is fitted for $C_4$. The fitted distributions are shown by dashed lines.

Structural commonalities of all RNs, irrespective of their composition, can be made visible by considering network properties as a function of the network size, $n$, across all

homologous series. Figure 3 shows the dependence of the three selected network properties: average shortest path length, $\bar{l}$ (metric property), average square clustering coefficient, $\overline{C_4}$ (local property), and fractal dimension $d_\mathrm{f}$ (global property) as a function of $n$. The value of $\bar{l}$ varies approximately logarithmically with the network size, $\bar{l} \sim \ln n$, a behavior characteristic of small-world networks,[26–28,31,32] see Figure 3(a). The linear fit of $\bar{l}$ against $\ln n$ yields the correlation coefficient $R^2 = 0.973$. However, water-rich RNs have higher values of $\bar{l}$ compared to the logarithmic relationship, likely due to their more regular structure, which deviates from the small-world behavior. The $\overline{C_4}$ values are close to inversely proportional to $\ln n$, as shown in Figure 3(b). The correlation coefficient of $1/\overline{C_4}$ against $\ln n$ is $R^2 = 0.995$. The logarithmic scaling of the value of $d_\mathrm{f}$ with the network size was already observed in CHO networks.[14] As Figure 3(c) shows, the universal relationship also holds for CHN and mixed CHNO networks, despite the addition of rule sets of carbon–nitrogen reactivity (Table 1). The linear fit of $d_\mathrm{f}$ against $\ln n$ has $R^2 = 0.992$. The water-rich RNs deviate from the scaling relationships for the average square clustering coefficients, $\overline{C_4}$, and $d_\mathrm{f}$ yielding $\overline{C_4}$ slightly above and $d_\mathrm{f}$ below the general trend.

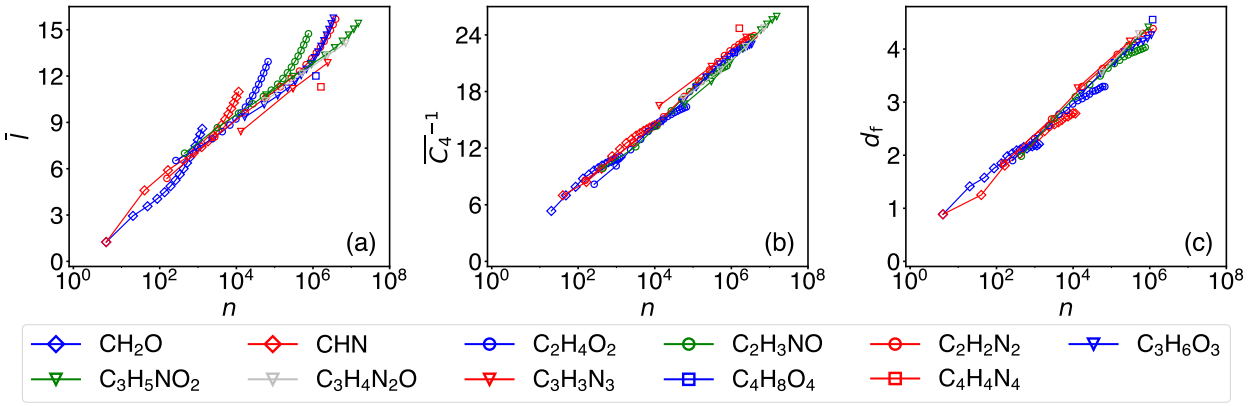



Figure 3: Variation of network properties with the number of nodes, $n$: (a) average shortest path length, $\bar{l}$, (b) inverse average square clustering coefficient, $\overline{C_4}^{-1}$ (excluding CHN and $CH_2O$ networks), (c) fractal dimension, $d_\mathrm{f}$. Homologous series are indicated by lines.

## 3.2 Comparisons with Generative Models

We selected ten CHNO networks of varying sizes and compositions for comparisons with sized-matched generative models. The comparison RNs are grouped into three size classes: small ($C_2H_5NO_2$, $CH_{19}NO_9$, and $C_2H_{10}O_5$) with $n \approx 4000$ nodes; medium ($C_3H_6O_3$, $C_2H_6N_2O_2$, and $C_2H_{18}O_9$) with $n \approx 17{,}000$; and large ($C_3H_8O_4$, $C_3H_4N_2O$, $C_2H_8N_2O_3$, and $C_3H_5NO_2$) with $n \approx 56{,}000$. These RNs are compared to generative models consisting of Erdős–Rényi networks (ER), bipartite random (BR) networks, Watts–Strogatz small-world networks (WS), regular lattices with non-periodic (LNP) and periodic (LP) boundary conditions as well as regular lattice networks with rewiring (LNP-R and LP-R).

As discussed in Section 3.1, the properties of RNs are strongly correlated with the number of network nodes, $n$. For brevity, we thus show only one RN for each size class in the following: $CH_{19}NO_9$ for small comparison RNs, $C_2H_6N_2O_2$ for medium comparison RNs, and $C_3H_4N_2O$ for large comparison RNs. Their properties are summarized in Table 2. The properties of all comparison RNs are listed in Section S4 of the SI.

Table 2: Selected network properties of $CH_{19}NO_9$ (small), $C_2H_6N_2O_2$ (medium), and $C_3H_4N_2O$ (large) comparison RNs: number of nodes, $n$; number of edges, $e$; average node degree, $\overline{k}$; average square clustering coefficient, $\overline{C_4}$; average shortest path length, $\bar{l}$; fractal dimension, $d_\mathrm{f}$.

| RN | $n$ | $e$ | $\overline{k}$ | $\overline{C_4}$ | $\bar{l}$ | $d_\mathrm{f}$ |
|---|---|---|---|---|---|---|
| $CH_{19}NO_9$ | 4131 | 15,404 | 7.46 | $7.37 \times 10^{-2}$ | 8.84 | 2.62 |
| $C_2H_6N_2O_2$ | 17,014 | 61,762 | 7.26 | $6.53 \times 10^{-2}$ | 9.62 | 3.29 |
| $C_3H_4N_2O$ | 56,485 | 208,955 | 7.40 | $5.78 \times 10^{-2}$ | 10.39 | 3.55 |

A comprehensive comparison of RNs and generative network models should consider their local, metric, and global properties. In Figure 4 we present pairwise plots of average square clustering coefficients, $\overline{C_4}$, average shortest path lengths, $\bar{l}$, and fractal dimensions, $d_\mathrm{f}$, in RNs and size-matched generative network models. The plots exclude the random networks (ER and BR) to improve the readability as these networks differ substantially from RNs with respect to all three properties. The full pairwise plots can be found in Section S3 of the

SI. In the following, we briefly discuss the properties of the size-matched generative models in comparison to the RNs. The complete data set of generative model networks is available in Section S5 of the SI.

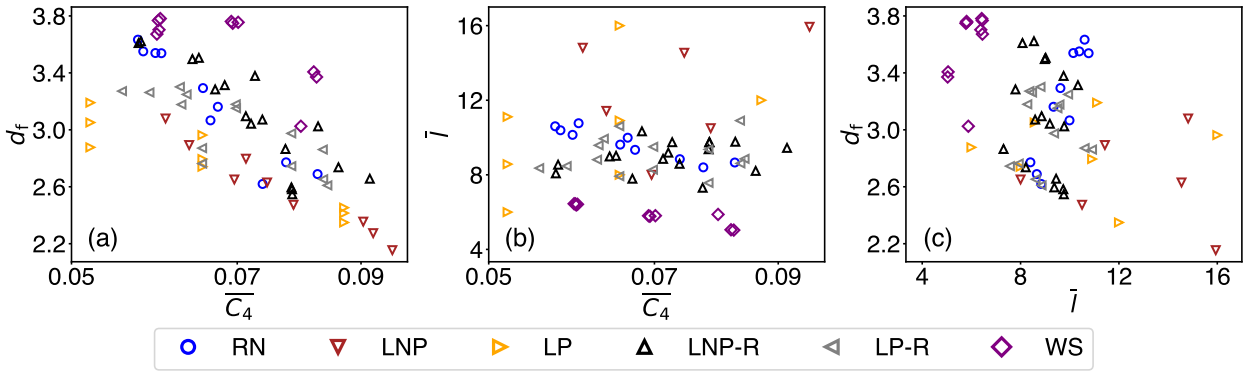



Figure 4: Comparisons between network properties of RNs (circles) and generative models: Watts–Strogatz networks (WS, diamonds), regular lattices with $d = 3$–5, non-periodic boundary conditions (LNP, down triangles) and periodic boundary conditions (LP right-facing triangles), LNP networks with rewiring (LNP-R, up triangles), and LP networks with rewiring (LP-R, left-facing triangles): (a) fractal dimensions, $d_f$, against average square clustering coefficient, $\overline{C_4}$, (b) average shortest path length, $\bar{l}$, against $\overline{C_4}$, and (c) $d_f$ against $\bar{l}$. See text for more details.

