Complex Reaction Networks in High Temperature Hydrocarbon Chemistry

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Physical Chemistry Chemical Physics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>CP-ART-10-2014-004736.R1</td>
</tr>
<tr>
<td>Article Type:</td>
<td>Paper</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>17-Oct-2014</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Mutlay, Ibrahim; Grafen Chemical Industries Co., Restrepo, Albeiro; Universidad de Antioquia UdeA,</td>
</tr>
</tbody>
</table>
Complex Reaction Networks in High Temperature Hydrocarbon Chemistry

IBRAHIM MUTLAY\(^1\), ALBEIRO RESTREPO\(^2\)

\(^1\)Grafen Chemical Industries Co., Ankara, Turkey
\(^2\)Instituto de Quimica, Universidad de Antioquia UdeA, Calle 70 No. 52-21, Medellin, Colombia

January 30, 2015

Abstract

Complex chemical reaction mechanisms of high temperature hydrocarbon decomposition are represented as networks and their underlying graph topology are analyzed as a dynamic system. As model reactants, 1,3-butadiene, acetylene, benzene, ethane, ethylene, methane, methyl isobutyl ketone (MIBK) and toluene are chosen in view of their importance for global environment, energy technologies as well as quantum chemical properties. Accurate kinetic mechanisms are computationally simulated and converted to bipartite graphs for incremental conversion steps of the main reactant. Topological analysis of the resulting temporal networks reveals novel features unknown to classical chemical kinetics theory. Time-dependent percolation behavior of the chemical reaction networks shows infinite order phase transition and a unique correlation between the percolation thresholds and electron distribution of the reactants. These observations are expected to find important applications in development of a new theoretical perspective to chemical reactions and technological processes e.g. inhibition of greenhouse gases, efficient utilization of fossil fuels, and large scale carbon nanomaterial production.

Keywords: Chemical reaction networks, graph theory, combustion, pyrolysis.

PACS: 82.20.-w, 82.40.-g, 82.20.Wt, 64.60.aq.

I. INTRODUCTION

Chemical reactions are one of the primitive processes ruling the entire physical world. From biological life, environmental cycles to industrial technologies, energy generation, the majority of the evolution of matter can be reduced to chemical reactions. The extremely complex and hierarchic nature of chemical reactions involving quantum mechanical scattering of molecules, statistical physics of particle ensembles and non-equilibrium thermodynamics attracts still many researchers suggesting vast variety of models. Quantum mechanical techniques [1, 2, 3] are able to explain chemical reactions in a great detail, however their application to real-life chemical kinetic mechanisms containing thousands of species and reactions is a computationally difficult (or unfeasible) task. Although system wide calculations, e.g. for mass flux by classical mass action kinetics are possible, they still require solution of large number of ordinary differential equations (ODEs).

Theory of complex networks brings a new understanding on the large scale organization of interacting entities. Abstraction of the systems as topological graphs by representing their units as vertices and interactions as edges, one acquires extensive information of the system organization by minimal computational load. Complex network approach has been successfully implemented for surprisingly different systems including social organizations, the World Wide Web, earthquakes, electrical power grids, protein interactions, citations to scientific articles and metabolic chemistry [4, 5, 6]. Chemical reactions with their native pathways have been always viewed as networks. In this context, a genuine research direction on kinetic stability of chemical reaction networks (CRNs) has been developed by Horn-Jackson [7], Clarke [8] and Feinberg [9]. This research provides invaluable results however they are limited to small mechanisms and traditional view of mass action law. Application of modern complex network theory on CRNs was first studied in the field of biochemistry, such as in [10, 11]. However, in spite of the subsequent success in biochemical CRNs [12], out of life sciences, there are still only a few works concerning for example organic chemistry [13, 14, 15] and the Diels-Alder reaction [16].