**Erdős–Rényi Networks** The ER networks are constructed by choosing the number of nodes, $n$, and creating edges randomly between pairs of nodes with a fixed probability $p_c$. The properties of the size-matched ER networks are shown in Table 3. These models behave very differently from their RN counterparts. We note that the average shortest path length, $\bar{l}$, of ER networks is considerably smaller, being nearly half that of the corresponding RN. For example, the $C_3H_4N_2O$ network has $\bar{l} = 10.39$, while the size-matched ER network has only $\bar{l} = 5.69$ on average. The average square clustering coefficients, $\overline{C_4}$, of ER networks are two to three orders of magnitude smaller than those in RNs, and the fractal dimensions, $d_f$, are higher by 0.6 to 1.1 units than those of the corresponding RNs. Properties of the other six size-matched ER networks and details of the construction are given in Section S5.1 of the SI.

Table 3: Selected properties of sized-matched Erdős–Rényi (ER) networks. See Table 2 for notation.

| ER | $n$ | $e$ | $\overline{k}$ | $\overline{C_4}$ | $\bar{l}$ | $d_\mathrm{f}$ |
|---|---|---|---|---|---|---|
| ER($CH_{19}NO_9$) | 4131 | 15,395.2 | 7.45 | $8.50 \times 10^{-4}$ | 4.37 | 3.71 |
| ER($C_2H_6N_2O_2$) | 17,014 | 61,701.4 | 7.25 | $1.94 \times 10^{-4}$ | 5.14 | 3.93 |
| ER($C_3H_4N_2O$) | 56,485 | 209,125.4 | 7.40 | $6.60 \times 10^{-5}$ | 5.69 | 4.24 |

**Bipartite Random Networks** The BR networks reflect the bipartite structure of RNs[14] while introducing randomness similar to ER networks. The construction of BR networks uses $n$ nodes, which are split into two partitions according to the fraction $b_\mathrm{s}$. Subsequently, $e$ edges are randomly assigned between the partitions. As the RNs are all very close to parity ($b_\mathrm{s}(CH_{19}NO_9) = 0.502$, $b_\mathrm{s}(C_2H_6N_2O_2) = 0.528$, and $b_\mathrm{s}(C_3H_5NO_2) = 0.505$), we chose $b_\mathrm{s} = 0.5$ in all constructed BR networks. Table 4 shows the properties of the size-matched BR networks, which are similar to those of the ER networks and very unlike the corresponding RNs. The average square clustering coefficients, $\overline{C_4}$, of BR networks are about one to two orders of magnitude larger than those of ER networks but still much lower than those of the corresponding RNs. Similar to the ER networks, the fractal dimensions, $d_\mathrm{f}$, of BR networks are higher by 0.6 to 1.0 units compared to the corresponding RNs, while the average shortest path lengths, $\bar{l}$, are nearly half of those in RNs. Properties of the other six size-matched BR networks are given in Section S5.2 of the SI.

Table 4: Selected properties of size-matched bipartite random (BR) networks with partition fraction $p_\mathrm{s} = 0.5$. See Table 2 for notation.

| BR | $n$ | $e$ | $\overline{k}$ | $\overline{C_4}$ | $\bar{l}$ | $d_\mathrm{f}$ |
|---|---|---|---|---|---|---|
| BR($CH_{19}NO_9$) | 4131 | 15,404 | 7.46 | $1.75 \times 10^{-3}$ | 4.58 | 3.57 |
| BR($C_2H_6N_2O_2$) | 17,014 | 61,762 | 7.26 | $4.24 \times 10^{-4}$ | 5.35 | 3.98 |
| BR($C_3H_4N_2O$) | 56,485 | 208,955 | 7.40 | $1.26 \times 10^{-4}$ | 5.91 | 4.29 |

**Watts–Strogatz Networks** The WS networks allow the tuning of the randomness in a network by the $p_\mathrm{r}$ parameter, which corresponds to the probability of random edge rewiring. For moderate values of $p_\mathrm{r}$, WS networks exhibit the small-world behavior, that is, $\bar{l} \sim \ln n$, as

well as significant local ordering, expressed by high values of clustering coefficients.[26–28,31,32] Table 5 shows the properties of WS networks, which most closely approximate the average square clustering coefficients, $\overline{C_4}$, and the number of edges, $e$, of the corresponding RNs.

The variation of the values of the network properties as a function of $p_\mathrm{r}$ shows that the best match between the $\overline{C_4}$ value and the fractal dimension, $d_\mathrm{f}$, of the WS networks and their respective RN counterparts is obtained with a moderate amount of rewiring, $p_\mathrm{r} = 0.25$–$0.3$. However, this level of randomness creates much smaller average shortest path lengths, $\bar{l}$, relative to the corresponding RN. The value of $d_\mathrm{f}$ of WS networks is lower than that of random networks but 0.2 to 0.7 units higher than those of the corresponding RNs, and the value of $d_\mathrm{f}$ shows a much slower increase with the network size. On the other hand, the closest match of the average shortest path length, $\bar{l}$, is obtained for a significantly reduced amount of rewiring. For example, the $C_3H_4N_2O$ network with $p_\mathrm{r} = 0.05$ results in $\bar{l} = 10.69$, an overestimation of the average square clustering coefficient, $\overline{C_4} = 2.5 \times 10^{-1}$, and an underestimation of the fractal dimension, $d_\mathrm{f} = 3.16$. Properties of the other six size-matched WS networks and other choices of the WS network parameters $K$ and $p_\mathrm{r}$ are given in and details of the construction Section S5.3 of the SI.

Table 5: Selected properties of size-matched Watts–Strogatz (WS) networks with $K = 8$ nearest neighbors and rewiring probability $p_\mathrm{r}$. See Table 2 for notation.

| WS | $p_\mathrm{r}$ | $n$ | $e$ | $\overline{k}$ | $\overline{C_4}$ | $\bar{l}$ | $d_\mathrm{f}$ |
|---|---|---|---|---|---|---|---|
| WS($CH_{19}NO_9$) | 0.25 | 4131 | 16,524 | 8.00 | $8.22 \times 10^{-2}$ | 5.03 | 3.37 |
| WS($C_2H_6N_2O_2$) | 0.275 | 17,014 | 68,056 | 8.00 | $7.01 \times 10^{-2}$ | 5.81 | 3.76 |
| WS($C_3H_4N_2O$) | 0.3 | 56,485 | 225,940 | 8.00 | $5.99 \times 10^{-2}$ | 6.42 | 3.78 |

**Non-Periodic Regular Lattices** The LNP networks are constructed for each given dimensionality, $d$, by choosing the side length $L$ to match the number of nodes, $n$, to the corresponding RN. The best match for the number of edges, $e$, is found if the dimensionalities of the LNP and RN networks are similar, that is, $d = 3, 4$. Since the dimensionality $d$ of LNP networks is integer, only approximate matching of both $n$ and $e$ is possible. The

calculations of the fractal dimensions, $d_{\mathrm{f}}$, of LNP networks help identify the systematic errors of the CBB approximation relative to the underlying dimensionality $d$. Table 6 shows the properties of a one-dimensional LNP network with $n = 5000$, a two-dimensional LNP network with $n = 500$, and size-matched regular lattices with $d = 3, 4$. The properties of additional LNP networks up to $d = 5$ can be found in Section S5.4 of the SI.

The regular lattices show a high degree of clustering, as indicated by high average square clustering coefficients, $\overline{C_4}$, which are similar to that of RNs and WS networks, see Figure 4. For example, $\overline{C_4} = 7.12 \times 10^{-2}$ , for the LNP network with $d = 4$ and $L = 16$. But LNP networks lack shortcuts typical of WS networks and, as a result, have long average shortest path lengths, $\bar{l}$, above those of RNs and far larger than random and WS networks. The CBB method underestimates the dimension of the underlying space; for example, for the one-dimensional chain of $L = 5000$, the CBB method gives $d_{\mathrm{f}} = 0.86$ , and for the two-dimensional $L \times L$ lattice with $L = 250$, the computed fractal dimension is $d_{\mathrm{f}} = 1.73$. The larger lattice with $L = 500$ has $d_{\mathrm{f}} = 1.75$, which shows that $d_{\mathrm{f}}$ increases towards $d = 2$ but the convergence is slow. We also find convergence toward the underlying dimensionality for $d = 3$, where the sized-matched lattice with $L = 15$ has $d_{\mathrm{f}} = 2.15$, while for $L = 50$, we find $d_{\mathrm{f}} = 2.40$. In LNP networks with $d = 4$, we obtain $d_{\mathrm{f}} = 2.41$ for the size-matched lattice with $L = 7$ and $d_{\mathrm{f}} = 2.88$ for $L = 20$. Although the theoretical dimensionality of the networks is underestimated by the CBB method, comparisons between networks of similar size and approximate dimensionality can be expected to be accurate.