In this work, we study complex network features of the CRNs emerging from high temperature reactions of hydrocarbons. This reaction class is im-
mensely important not only because of technological and industrial processes which rely on fossil fuels, but also because of atmospheric and environmental phenomena, especially global warming, which is governed by hydrocarbon chemistry. For example, combustion of hydrocarbon fuels is the universal energy source of the modern world from power plants to internal combustion engine and pyrolytic reactions of hydrocarbons are the backbone of petroleum industry by converting oil to fractions (i.e. Eq. (1)) \( s_{ij} = 0 \) for absent species. A simple example can be found at Fig. 1a showing small portion of a giant reaction mechanism along with its stoichiometric matrix (Fig. 1b).

\[ \frac{d\mathbf{M}}{dt} = \mathbf{S} \cdot \nu(x) \]  

where \( \mathbf{M}(t) \) is species concentration and \( \nu(x) \) is flux vector. \( \mathbf{S} \) is stoichiometric matrix of entries \( s_{ij} \) corresponding to stoichiometric coefficient of the molecule \( M_i \) (ith row) at the reaction \( R_j \) (jth column). In the traditional notation of chemical reactions (i.e. Eq. (1)) \( s_{ij} \), however stoichiometric matrix definition requires that \( s_{ij} > 0 \) for produced (consumed) molecules while \( s_{ij} = 0 \) for absent species. A simple example can be found at Fig. 1a showing small portion of a giant reaction mechanism along with its stoichiometric matrix (Fig. 1b).
Figure 1: Fundamental definitions for bipartite representation of CRNs. a) initial reactions of CH₄ pyrolysis as a symbolic reaction mechanism. The species with asterisk are radicals. b) stoichiometric matrix of the reactions given in a). c) full bipartite CRN of CH₄ pyrolysis at 0.99 conversion. d) unweighted and undirected simple bipartite representation of CRNs which is in this case a subgraph sampled from c) and also is the exact bipartite graph of a) and b). e) unipartite species projection of the bipartite graph of d).

Stoichiometric matrix $S$ which is an inherent incidence matrix of directed hypergraphs lies at the heart of topology of chemical reaction networks. Directed hypergraphs are the intuitive network representation of chemical reactions by their multiple edges between two vertices. Unfortunately, considering lack of current understanding on hypergraphs, one would not expect much functionality. In order to handle chemical reaction networks efficiently we choose bipartite graphs because they are among the most favored and widely employed representations for reaction networks [31, 32]. Bipartite graphs delineate interactions of two separate vertex sets, i.e. participation of “agent” vertices in “event” vertices which are allowed to connect only opposite vertex sets not themselves. Bipartite graphs are actually König representation of hypergraphs [31] and there is a direct conversion between them by using stoichiometric matrix as a bipartite adjacency matrix, $B$ in the dimensions of agents × events [33]:

$$B_{k,l} = \begin{cases} b_{k,l} & \text{if agent participates in event} \\ 0 & \text{if not} \end{cases}$$

After this transformation, agents are species, events are reactions, and $b_{k,l}$ is the stoichiometric coefficients obviously (see Fig. 1d). The numerical value and sign of $b_{k,l}$ reflect weight and direction of edges in bipartite CRNs, respectively. Frequently, two types of the nodes are called top and bottom vertices and can be projected into two unipartite graphs, i.e. “species graph” and “reaction graph” respectively which can be analyzed just like ordinary graphs. Fig. 1e illustrates this operation; if the vertices of agents (events) in the original bipartite graph have links to an opposite vertex, then these agent (event) vertices connect inter se.

In this study, we work on undirected and unweighted simple bipartite graphs to minimize computational workload. This is a reasonable approach since the reactions here are reversible which do not need any direction in the graph edges. Unfortunately, we are also limited to available theoretical tools of the literature covering mainly simple graphs (for definition of the “simple graphs” see Supplementary Information). In the preparation of CRNs, all the nonzero entries in the stoichiometric matrices are reduced to 1 and edge directions are omitted. We also divide bipartite CRNs into their projections and focus only on species graphs because reaction projection affords excessive number of edges and uncertain vertex identity. However this issue is due to our computational configurations and should not be generalized. Holme and Huss already suggested that species projection is the most convenient graph to interpret the parent CRN [34, 35].
Figure 2: How CRNs grow in time? For CH₄ combustion and pyrolysis, time evolution (in reactant conversion dimension) of bipartite CRN graphs and adjacency matrices of their corresponding unipartite species graph projections. In bipartite CRN graphs, species and reaction vertices are salmon and light blue colored respectively. We emphasize species vertices with a nonlinear size proportionality to the degree (i.e. the total edge number connected to the vertex, for the definition see Supplementary Information). For the sake of visually enhanced representation of adjacency matrices we employ weighted unipartite species graphs only here: pixels of adjacency matrix plots are colored by continuously scaling weights of edges from blue ≡ 0 to red ≡ max(weight).

In Fig. 2, we depict time evolution of CRNs with the help of visualizations of bipartite CRNs and their accompanying unipartite species graph adjacency matrices. After total consumption of the reactant, we have an enormous kinetic mechanism with several “social” species of high degree represented by the larger vertices. In that example of CH₄ combustion mechanism has many high degree species while in pyrolysis there is only one prominent species. Note that, dominant blue color of the adjacency matrices in 0.99 conversion implies that those species graphs are sparse in accordance with the suggestion of [36] and sparsity increases as the reaction progresses.

A graph topology is defined by its metrics or measures which quantify the network structure (these concepts are extensively explained in Supplementary Information). We monitor the major graph metrics over the course of conversion for all the reactants in both combustion and pyrolysis regimes (see Fig. 3). Graph diameter, i.e. the maximum shortest path between vertices, determines the navigation or transport efficiency of networks [4, 5]. An intriguing point is that diameters of biochemical reaction networks in 43 different organisms were found to be universally ~3 [10, 38] smaller than our hydrocarbon mechanisms which could hint a difference in efficiency. In contrast to common temporal social networks [39], diameter of the species graphs does not fall, moreover increases slowly in time. The tendency of graphs to form cliques/clusters is measured by the clustering coefficient, especially transitivity i.e. the number ratio of triangles to triples [5, 38]:

\[ T_{3c} = \frac{3 \times n_A}{n_A} \]  \hspace{1cm} (3)

Triangle and triple are vertex sets of three at which vertices of the former connect simply like a triangle and the latter connect by ensuring at least one of the vertices connects to the rest [4, 5]. Since transitivity is not valid for bipartite graphs that cannot accommodate triangles Lind et al. preferred to use the rectangular vertex sets of four [40]:

\[ \langle C_4 \rangle = \frac{C_4(i)}{n_N} = \frac{n_{ij\square}}{n_i^n N} \]  \hspace{1cm} (4)

where \( n_{ij\square} \) is the number of squares of vertex \( i \) and \( n_i^n \) is the total number of possible squares of \( i \). Averaging over the nodes gives the global square clustering coefficient for bipartite graphs. We compare \( T_{3c} \) of projected species graphs with \( \langle C_4 \rangle \) of bipartite CRNs. Tracking through the conversion reveals declining \( \langle C_4 \rangle \) unlike constant \( T_{3c} \) as can be seen in Fig. 3. It is worth noting that there is little data on temporal profile of the clustering coefficients, for example [41] confirming our results. Assortative mixing is graph theory counterpart of the “like dissolves like” principle of chemistry and assortativity characterizes to what extent similar vertices connect. In assortative mixing, vertices select the other vertices to connect according to the similarity of their individual network measures [5]. It is possible to evaluate assortative mixing using any of the graph measures however in scientific literature the prevalent method is to calculate whether the
Figure 3: Time evolution of major graph metrics. All the curves are plotted for projected unipartite species graphs except square clustering coefficient which is for the original bipartite CRNs. Average vertex degree $\langle k \rangle$ defined on all vertex space is the simplest graph metric. Graph density is the ratio of total number to maximum number of edges. The density drops steadily by the conversion which is another reflection of sparsity increase [37]. Assortativity coefficient is calculated based on degree. See the main text and Supplementary Information for the detailed definitions and discussions.
vertices select other vertices to connect by considering their degree (i.e. degree based assortativity). In this study we also follow the same method to assess the time-dependent degree assortativity of reaction networks and the results are presented in Fig. 3 exhibiting surprising behavior: species graphs start as disassortative (negative values) but then turn out to be assortative (positive values) in late stages of the conversion, especially in the case of pyrolysis. This is highly distinctive in view of [42] reporting positive assortativity only for social networks and negative assortativity for biochemical networks. Modularity defines how communities, i.e. more densely connected vertex clusters are well separated [43]. Identification of communities requiring for further modularity assessment is accomplished by the WalkTrap algorithm [44] in this study. We find a generally increasing trend in Fig. 3 and the magnitude of modularity are significantly high, achieving nearly practical maximum [43] for 1,3-butadiene indicating strong communities in CRNs. We think that communities and modular topology correspond to reaction subclasses and have important applications in “lumping” techniques. Lumping is a model reduction method grouping compounds according to their physical/chemical properties in the form of pseudospecies which is able to save large amount of computational load [45]. In this regard, communities might help to find novel lumping techniques.