**Periodic Regular Lattices** The LP networks better approximate an infinite lattice as they include periodic boundary conditions, removing edge effects. Table 7 shows the network properties for a one-dimensional LP network (ring) with $n = 5000$ nodes, a two-dimensional LP network with $n = 500$, and size-matched LP networks with $d = 3, 4$. The properties of additional LP networks up to $d = 5$ can be found in Section S5.5 of the SI. For the one-dimensional ring, the CBB method yields $d_{\mathrm{f}} = 0.93$, which indicates that the error of the CBB

Table 6: Selected properties of non-periodic regular lattices (LNP) with dimensionality $d$ and side length $L$. See Table 2 for notation.

| Lattice | $n$ | $e$ | $\overline{k}$ | $\overline{C_4}$ | $\bar{l}$ | $d_{\mathrm{f}}$ |
|---|---|---|---|---|---|---|
| $d = 1$, $L = 5000$ | 5000 | 4999 | 2.00 | 0 | 1666.67 | 0.86 |
| $d = 2$, $L = 500$ | 250,000 | 499,000 | 3.99 | $1.25 \times 10^{-1}$ | 333.33 | 1.75 |
| $d = 3$, $L = 15$ | 3375 | 9450 | 5.60 | $9.66 \times 10^{-2}$ | 14.93 | 2.15 |
| $d = 3$, $L = 26$ | 17,576 | 50,700 | 5.77 | $9.22 \times 10^{-2}$ | 25.96 | 2.27 |
| $d = 3$, $L = 39$ | 59,319 | 173,394 | 5.85 | $9.04 \times 10^{-2}$ | 38.97 | 2.36 |
| $d = 4$, $L = 7$ | 2401 | 8232 | 6.86 | $8.08 \times 10^{-2}$ | 9.14 | 2.41 |
| $d = 4$, $L = 12$ | 20,736 | 76,032 | 7.33 | $7.35 \times 10^{-2}$ | 15.89 | 2.68 |
| $d = 4$, $L = 16$ | 65,536 | 245,760 | 7.50 | $7.12 \times 10^{-2}$ | 21.25 | 2.80 |

method is less than 10%. However, the relative errors seem to grow as the dimensionality increases and the side length of the size-matched LP lattices decreases. In comparison to the corresponding RNs, the fractal dimensions of LP networks with $d = 3, 4$ are too low by 0.1–0.5 units. Periodic boundary conditions slightly reduce average shortest path lengths, $\bar{l}$, of regular networks; however, they remain larger than those of the corresponding RNs, see Figure 4. The average square clustering coefficients, $\overline{C_4}$, are somewhat reduced compared to the LP networks; for example, $\overline{C_4} = 6.52 \times 10^{-2}$ , in the LP network with $d = 4$ and $L = 16$. Moreover, the $\overline{C_4}$ values are independent of the side length, $L$.

Table 7: Selected properties of periodic regular lattices (LP) with dimensionality $d$ and side length $L$. See Table 2 for notation.

| Lattice | $n$ | $e$ | $\overline{k}$ | $\overline{C_4}$ | $\bar{l}$ | $d_{\mathrm{f}}$ |
|---|---|---|---|---|---|---|
| $d = 1$, $L = 5000$ | 5000 | 5000 | 2 | 0 | 1250.00 | 0.93 |
| $d = 2$, $L = 500$ | 250,000 | 500,000 | 4 | $1.25 \times 10^{-1}$ | 250.00 | 1.81 |
| $d = 3$, $L = 15$ | 3375 | 10,125 | 6 | $8.70 \times 10^{-2}$ | 11.20 | 2.29 |
| $d = 3$, $L = 26$ | 17,576 | 52,728 | 6 | $8.70 \times 10^{-2}$ | 19.50 | 2.42 |
| $d = 3$, $L = 39$ | 59,319 | 177,957 | 6 | $8.70 \times 10^{-2}$ | 29.23 | 2.45 |
| $d = 4$, $L = 7$ | 2401 | 9604 | 8 | $6.52 \times 10^{-2}$ | 6.86 | 2.63 |
| $d = 4$, $L = 12$ | 20,736 | 82,944 | 8 | $6.52 \times 10^{-2}$ | 12.00 | 2.87 |
| $d = 4$, $L = 16$ | 65,536 | 262,144 | 8 | $6.52 \times 10^{-2}$ | 16.00 | 2.96 |

**Regular Lattices with Rewiring** As Figure 4 illustrates, the RNs are located in the same range of average square clustering coefficients, $\overline{C_4}$, as WS networks and regular lattices, and at intermediate values of fractal dimension, $d_{\mathrm{f}}$, and average shortest path length, $\bar{l}$, between these two generative network models. In analogy to the edge randomization in WS networks, we thus modify LNP and LP networks by randomly rewiring edges with probability $p_{\mathrm{r}}$. We denote the resulting networks with tunable "randomness" as LNP-R and LP-R networks, respectively.

Table 8 shows the properties of the size-matched LNP-R and LP-R networks for $d = 3, 4$. As in WS networks, a small amount of rewiring increases the value of $d_{\mathrm{f}}$ and decreases the value of $\bar{l}$ compared to the parent regular lattices, see Figure 4. A modest degree of rewiring suffices to match the fractal dimension $d_{\mathrm{f}}$ of the LNP-R and LP-R networks to that of the corresponding RNs. The LNP-R network with $d = 3$, $L = 39$, and $p_{\mathrm{r}} = 0.1$ has $d_{\mathrm{f}} = 3.62$ , while the corresponding LP-R network has $d_{\mathrm{f}} = 3.27$. The average square clustering coefficient is also well matched, $\overline{C_4} = 5.54 \times 10^{-2}$. Smaller RNs need smaller values of $p_{\mathrm{r}}$ to match the value of $d_{\mathrm{f}}$ of the corresponding RNs. The LNP-R networks with $d = 3$ and $L = 15$ matches the corresponding RN at $p_{\mathrm{r}} = 0.05$, and the best matching LP-R network with $d = 3$ and $L = 15$ has $p_{\mathrm{r}} = 0.025$. With dimensionality $d = 4$, an even smaller amount of rewiring is sufficient; the LNP-R network with $L = 16$ and $p_{\mathrm{r}} = 0.05$ achieves $\overline{C_4} = 5.73 \times 10^{-2}$ and $d_{\mathrm{f}} = 3.61$.

The average shortest path lengths, $\bar{l}$, in these LNP-R and LP-R networks underestimate those of the corresponding RNs somewhat. Smaller amounts of rewiring produce a better match for the $\bar{l}$ value of the RNs. For example, the LNP-R network with $L = 16$ and $p_{\mathrm{r}} = 0.05$ has $\bar{l} = 8.08$, compared to $\bar{l} = 10.39$ in the corresponding RN, see also Table 8. At the same time, choosing $p_{\mathrm{r}} = 0.01$ gives $\bar{l} = 10.33$ but overestimates the values of $\overline{C_4}$ and $d_{\mathrm{f}}$, as shown in Figure 4. An intermediate value of $p_{\mathrm{r}} = 0.025$ offers the best trade-off between matching the metric, local, and global properties accurately. Generally, LNP-R networks match the corresponding RNs better than LP-R networks.

Table 8: Selected properties of periodic regular lattices with rewiring (LP-R) and non-periodic regular lattices with rewiring (NLP-R). $d$ denotes the lattice dimensionality, $L$ is the side length, and $p_\mathrm{r}$ is the rewiring probability. See Table 2 for notation.

| Lattice | | $p_\mathrm{r}$ | $n$ | $e$ | $\overline{k}$ | $\overline{C_4}$ | $\bar{l}$ | $d_\mathrm{f}$ |
|---|---|---|---|---|---|---|---|---|
| LNP | $d = 3$, $L = 15$ | 0.05 | 3375 | 9450 | 5.60 | $7.72 \times 10^{-2}$ | 7.11 | 2.79 |
| LP | $d = 3$, $L = 15$ | 0.025 | 3375 | 10,125 | 6.00 | $7.82 \times 10^{-2}$ | 7.33 | 2.75 |
| LNP | $d = 3$, $L = 26$ | 0.075 | 17,576 | 50,700 | 5.77 | $6.61 \times 10^{-2}$ | 8.00 | 3.27 |
| LP | $d = 3$, $L = 26$ | 0.05 | 17,576 | 52,728 | 6.00 | $6.98 \times 10^{-2}$ | 8.28 | 3.18 |
| LNP | $d = 3$, $L = 39$ | 0.1 | 59,319 | 173,394 | 5.85 | $5.76 \times 10^{-2}$ | 8.54 | 3.62 |
| LP | $d = 3$, $L = 39$ | 0.1 | 59,319 | 177,957 | 6.00 | $5.54 \times 10^{-2}$ | 8.36 | 3.27 |
| LNP | $d = 4$, $L = 7$ | 0.0075 | 2401 | 8232 | 6.86 | $7.81 \times 10^{-2}$ | 7.17 | 2.68 |
| LP | $d = 4$, $L = 7$ | 0.0001 | 2401 | 9604 | 8.00 | $6.52 \times 10^{-2}$ | 6.81 | 2.64 |
| LNP | $d = 4$, $L = 12$ | 0.025 | 20,736 | 76,032 | 7.33 | $6.60 \times 10^{-2}$ | 8.06 | 3.29 |
| LP | $d = 4$, $L = 12$ | 0.0001 | 20,736 | 82,944 | 8.00 | $6.52 \times 10^{-2}$ | 11.52 | 2.95 |
| LNP | $d = 4$, $L = 16$ | 0.05 | 65,536 | 245,760 | 7.50 | $5.73 \times 10^{-2}$ | 8.08 | 3.61 |
| LP | $d = 4$, $L = 16$ | 0.025 | 65,536 | 262,144 | 8.00 | $5.86 \times 10^{-2}$ | 8.47 | 3.26 |