Figure 4: Log-log plot of the complementary cumulative distribution function (CCDF*) of the vertex degree distribution for species graphs of CH₄ combustion and pyrolysis along with their power-law analysis. Filled blue circles are the empirical degree distributions and vertical dashed lines are the average degrees. Both of them are shown with an increasing opacity order of the line color from the conversion of 0.2 to 0.99 in the interval of 0.2. Dash-point black curves are theoretical fit by the exponential degree distribution model to the final network (of the conversion 0.99). The insets are the time evolution of Vuong’s Test results. *Note: CCDF curve at x = k point represents the quantity of vertices with a degree of k or more.
Vertex degree distribution is one of the most informative topological features for complex networks (see Supplementary Information for underlying mathematics and the models). Commonly it is plotted as a cumulative distribution of vertex degrees since this offers visually more robust graphics [46]. What makes degree distribution so special is that its form is unique for each kind of network topology. In this context, it practically illuminates most of the network connectivity. Cumulative degree distributions of species graphs of our reaction mechanisms throughout the conversion window are presented in Fig. 4 and Fig. S1 (see Supplementary Information). After a qualitative examination of the shapes of the plots, one can easily deduct from their nonlinear decay profile in log-log axes that the best fitting theoretical probability distribution is the exponential distribution [47]. In order to prove this, we estimate the degree distribution fit of the species graphs for 0.99 conversion using the exponential distribution model [46]:

$$p(k) = \tau e^{k_{\min} - \gamma k}$$  \hspace{1cm} (5)

where $\tau$ is the rate parameter and $k_{\min}$ is the vertex degree cut-off value. Indeed, the fitting is acceptable for most of the reaction systems. However we need solid proof, hence we invoke Vuong’s Test which is a model selection test based on likelihood ratio and the Kullback-Leibler criteria. The Vuong’s Test of exponential distribution here is implemented against classical power-law model $p(k) \sim k^{-\gamma}$, $\gamma$ is the scaling parameter. Sign and magnitude of the Vuong’s Test results define which model is more suitable for the data. Negative and significantly low values of Vuong’s Test are the distinctive topological features for complex networks. In fact, exponential degree distribution in temporal networks is considerably prevalent observation in the literature [48, 49] even including protein [50] and online social networks [51]. The rationale is that in non-equilibrium and/or time-varying networks, preferential attachment which is the major process leading to power-law degree distribution is hindered [49, 52, 53]. In the absence of scale-free topology, i.e. the graphs with power-law degree distribution, we seek for the other common topology; the small-worlds which are characterized by small diameter and high clustering coefficient.

For the sake of objective evaluation, we follow the method of [54] suggesting a small-worldness index. Because of the unsuccessful calculations in some cases, we do not give the whole data however, generally speaking, the index indicates that all the species graphs here are small-worlds and their small-worldness increase by the conversion.