Figure 5 illustrates the effect of random rewiring on the average square clustering coefficient, $\overline{C_4}$, and fractal dimension, $d_\mathrm{f}$, over a larger range of $p_\mathrm{r}$ values. Similar to WS networks, two ranges of behavior are observed. For small rewiring probabilities, $p_\mathrm{r}$, the value of $d_\mathrm{f}$ decreases by several units due to the occurrence of shortcuts, with only small changes in the average square clustering coefficient, $\overline{C_4}$. The networks are thus still relatively structured but gain the small-world property. For larger amounts of rewiring, the networks become increasingly random, and the values of $\overline{C_4}$ have a much stronger dependence on the rewiring probability, $p_\mathrm{r}$, while the fractal dimension levels off. This behavior is observed for both LNP-R and LP-R networks with $d = 3$ and $d = 4$. Interestingly, the networks that most closely approximate RNs in Figure 5 are located close to the transition between the two ranges. The significance of this finding is not yet clear. The complete set of data on LNP-R and LP-R networks is included in Sections S5.6 and S5.7 of the SI, respectively.

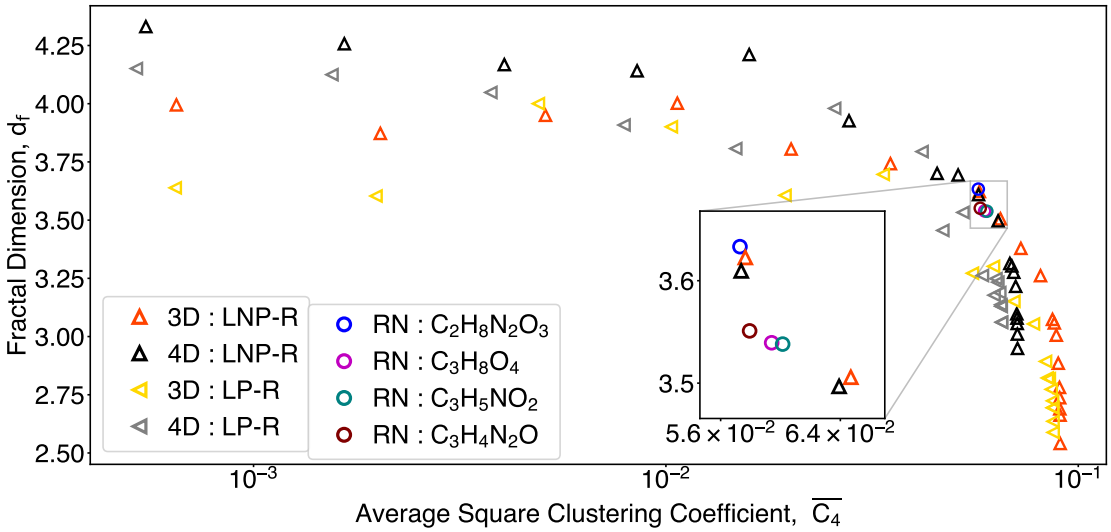



Figure 5: Plot of fractal dimension, $d_f$, against average square clustering coefficient, $\overline{C_4}$, for LNP-R networks with $d = 3$ (black triangles) and $d = 4$ (red triangles) and LP-R networks with $d = 3$ (gray triangles) and $d = 4$ (yellow triangles) in comparison to RNs (circles). The LNP-R and LP-R networks have $L = 39$ and $L = 16$ for $d = 3$ and $d = 4$, respectively.

# 4 Discussion

## 4.1 Reaction Network Characterization

As Figure 1 shows, the properties of the RNs within each homologous series follow similar profiles as $H_2O$ molecules are added to the base RN, with a characteristic crossover point between water-poor (smaller) and water-rich (larger) RNs. For a given homologous series, the number of nodes, $n$, of water-poor networks scales polynomially with the number of atoms, $N$, in line with our previous observations for CHO networks.[14] The size scaling of water-rich RNs is also polynomial in $N$, however, with a reduced slope. The crossover happens approximately between the third and sixth members of each homologous series and is reflected in local properties, such as average node degree, $\overline{k}$, and average square clustering coefficient, $\overline{C_4}$, metric properties, such as average shortest path length, $\bar{l}$, and diameter, $D$, and fractal dimension, $d_f$. The plots of all RN properties against $N$ and an illustration of the crossover point for selected networks and properties are given in Sections S1.1 and S2 of the SI, respectively.

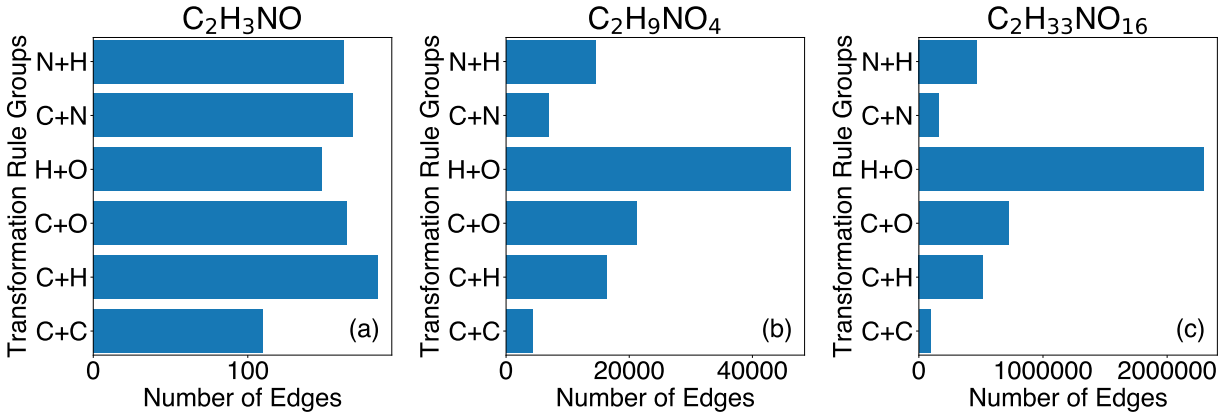



Figure 6: Numbers of network edges grouped by transformation rule in RNs of the $C_2H_3NO$ homologous series: (a) Base RN ($C_2H_3NO$), (b) RN containing three $H_2O$ molecules ($C_2H_9NO_4$), and (c) RN containing 15 $H_2O$ molecules ($C_2H_{33}NO_{16}$).

The water-poor RNs contain a wide distribution of transformation rules, including C−C, C−O and/or C−N, and E−H (E = C, O, N) rules and their multiple-bond counterparts (see Table 1). For example, Figure 6 shows the numbers of edges grouped by transformation rule in three RNs of the $C_2H_3NO$ homologous series. As Figure 6(a) shows, the base RN $C_2H_3NO$ has 936 edges, of which 11.8% correspond to C−C, 19.7% to C−H, 17.9% to C−N, 17.5% to C−O, 17.3% to N−H, and 15.8% to O−H transformation rules. In contrast, the transformations in water-rich RNs are dominated by the breaking and formation of O−H bonds in the solvent, masking the differences between RNs of different stoichiometries. In the water-rich $C_2H_{33}NO_{16}$ RN, which results from adding 15 $H_2O$ molecules to the above network, over 53.9% of its $4.25 \times 10^6$ edges, as shown in Figure 6(c), are due to O−H bond breaks and bond formations. For the RNs above the crossover point, no new chemical transformations are added; instead, each network node increases by an $H_2O$ unit. As a result, the RN with $N_w + 1$ $H_2O$ molecules (where $N_w$ is above the crossover point) is the product (Kronecker) graph of the preceding RN, containing $N_w$ $H_2O$, and the 2-node graph describing water dissociation.[14] We thus expect RNs to become more regular with increasing dilution. Distributions of edges by transformation rule in all RNs are given in Section S8 of the SI.

The increase in RN size as a result of adding $CH_2O$ and/or HCN monomers is exponential, as Figure 1(a) shows. In comparison with CHO networks,[14] CHN networks show a faster increase in network size, while mixed CHNO networks fall in between. CHN networks also grow faster in the water-poor regime with respect to their average shortest path length, $\bar{l}$, and average node degree, $\overline{k}$, compared to CHO networks. In water-rich RNs, the value of $\bar{l}$ and the diameter, $D$, increase by a constant for each additional $H_2O$ molecule. The increase of the average node degrees, $\overline{k}$, of the CHN and CHO networks also become similar in the water-rich regime. As for the network sizes, the mixed CHNO networks are intermediate in their increase between CHN and CHO networks.