Centrality of a node is the quantitative measure of its influence and importance in the network. Intuition tells us that the more central a vertex is, the more prominent and functionally crucial a species is. This clearly serves as a solution to one of the most pursued questions in the theory of chemical kinetics, i.e. finding key species and pathways of reaction systems. Fig. 5 and Fig. S2 (see Supplementary Information) display bubble charts of fundamental centrality measures against gradually progressing conversion. Authority (hub) is an eigenvalue centrality which is determined by HITS algorithm calculating eigenvectors of the following matrices [55]:

$$M_{\text{Authority}} = A^T A \quad \text{and} \quad M_{\text{Hub}} = AA^T$$  \hspace{1cm} (6)

where $A$ is the adjacency matrix of, in our case, species graphs. Authority in conjunction with hub centrality, not surprisingly, delineates the vertices governing information transfer along the network. In order to see the performance of predictive capabilities of centrality measures, we compare them with chemical kinetics literature. The highest authority centrality in total burning of methane is detected for C2HO* ketenyl radical which is also known as a critical species in this reaction [56]. The ruling role of C2HO* in hydrocarbon combustion was already reported [57] particularly for acetylene proving our results. For combustion of all the reactants here except 1,3-butadiene, we observe that oxygen bounded primitive species are authorities. Monitoring authority in pyrolysis reactions unveils the importance of atomic hydrogen and methyl radical CH3* mentioned in the literature [58, 59]. Richter-Howard molecular growth approach to polynaromatic hydrocarbon (PAH) and soot formation [60] is valid for our benzene and toluene pyrolysis. Betweenness and closeness both of which are geometric centralities based on shortest paths are analyzed next. Betweenness is an indicator of junction vertices conveying main information “traffic” in the network. This centrality therefore has
very important functions in graphs, particularly matter/information flow capacity and robustness. Betweenness centrality of a vertex $i$ mathematically is as follows [6]:

$$C_B^i = \sum_{j \in V} \sum_{k \in V} \frac{\zeta_{jk}(i)}{\zeta_{jk}}$$  \hspace{1cm} (7)

where $\zeta_{jk}(i)$ is the number of shortest paths starting from vertex $j$, crossing $i$ and going to $k$ and $\zeta_{jk}$ is the total number of the possible paths between $j$ and $k$. CH$_3$* radical is a species of high betweenness in variety of mechanisms (see Fig. 5 and Fig. S2, Supplementary Information). We find acrolein (SPC$_{459}$) and C$_3$H$_2$** diradical as the key species of methane chemistry referred also by the literature [61, 62] but not studied in depth. Acrolein has a strong betweenness centrality in acetylene combustion. Maximum betweenness molecule hexadiyne (SPC$_{108}$) of 1,3-butadiene combustion at 0.99 conversion is an interesting molecule as a benzene precursor [63]. Ethylene chemistry is notably governed by radicals in view of betweenness, e.g. simple vinyl radical (C$_2$H$_3^*$) form of ethylene (C$_2$H$_4$). Betweenness analysis of the aromatic reactants gives inherent species: C$_6$H$_5^*$ phenyl radical for benzene and C$_7$H$_7^*$ benzyl radical for toluene as well as O$_2$ for their combustion. Betweenness centrality, unlike authority/hub is variable, roughly increasing by conversion. Closeness, in a sense, is the center point of the network because it tells us how far away from other graph regions a vertex is. Calculation of closeness for a vertex $i$ is quite straightforward [6]:

$$C_C^i = \frac{N - 1}{\sum_{j \neq i} d_{ij}}$$  \hspace{1cm} (8)

i.e. inverse average of geodesic distances from $i$ to $j$. The species with maximum closeness are almost same as that of the maximum betweenness.
However the time-dependent profile of closeness appears to diminish in contrast to betweenness. Investigation of the brief results above and detailed kinetic mechanisms which are provided in Supplementary Information reveals a practical, universal principle: a species is said to be central to its species graph if it participates in any of the first three reactions of its kinetic mechanism.