We observe universal relationships for metric, local, and global network properties in a wide variety of CHNO networks shown in Figure 3. For the CHO networks, we have previously shown that the universal relationships for the fractal dimension, $d_{\mathrm{f}}$, are also independent of the formal carbon oxidation state.[14] These findings indicate that the underlying structural principles of RNs are independent of the specific transformation rules and are the result of the octet rule and the use of polar transformations (electron-pair shifts), see Table 1. Moreover, algorithms based on structural features of RNs, including embedding, sampling, and search, are likely to be broadly applicable. Investigations of elements that form hypervalent compounds, for example, phosphorous and sulfur, and of radical reactions will help shed light on this hypothesis. As the current study exclude some chemistry, if the rule set is expanded to include radical reactions, the bipartite nature of the RNs will be lost. The structural properties of RNs with radicals will be explored in future work. The small differences observed between CHO and CHNO networks implies that incorporating additional types of atoms and a large set of rules is unlikely to significantly alter the structural properties of the RNs.

As reported previously, the fractal dimension, $d_{\mathrm{f}}$, of RNs is considerably smaller than the dimensionality of the underlying PES, $d_{\mathrm{PES}} = 3N - 6$.[14] The computed $d_{\mathrm{f}}$ values range from 0.89 to 4.55 for the RNs, for which these calculations were feasible, see Figure 1(f).

The fractal dimension $d_f = 4.55$ was obtained for the $C_4H_8O_4$ RN, whose corresponding PES has dimensionality $d_{PES} = 42$. The dimensionality reduction is nearly tenfold. Despite the expanded rule set, we obtain similar reductions for CHN and mixed CHNO networks. For the $C_3H_3N_3$ and $C_3H_9NO_4$ RNs, the fractal dimensions $d_f = 3.55$ and $d_f = 4.41$ were obtained, with the corresponding PES dimensionality being $d_{PES} = 21$ and $d_{PES} = 45$, respectively, showing the nearly tenfold dimensionality reduction is irrespective of the stoichiometry for a range of RNs built with polar transformation rules as heuristics.

## 4.2 Comparisons to Generative Network Models

Generative network models are of interest because they can represent a multitude of network properties with just a few parameters. The dimensionalities of many such models have been studied in the literature. A particularly well-studied class of networks are scale-free (SF) networks, characterized by their highly heterogeneous structure, featuring a few highly connected nodes (hubs) alongside many nodes with low connectivity.[80,81] In SF networks, the dimension is influenced by the scaling exponent, $\lambda$ , which is a key parameter in the power-law distribution that describes the probability $P(k)$ that a node has $k$ connections, following the relationship $P(k) \sim k^{-\lambda}$. When $\lambda > 4$, the fractal dimension, $d_f$, in ER and SF networks, stabilizes at 6, indicating a departure from scale-free behavior due to a rapid decline in degree distribution. Typical SF networks studied have a $\lambda$ between 2 and 3. As $\lambda$ approaches 3, $d_f$ tends toward infinity. However, this holds only in the limit of infinitely large networks. For finite networks, an empirical $d_f$ can be estimated, but it would, in effect, only show the finite-size effects of SF networks, which is why they are not included in this study for comparison.[82]

The fractal nature of ER and SF networks at the boundaries (length scales above the average shortest path length, $\bar{l}$) has been explored and found to have a fractal dimension of $d_f \approx 2$. The nodes at the edges have a low degree, causing the low dimensionality.[83] For ER and other random graph models, analytical expressions for $\bar{l}$ and the diameter, $D$, have been

found.[84,85]

Dimensionality calculations of regular lattices using approximate box-covering methods have been reported in the literature.[86–88] Depending on the method to calculate the dimensionality of regular lattices, the dimensionality is either found to be exact or approach the dimensionality of the underlying space. The spectral dimensions of linear, square, and triangular lattices are found to be 1.0, 2.0, and 3.0 dimensional, respectively.[86] Using a random sequential box-covering algorithm, the fractal dimension, $d_f \approx 2.0$ was obtained for a $500 \times 500$ regular lattice.[87] The $50 \times 50$ LNP network was computed to have $d_f = 1.649$ instead of 2 using an approximate method because of the finite size,[88] which is comparable to $d_f = 1.68$ obtained in this work for a $100 \times 100$ LNP network.

The effective dimension of WS-like networks varies as a function of the length scale on which it is measured. A transition from a linear to logarithmic scaling of $\bar{l}$ is observed as the typical length of shortcuts changes, similar to the transition seen in Figure 5 for LNP-R and LP-R networks.[89] A transition in $\bar{l}$ from linear scaling to logarithmic scaling was also found on increasing edge rewiring probability, $p_r$.[90] Higher values of $p_r$ were shown to reduce the characteristic length scale in WS networks, which increases $d_f$.[91] In the random limit ($p_r = 1$), $d_f$ increases with the number of nodes, $n$.[92] Similarly, in this work, the fractal dimension, $d_f$, of WS, LNP-R, and LP-R networks increases with larger $p_r$ as shown for LNP-R and LP-R networks in Figure 5. The initial increase of $d_f$ with $p_r$ is quite steep but becomes shallower at $p_r \geq 0.2$, as shown in Figure 5. The increase of the fractal dimension, $d_f$, due to random rewiring in regular lattices was previously reported.[93,94]

Figure 4 shows the comparison of the average shortest path length, $\bar{l}$ (metric property), the average square clustering coefficient, $\overline{C_4}$ (local property), and the fractal dimension, $d_f$, between three size classes of RNs: small ($C_2H_5NO_2$, $CH_{19}NO_9$, and $C_2H_{10}O_5$), medium ($C_3H_6O_3$, $C_2H_6N_2O_2$, and $C_2H_{18}O_9$), and large ($C_3H_8O_4$, $C_3H_4N_2O$, $C_2H_8N_2O_3$, and $C_3H_5NO_2$) and size-matched generative network models excluding the random networks (ER and BR). We note that most of these networks belong to the water-poor regime. The water-poor RNs

provide a more stringent test of structural similarity to generative models since the networks become more regular with increasing dilution, as discussed above.

The existing body of work has demonstrated that randomness increases fractal dimension, $d_\mathrm{f}$, reduces average shortest path length, $\bar{l}$, and suppresses local clustering.[25,83,89–97] Consequently, the size-matched random ER and BR networks have values of $d_\mathrm{f}$ higher by 0.6 to 1.1 units and values of $\overline{C_4}$ smaller by one to three orders of magnitude in comparison to the RNs. BR networks show higher local clustering but are still much less ordered than RNs. Although both random networks and RNs show the same scaling of $\bar{l}$ with the network size, that is, belong to small-world networks, the random networks are "smaller" with respect to their metric properties by a factor of two. On the other end of the spectrum, size-matched regular lattice networks with $d = 3, 4$ are similar to the corresponding RNs with respect to their average square clustering coefficients but have longer average shortest path lengths and lower CBB fractal dimensions.

The WS networks demonstrate the effect of randomness on network properties within the same generative network model. The dependence of the clustering coefficient and the diameter on the rewiring probability, $p_\mathrm{r}$, was already described by Watts and Strogatz in their seminal paper.[25] Additionally, random rewiring increases the fractal dimension, $d_\mathrm{f}$, bringing it closer to the values observed in ER and BR networks. However, the WS model cannot simultaneously satisfy the matching conditions for metric, global, and local network properties.

The LNP-R and LP-R networks apply the random rewiring to low-dimensional regular lattices, as shown in Figure 5. These models with $d = 3, 4$ require a smaller amount of rewiring than WS networks to match the average square clustering coefficient, $\overline{C_4}$, and the fractal dimension, $d_\mathrm{f}$, of the corresponding RNs with good accuracy. Therefore, it is possible to simultaneously match metric, local, and global properties of RNs by regular lattices with rewiring. The necessary amount of rewiring in LNP-R and LP-R models is small, $p_\mathrm{r} \leq 0.1$, indicating that RNs are more regular than random. Thus, RNs are not only low-dimensional

*metric spaces*, but moreover, they resemble low-dimensional *Euclidean spaces*, represented by the regular lattices.

The similarities between RNs and LNP-R networks with respect to the local and metric properties suggest the presence of inherent symmetries within the RNs. The symmetry arises from the fact that the order of bond formations or breaking often does not matter, which creates regular patterns in the RNs, which have been observed in our previous work.[14,20] The regular patterns found in the RNs are also reflected in the relatively narrow distributions of the node degrees, $k$, and the square clustering coefficients, $C_4$, shown in Figure 2. Most nodes are thus embedded in similar local neighborhoods, as is the case in regular lattices. The above analysis was performed on water-poor RNs and applies even more to the water-rich RNs, which exhibit a more regular structure. Finally, we note that the RNs studied in this work did not take into consideration thermodynamic or kinetic feasibility of the reaction paths and constitute the superset of the experimentally observed reactions.[38] The dimensionality of the LNP-R and LP-R networks is thus an upper bound to the subnetworks of the RN that represent feasible reactions. To gain further insights into RNs, methods such as disconnectivity graphs,[98] which visualize energy landscapes, and persistent homology,[99] which captures the global topological features of the network, could be employed.

# 5 Conclusions

In this paper, we investigated local, metric, and global properties of RNs and showed that they share common structural characteristics irrespective of their size, composition, and rule set and can be modeled by low-dimensional non-periodic regular lattices with a small amount of random rewiring. Thus, in addition to being low-dimensional metric spaces,[14] RNs are also close to low-dimensional Euclidean spaces. The coordinate representation of RNs can be constructed by embedding into $\mathbb{R}^d$ with $d = 3, 4$ for the networks considered in this work, generating a tenfold dimensionality reduction with a small error. The reduced-dimensionality

representations are useful for speeding up sampling and search of RNs. Network embedding is widely used in similarity search, classification, clustering, and visualization tasks.[97,100–102]

RNs have the potential to be effectively utilized for identify transition states. Transition state searches would first locate a potentially feasible reaction pathway within a RN, and then interpolation between neighboring nodes would be employed to estimate the location of the corresponding transition state. RNs offer a practical starting point for initiating the transition state search. At the same time, the rewiring probability $p_\mathrm{r}$ in the LNP-R and LP-R networks offers an *a posteriori* measure of the average deviation from Euclidean structure and is related to the error of the dimensionality reduction between the parent PES and the low-dimensional Euclidean space.

The limitation of the current approach is that it requires explicit RN construction, which is limited by the computational cost and storage needs. However, the understanding of the dimension and amount of rewiring as a function of the RN composition, oxidation state, and rule set developed in this work provides the target dimensionality and error measures for the network embedding. These parameters can then be applied to embed subnetworks constructed around a compound or reaction of interest. On-the-fly calculations of thermodynamic and kinetic feasibility[38] further restrict the size of a subnetwork.

Currently, the node representation ignores stereoisomers and conformers, as the three-dimensional molecular structures are not taken into consideration. This could be partially remedied by using low-dimensional representations of conformer equilibria, which have been recently developed.[103] Finally, the expressiveness of RNs is limited by their reliance on simple graph-based rules such as bond breaks and bond formations. This representation, like other graph-based molecular models, excludes reactions of transition-metal compounds and intermolecular complexes, which cannot be expressed by composing discrete rules. Explicit dimensionality reduction methods might be of interest for treating these chemical systems.

## Acknowledgement

This work was supported in part by the National Science Foundation under Grant CHE-2227112 and DUE-1930546.

## Supporting Information Available

CHNO reaction network plots and tables of properties; generative model network tables of properties; comparison plots of CHNO comparison reaction networks and generative model networks; CHNO reaction network degree and local square clustering coefficient distributions; CHNO reaction network transformation rules distribution plots (PDF)

## References

(1) Jeong, H.; Tombor, B.; Albert, R.; Oltvai, Z. N.; Barabási, A.-L. The large-scale organization of metabolic networks. *Nature* **2000**, *407*, 651–654, DOI: 10.1038/35036627.

(2) Hatzimanikatis, V.; Li, C.; Ionita, J. A.; Henry, C. S.; Jankowski, M. D.; Broadbelt, L. J. Exploring the diversity of complex metabolic networks. *Bioinformatics* **2005**, *21*, 1603–1609, DOI: 10.1093/bioinformatics/bti213.

(3) Duarte, N. C.; Becker, S. A.; Jamshidi, N.; Thiele, I.; Mo, M. L.; Vo, T. D.; Srivas, R.; Palsson, B. Ø. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Nat. Acad. Sci.* **2007**, *104*, 1777–1782, DOI: 10.1073/pnas.0610772104.

(4) Montañez, R.; Medina, M. A.; Solé, R. V.; Rodríguez-Caso, C. When metabolism meets topology: Reconciling metabolite and reaction networks. *BioEssays* **2010**, *32*, 246–256, DOI: 10.1002/bies.200900145.

(5) Gao, C. W.; Allen, J. W.; Green, W. H.; West, R. H. Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms. *Comput. Phys. Commun.* **2016**, *203*, 212–225, DOI: 10.1016/j.cpc.2016.02.013.

(6) Vernuccio, S.; Broadbelt, L. J. Discerning complex reaction networks using automated generators. *AIChE J.* **2019**, *65*, e16663, DOI: 10.1002/aic.16663.

(7) Chodera, J. D.; Singhal, N.; Pande, V. S.; Dill, K. A.; Swope, W. C. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.* **2007**, *126*, 155101, DOI: 10.1063/1.2714538.

(8) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **2011**, *134*, 174105, DOI: 10.1063/1.3565032.

(9) Simm, G. N.; Vaucher, A. C.; Reiher, M. Exploration of Reaction Pathways and Chemical Transformation Networks. *J. Phys. Chem. A* **2019**, *123*, 385–399, DOI: 10.1021/acs.jpca.8b10007.

(10) Unsleber, J. P.; Reiher, M. The Exploration of Chemical Reaction Networks. *Annu. Rev. Phys. Chem.* **2020**, *71*, 121–142, DOI: 10.1146/annurev-physchem-071119-040123.

(11) Dewyer, A. L.; Argüelles, A. J.; Zimmerman, P. M. Methods for exploring reaction space in molecular systems. *WIREs Comput. Mol. Sci.* **2018**, *8*, e1354, DOI: 10.1002/wcms.1354.

(12) Wagner, A.; Fell, D. A. The small world inside large metabolic networks. *Proc. R. Soc. Lond. B. Biol. Sci.* **2001**, *268*, 1803–1810, DOI: 10.1098/rspb.2001.1711.

(13) Rappoport, D.; Galvin, C. J.; Zubarev, D. Y.; Aspuru-Guzik, A. Complex Chemi-

cal Reaction Networks from Heuristics-Aided Quantum Chemistry. *J. Chem. Theory Comput.* **2014**, *10*, 897–907, DOI: 10.1021/ct401004r.

(14) Rappoport, D. Reaction Networks and the Metric Structure of Chemical Space(s). *J. Phys. Chem. A* **2019**, *123*, 2610–2620, DOI: 10.1021/acs.jpca.9b00519.

(15) Song, C.; Havlin, S.; Makse, H. A. Self-similarity of complex networks. *Nature* **2005**, *433*, 392–395, DOI: 10.1038/nature03248.

(16) Song, C.; Gallos, L. K.; Havlin, S.; Makse, H. A. How to calculate the fractal dimension of a complex network: the box covering algorithm. *J. Stat. Mech.* **2007**, *2007*, P03006, DOI: 10.1088/1742-5468/2007/03/P03006.

(17) Gallos, L. K.; Song, C.; Makse, H. A. A review of fractality and self-similarity in complex networks. *Physica A* **2007**, *386*, 686–691, DOI: 10.1016/j.physa.2007.07.069.

(18) Clarkson, K. L. In *Nearest-Neighbor Methods in Learning and Vision*; Shakhnarovich, G., Darrell, T., Indyk, P., Eds.; MIT Press: Cambridge MA, 2005; pp 15–59.

(19) Andoni, A.; Indyk, P. In *Nearest Neighbors in High-dimensional Spaces*, 3rd ed.; Goodman, J. E., O'Rourke, J., Tóth, C. D., Eds.; CRC Press: Boca Raton FL, 2018; Chapter 43, pp 1135–1155, DOI: 10.1201/9781315119601.

(20) Rappoport, D. Statistics and Bias-Free Sampling of Reaction Mechanisms from Reaction Network Models. *J. Phys. Chem. A* **2023**, *127*, 5252–5263, DOI: 10.1021/acs.jpca.3c01430.

(21) Deza, M. M.; Deza, E. *Distances in Graph Theory*; Springer: Berlin Heidelberg, 2013; pp 249–278, DOI: 10.1007/978-3-642-30958-8_15.

(22) Dugundji, J.; Ugi, I. An algebraic model of constitutional chemistry as a basis for chemical computer programs. *Top. Curr. Chem.* **1973**, *39*, 19–64, DOI: 10.1007/bfb0051317.

(23) Kvasnička, V.; Pospíchal, J.; Baláž, V. Reaction and chemical distances and reaction graphs. *Theor. Chim. Acta* **1991**, *79*, 65–79, DOI: 10.1007/bf01113330.

(24) Erdős, P.; Rényi, A. On random graphs I. *Publ. Math. Debr.* **1959**, *6*, 290–297, DOI: 10.5486/PMD.1959.6.3-4.12.

(25) Watts, D. J.; Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **1998**, *393*, 440–442, DOI: 10.1038/30918.

(26) Newman, M. E. J. *Networks: An Introduction*, 1st ed.; Oxford University Press: Oxford, 2010; DOI: 10.1093/oso/9780198805090.001.0001.

(27) Cohen, R.; Havlin, S. *Complex Networks: Structure, Robustness and Function*; Cambridge University Press: Cambridge, 2010; DOI: 10.1017/CBO9780511780356.

(28) Estrada, E. *The Structure of Complex Networks: Theory and Applications*; Oxford University Press, 2011; DOI: 10.1093/acprof:oso/9780199591756.001.0001.

(29) Barabási, A.-L.; Albert, R. Emergence of Scaling in Random Networks. *Science* **1999**, *286*, 509–512, DOI: 10.1126/science.286.5439.509.

(30) Caldarelli, G. *Scale-Free Networks*; Oxford University Press: Oxford, 2007; DOI: 10.1093/acprof:oso/9780199211517.001.0001.