Network motifs is a notion of great interest believed to be abstract building blocks of networks each of which has specific real-life function. Motifs which are isomorphic subgraphs with high cardinality [6] have been a matter of debate to understand their correspondence relations with physical and especially biochemical phenomena, such as biochemical reactions and neural patterns in brain [38]. Although motifs do not have to be specific reactions or chemical entities in CRNs [64], it is tempting to explore possible relationships to the hierarchy of kinetic mechanisms. Toward that goal, we extract size-3 and size-4 motifs from species graphs of all reactants throughout the conversion range (see Fig. 6). Remarkably, we reveal that the most abundant motifs, M2, M5 and M8, denote the unipartite species graph projections of the following reaction system:

$$
X_i \xrightarrow{} X_{i+1} \quad (I)
$$
$$
X_{i+1} + X_{i+2} \xrightarrow{} X_{i+3} \quad (II)
$$
$$
X_{i+4} + X_{i+5} \xrightarrow{} X_{i+6} + X_{i+7} \quad (III)
$$

Bearing the rules of bipartite CRN building and unipartite projection in mind, reaction (II) produces M2 and combination of (I)-(II) M5. The last reaction (III) gives a closed triplet, i.e. M8. Our number distribution of motifs is unique in comparison to [65] acquiring completely different findings for several social, spatial and biological networks. The listed mechanism of (I), (II) and (III) is obviously among the fundamental reactions of chemical kinetics, hence this fact would have significant outcomes in the context of chemistry which will be discussed in Section 3.

III. PERCOLATION IN CRNs

Percolation, the evolution in properties of a thermodynamic system by the emergence of a large spanning cluster of the system units, is among the universal phase transitions. The recent decades have witnessed comprehensive research on this model finding a great variety of applications in phase transitions of superconductivity, magnetization, electrical/thermal conductivity of granular materials and composites, colloidal formations, epidemics, fracture, and even in socioeconomic organizations [66, 67, 68, 69]. Theoretical background of percolation phenomenon is inherently compatible with graphs. Connectivity of networks, under gradual node/edge addition or removal, experiences an abrupt transition at some point corresponding to the verge of appearance or disappearance of the giant connected component which is the largest cluster spreading on majority of the network. This critical onset is called percolation threshold at which global transport capacity and robustness, in other words, function and survival state of network is switched. Since a network growing (or shrinking by disconnections and attacks) beyond the percolation threshold would excel in its function (or disrupt and be annihilated), percolation plays a dramatic role for real-life CRNs by characterizing their chemical yield and efficiency.

In order to interpret percolation phenomenon for CRNs successfully, one would need to understand the analogy between abstract CRNs and real chemical reaction systems. Chemical reactions are the alteration of electronic and nuclear wavefunctions via different processes, particularly atomic redistribution of molecules. In another view adopted by biochemistry society referring holistic principles, CRNs are supposed to be mass flux networks. Clearly, both of the approaches point out that practical productivity of a chemical reaction mechanism relies on (information or mass) transport capability of its CRN and of course its percolation transition. Fundamental laws governing the percolation differentiate for static and dynamic networks. Static networks are in the class of equilibrium percolation obeying finite size scaling and power law $X \sim (p - p_c)^\omega$ at $N \to \infty$ where $X$ is any global network feature, $p$ is the fraction of nodes/edges added or removed and $p_c$ is the percolation threshold. Unfortunately, powerful techniques of finite size scaling fail for dynamic networks, e.g. CRNs and there is little work on non-equilibrium percolation. Callaway et al. [48] studied dynamic networks by adding a single vertex at each time step and
found that non-equilibrium percolation is in the class of infinite order phase transition. They also successfully approximated the percolative growth of dynamic graph by the relation below:

\[ S(\delta) \sim \delta^{\alpha(\delta_0 - \delta)} \beta \]  

as \( \delta \to \delta_c \), where \( \delta \) is edge formation probability, \( \delta_c \) percolation threshold, \( \alpha \) and \( \beta \) are critical exponents [48]. Percolation in growing protein networks evolved by mutations [70] and synthetically generated time-varying graphs [71] were also identified as infinite order.