(31) Kleinberg, J. The small-world phenomenon: An algorithmic perspective. **2000**, 163—-170, DOI: 10.1145/335305.335325.

(32) Barrat, A.; Barthelemy, M.; Vespignani, A. *Dynamical processes on complex networks*; Cambridge University Press: Cambridge, 2008; DOI: 10.1017/CBO9780511791383.

(33) Adamic, L. A.; Lukose, R. M.; Puniyani, A. R.; Huberman, B. A. Search in Power-Law Networks. *Phys. Rev. E* **2001**, *64*, 046135, DOI: 10.1103/physreve.64.046135.

(34) Adamic, L. A.; Lukose, R. M.; Huberman, B. A. In *Handbook of Graphs and Networks*; Bornholdt, S., Schuster, H. G., Eds.; Wiley–VCH: Weinheim, 2003; Chapter 13, pp 295–317, DOI: 10.1002/3527602755.ch13.

(35) de Moura, A. P. S.; Motter, A. E.; Grebogi, C. Searching in small-world networks. *Phys. Rev. E 68*, 036106, DOI: 10.1103/physreve.68.036106.

(36) Weng, T.; Small, M.; Zhang, J.; Hui, P. Lévy Walk Navigation in Complex Networks: A Distinct Relation between Optimal Transport Exponent and Network Dimension. *Sci. Rep.* **2015**, *5*, 17309, DOI: 10.1038/srep17309.

(37) Rosenberg, E. *Fractal Dimensions of Networks*; Springer: Cham, 2020; DOI: 10.1007/978-3-030-43169-3.

(38) Rappoport, D.; Aspuru-Guzik, A. Predicting Feasible Organic Reaction Pathways Using Heuristically Aided Quantum Chemistry. *J. Chem. Theory Comput.* **2019**, *15*, 4099–4112, DOI: 10.1021/acs.jctc.9b00126.

(39) Habershon, S. Automated Prediction of Catalytic Mechanism and Rate Law Using Graph-Based Reaction Path Sampling. *J. Chem. Theory Comput.* **2016**, *12*, 1786–1798, DOI: 10.1021/acs.jctc.6b00005.

(40) Ismail, I.; Stuttaford-Fowler, H. B. V. A.; Ochan Ashok, C.; Robertson, C.; Habershon, S. Automatic Proposal of Multistep Reaction Mechanisms using a Graph-Driven Search. *J. Phys. Chem. A* **2019**, *123*, 3407–3417, DOI: 10.1021/acs.jpca.9b01014.

(41) Robertson, C.; Habershon, S. Fast screening of homogeneous catalysis mechanisms using graph-driven searches and approximate quantum chemistry. *Catal. Sci. Technol.* **2019**, *9*, 6357–6369, DOI: 10.1039/C9CY01997A.

(42) Jensen, J. H. A graph-based genetic algorithm and generative model/Monte Carlo tree search for the exploration of chemical space. *Chem. Sci.* **2019**, *10*, 3567–3572, DOI: 10.1039/C8SC05372C.

(43) Laio, A.; Gervasio, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Progr. Phys.* **2008**, *71*, 126601, DOI: 10.1088/0034-4885/71/12/126601.

(44) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *WIREs Comput. Mol. Sci.* **2011**, *1*, 826–843, DOI: 10.1002/wcms.31.

(45) Pracht, P.; Grimme, S. Calculation of absolute molecular entropies and heat capacities made simple. *Chem. Sci.* **2021**, *12*, 6551–6568, DOI: 10.1039/d1sc00621e.

(46) Sameera, W. M. C.; Maeda, S.; Morokuma, K. Computational Catalysis Using the Artificial Force Induced Reaction Method. *Acc. Chem. Res.* **2016**, *49*, 763–773, DOI: 10.1021/acs.accounts.6b00023.

(47) Maeda, S.; Harabuchi, Y.; Saita, K. In *Molecular Technology: Life Innovation*; Yamamoto, H., Kato, T., Eds.; Wiley-VCH: Weinheim, 2018; Vol. 1; pp 173–199, DOI: 10.1002/9783527823987.vol3_c8.

(48) Maeda, S.; Harabuchi, Y. Exploring paths of chemical transformations in molecular and periodic systems: An approach utilizing force. *WIREs Comput. Mol. Sci.* **2021**, *11*, e1538, DOI: 10.1002/wcms.1538.

(49) Steiner, M.; Reiher, M. A human-machine interface for automatic exploration of chemical reaction networks. *Nat. Commun.* **2024**, *15*, 3680, DOI: 10.1038/s41467-024-47997-9.

(50) Wang, L.-P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martínez, T. J. Discov-

ering chemistry with an ab initio nanoreactor. *Nat. Chem.* **2014**, *6*, 1044–1048, DOI: 10.1038/nchem.2099.

(51) Wang, L.-P.; McGibbon, R. T.; Pande, V. S.; Martinez, T. J. Automated Discovery and Refinement of Reactive Molecular Dynamics Pathways. *J. Chem. Theory Comput.* **2016**, *12*, 638–649, DOI: 10.1021/acs.jctc.5b00830.

(52) Ford, J.; Seritan, S.; Zhu, X.; Sakano, M. N.; Islam, M. M.; Strachan, A.; Martínez, T. J. Nitromethane Decomposition via Automated Reaction Discovery and an Ab Initio Corrected Kinetic Model. *J. Phys. Chem. A* **2021**, *125*, 1447–1460, DOI: 10.1021/acs.jpca.0c09168.

(53) Xu, R.; Meisner, J.; Chang, A. M.; Thompson, K. C.; Martínez, T. J. First principles reaction discovery: from the Schrodinger equation to experimental prediction for methane pyrolysis. *Chem. Sci.* **2023**, *14*, 7447–7464, DOI: 10.1039/D3SC01202F.

(54) Zhao, Q.; Savoie, B. M. Simultaneously improving reaction coverage and computational cost in automated reaction prediction tasks. *Nat. Comput. Sci.* **2021**, *1*, 479–490, DOI: 10.1038/s43588-021-00101-3.

(55) Zhao, Q.; Xu, Y.; Greeley, J.; Savoie, B. M. Deep reaction network exploration at a heterogeneous catalytic interface. *Nat. Commun.* **2022**, *13*, 4860, DOI: 10.1038/s41467-022-32514-7.

(56) Zhao, Q.; Savoie, B. M. Algorithmic Explorations of Unimolecular and Bimolecular Reaction Spaces. *Angew. Chem. Int. Ed.* **2022**, *61*, e202210693, DOI: 10.1002/anie.202210693.

(57) Zhao, Q.; Savoie, B. M. Deep reaction network exploration of glucose pyrolysis. *Proc. Nat. Acad. Sci.* **2023**, *120*, e2305884120, DOI: 10.1073/pnas.2305884120.

(58) van Deursen, R.; Reymond, J.-L. Chemical Space Travel. *ChemMedChem* **2007**, *2*, 636–640, DOI: 10.1002/cmdc.200700021.

(59) Kawai, K.; Nagata, N.; Takahashi, Y. De Novo Design of Drug-Like Molecules by a Fragment-Based Molecular Evolutionary Approach. *J. Chem. Inf. Model.* **2014**, *54*, 49–56, DOI: 10.1021/ci400418c.

(60) Hoksza, D.; Škoda, P.; Voršilák, M.; Svozil, D. Molpher: a software framework for systematic chemical space exploration. *J. Cheminf.* **2014**, *6*, 7, DOI: 10.1186/1758-2946-6-7.

(61) Schwaller, P.; Vaucher, A. C.; Laplaza, R.; Bunne, C.; Krause, A.; Corminboeuf, C.; Laino, T. Machine intelligence for chemical reaction space. *WIREs Comput. Mol. Sci.* **2022**, *12*, e1604, DOI: 10.1002/wcms.1604.

(62) Ohno, K.; Maeda, S. A scaled hypersphere search method for the topography of reaction pathways on the potential energy surface. *Chemical Physics Letters* **2004**, *384*, 277–282, DOI: https://doi.org/10.1016/j.cplett.2003.12.030.

(63) Hill, E. A. On a system of indexing chemical literature; adopted by the classification division of the US Patent Office. *J. Am. Chem. Soc.* **1900**, *22*, 478–494, DOI: 10.1021/ja02046a005.

(64) Grossman, R. B. *The Art of Writing Reasonable Organic Reaction Mechanisms*, 3rd ed.; Springer: Cham, 2019; DOI: 10.1007/978-3-030-28733-7.

(65) Levy, D. E. *Arrow-Pushing in Organic Chemistry. An Easy Approach to Understanding Reaction Mechanisms*; Wiley: Hoboken NJ, 2008; DOI: 10.1002/9780470378885.

(66) Rappoport, D.; Stulajter, M. colibri2 is a distributed computational engine for chemical exploration. https://bitbucket.org/rappoport/colibri2/, (accessed 2024-02-26).

(67) Weininger, D. SMARTS – A Language for Describing Molecular Patterns. https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html, (accessed 2024-02-26).

(68) Landrum, G. et al. RDKit: Open-Source Cheminformatics Software; ver. 2021.09.2. https://zenodo.org/record/7541264, (accessed 2024-02-26).

(69) Brandes, U.; Eiglsperger, M.; Herman, I.; Himsolt, M.; Marshall, M. S. GraphML Progress Report. Structural Layer Proposal. Graph Drawing 2001. Berlin Heidelberg, 2002; pp 501–512, DOI: 10.1007/3-540-45848-4_59.

(70) Staudt, C. L.; Sazonovs, A.; Meyerhenke, H. NetworKit: A tool suite for large-scale complex network analysis. *Netw. Sci.* **2016**, *4*, 508–530, DOI: 10.1017/nws.2016.20.

(71) Lind, P. G.; Gonzalez, M. C.; Herrmann, H. J. Cycles and clustering in bipartite networks. *Phys. Rev. E* **2005**, *72*, 056127, DOI: 10.1103/physreve.72.056127.

(72) Zhang, P.; Wang, J.; Li, X.; Li, M.; Di, Z.; Fan, Y. Clustering coefficient and community structure of bipartite networks. *Physica A* **2008**, *387*, 6869–6875, DOI: 10.1016/j.physa.2008.09.006.

(73) Kovacs, P. T.; Nagy, M.; Molontay, R. Comparative analysis of box-covering algorithms for fractal networks. *Appl. Netw. Sci.* **2021**, *6*, 73, DOI: 10.1007/s41109-021-00410-6.

(74) Wen, T.; Cheong, K. H. The fractal dimension of complex networks: A review. *Inf. Fusion* **2021**, *73*, 87–102, DOI: 10.1016/j.inffus.2021.02.001.

(75) Kim, J. S.; Goh, K.-I.; Kahng, B.; Kim, D. Fractality and self-similarity in scale-free networks. *New J. Phys.* **2007**, *9*, 177, DOI: 10.1088/1367-2630/9/6/177.

(76) Batagelj, V.; Brandes, U. Efficient generation of large random networks. *Phys. Rev. E* **2005**, *71*, 036113, DOI: 10.1103/PhysRevE.71.036113.

(77) Hagberg, A. A.; Schult, D. A.; Swart, P. J. Exploring network structure, dynamics, and function using NetworkX. Proceedings of the 7th Python in Science Conference (SciPy2008). Pasadena CA, 2008; pp 11–15.

(78) Stulajter, M. Reaction Networks Resemble Low-Dimensional Regular Lattices : Generation and Analysis Code. `https://github.com/mstulajter/Reaction-Network-Structure-Paper`, (accessed 2024-05-07).

(79) Jarque, C. M.; Bera, A. K. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters* **1980**, *6*, 255–259, DOI: `10.1016/0165-1765(80)90024-5`.

(80) Barabási, A.-L.; Albert, R. Emergence of Scaling in Random Networks. *Science* **1999**, *286*, 509–512, DOI: `10.1126/science.286.5439.509`.

(81) Newman, M. E. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* **2005**, *46*, 323–351, DOI: `10.1080/00107510500052444`.

(82) Cohen, R.; Havlin, S. Fractal dimensions of percolating networks. *Physica A* **2004**, *336*, 6–13, DOI: `10.1016/j.physa.2004.01.005`.

(83) Shao, J.; Buldyrev, S. V.; Cohen, R.; Kitsak, M.; Havlin, S.; Stanley, H. E. Fractal boundaries of complex networks. *EPL* **2008**, *84*, 48004, DOI: `10.1209/0295-5075/84/48004`.

(84) Albert, R.; Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **2002**, *74*, 47–97, DOI: `10.1103/RevModPhys.74.47`.

(85) Fronczak, A.; Fronczak, P.; Hołyst, J. A. Average path length in random networks. *Phys. Rev. E* **2004**, *70*, 056110, DOI: `10.1103/PhysRevE.70.056110`.

(86) Rudra, J. K.; Kozak, J. J. Spectral dimension of regular and fractal lattices. *Phys. Lett. A* **1990**, *151*, 429–432, DOI: `10.1016/0375-9601(90)90917-D`.

(87) Kim, J.; Goh, K.-I.; Kahng, B.; Kim, D. A box-covering algorithm for fractal scaling in scale-free networks. *Chaos* **2007**, *17*, DOI: 10.1063/1.2737827.

(88) Long, G.; Cai, X. The fractal dimensions of complex networks. *Chin. Phys. Lett.* **2009**, *26*, 088901, DOI: 10.1088/0256-307X/26/8/088901.

(89) Newman, M. E.; Watts, D. J. Scaling and percolation in the small-world network model. *Phys. Rev. E* **1999**, *60*, 7332–7342, DOI: 10.1103/PhysRevE.60.7332.

(90) Lochmann, A.; Requardt, M. An analysis of the transition zone between the various scaling regimes in the small-world model. *J. Stat. Phys.* **2006**, *122*, 255–278, DOI: 10.1007/s10955-005-8083-x.

(91) Yang, X. Fractals in small-world networks with time-delay. *Chaos, Solitons Fractals* **2002**, *13*, 215–219, DOI: 10.1016/S0960-0779(00)00265-4.

(92) Pavon-Dominguez, P.; Moreno-Pulido, S. Sandbox fixed-mass algorithm for multifractal unweighted complex networks. *Chaos, Solitons Fractals* **2022**, *156*, 111836, DOI: 10.1016/j.chaos.2022.111836.

(93) Shanker, O. Defining dimension of a complex network. *Mod. Phys. Lett. B* **2007**, *21*, 321–326, DOI: 10.1142/S0217984907012773.

(94) Chang, K.; Choi, B.; Yoon, S.-M.; Kim, K. Multifractal measures on small-world networks. *Fractals* **2006**, *14*, 119–123, DOI: 10.1142/S0218348X0600312X.

(95) Maier, B. F. Generalization of the small-world effect on a model approaching the Erdős–Rényi random graph. *Sci. Rep.* **2019**, *9*, 9268, DOI: 10.1038/s41598-019-45576-3.

(96) Katzav, E.; Biham, O.; Hartmann, A. K. Metric properties of subcritical Erdős–Rényi networks. *Phys. Rev. E* **2018**, *98*, 012301, DOI: 10.1103/physreve.98.012301.

(97) Zhang, D.; Yin, J.; Zhu, X.; Zhang, C. Network Representation Learning: A Survey. *IEEE Trans. Big Data* **2020**, *6*, 3–28, DOI: 10.1109/TBDATA.2018.2850013.

(98) Becker, O. M.; Karplus, M. The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *The Journal of Chemical Physics* **1997**, *106*, 1495–1517, DOI: 10.1063/1.473299.

(99) Murayama, B.; Kobayashi, M.; Aoki, M.; Ishibashi, S.; Saito, T.; Nakamura, T.; Teramoto, H.; Taketsugu, T. Characterizing Reaction Route Map of Realistic Molecular Reactions Based on Weight Rank Clique Filtration of Persistent Homology. *Journal of Chemical Theory and Computation* **2023**, *19*, 5007–5023, DOI: 10.1021/acs.jctc.2c01204.

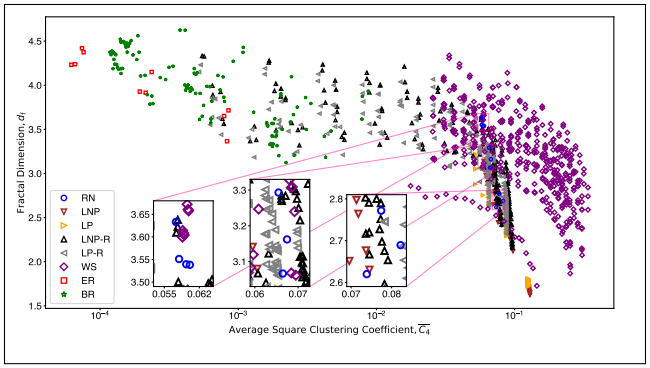
(100) Yi, H.-C.; You, Z.-H.; Huang, D.-S.; Kwoh, C. K. Graph representation learning in bioinformatics: trends, methods and applications. *Brief. Bioinform.* **2022**, *23*, bbab340, DOI: 10.1093/bib/bbab340.

(101) Rozenfeld, A. F.; Cohen, R.; Ben-Avraham, D.; Havlin, S. Scale-free networks on lattices. *Phys. Rev. Lett.* **2002**, *89*, 218701, DOI: 10.1103/PhysRevLett.89.218701.

(102) Kovács, B.; Palla, G. Model-independent embedding of directed networks into Euclidean and hyperbolic spaces. *Commun. Phys.* **2023**, *6*, 28, DOI: 10.1038/s42005-023-01143-x.

(103) Jing, B.; Corso, G.; Chang, J.; Barzilay, R.; Jaakkola, T. Torsional Diffusion for Molecular Conformer Generation. Advances in Neural Information Processing Systems. Red Hook NY, 2022; pp 24240–24253.

# TOC Graphic




The vexingly high dimensionality of potential energy surfaces of complex chemical reactions can be represented by low dimensional regular lattice graphs under a suitably chosen discretization.