We study percolation analysis of our species graphs projected from bipartite CRNs in Fig. 7 displaying peculiar percolation behavior. The giant connected component, GCC emerges suddenly just at the beginning of conversion. This is surprising because in classical percolation theory there is no GCC before percolation threshold. Moreover, size of the GCC exactly equals to network size throughout the whole range of conversion in contrast to classical GCC that can cover a graph only partially. Nevertheless, percolation trends of our species graphs are quite similar to the results of the conventional systems and experiences a critical point as a growth burst in network size. Although we are not able to simulate conversion \( \xi \) at \( \xi \to 0 \) limit, this unusual percolation is prevalent for all the reactants in the range of \( \xi > 0.01 \). In fact, CRNs are not the only systems showing such an always percolated phase.
Figure 7: Percolation analysis of species graphs of combustion and pyrolysis CRNs. The plots are GCC size (a)-(b), largest eigenvalue (c)-(d) and average path length (e)-(f) against conversion. Data points are empirical data whilst solid curves are theoretical fit based on Eq. (10). First plots (a, c, e) are for combustion, second plots (b, d, f) are for pyrolysis. Intensity of the percolative graph measures (y-axes) for all graphics are normalized over their maximum values. In order to estimate $\alpha$ and $\xi_c$ of Eq. (10) which is needed to plot $S(\xi)$ time-dependent GCC size curves of (a)-(b), we fit Eq. (10) by a linearized log-log model for each reactant separately. Right column $\xi_c$ is the percolation thresholds for combustion (orange box, upper) and pyrolysis (gray box, lower) of the corresponding reactants.
whose size equals to full network [75]. Facebook, as
Fig. 7a-b). In Eq. (10), S(ξ) is the size of
the reaction system is thoroughly evolved as
previous Section comes into play at this point. As
self-similar networks [73] and graphs from Achliop-
tas process [74] all exhibit a percolation threshold
at the stage zero of their evolution. Similarly, GCCs
in the size of their parent networks require special
conditions to grow. For instance, in particle con-
tact networks of granular materials, gravity ensures
close packing of the particles leading to a GCC
whose size equals to full network [75]. Facebook,
thanks to its extremely connected social web, is an-
other interesting example at which GCC to network
size ratio achieves 0.999 [76].

Let us recall our main goal in this paper, we try
to comprehend how and why percolation of CRNs
belongs to that strange phase transition class. The
graph motif distribution we have discussed in the
previous Section comes into play at this point. As
one can see from the tabulated reaction mechan-
isms in Supplementary Information, almost all
of the reactions are in the forms of \( X + Y \rightleftharpoons Z \) or
\( X + Y \rightleftharpoons W + Z \) which also correspond to the
primary motifs M2 and M8, respectively. Furthermore,
in those mechanisms, every new reaction added
in each step contains species from previous steps.
Intuitively, this process is supposed to be as follows:
by increasing \( \xi \), in each step, species graph grows
by M2 and M8 motifs without any isolated vertex
or cluster considering absence of quasi-reactions
e.g. \( 0 \rightleftharpoons X \), and the fact that each new reaction is
tied to previous ones by their reactants or products.
Thus, the reaction system is thoroughly evolved as
a single unique cluster, i.e. the GCC. The growth
mechanism of CRNs depicted above clarifies the
strange percolation trend observed here. Among
all other complex networks, it is fairly natural to
assume that such kind of percolation behavior pertains
to CRNs which are able to ensure a strongly
connected growth based on non-isolated triplets.
Dynamic percolation of CRNs is investigated via
infinite order phase transition model of [48]. We
utilize asymptotic form of Eq. (9) by a slight abuse
of notation:

\[
S(\xi) \sim \xi^a / \sqrt{(\xi - \xi_c)} \tag{10}
\]
to fit the time-dependent empirical GCC size data
(see Fig. 7a-b). In Eq. (10), \( S(\xi) \) is the size of
giant connected component, \( \xi \) is the conversion,
\( \xi_c \) is the conversion at the percolation threshold
and \( a \) is the critical exponent. Agreement between
the fit and data points indicates that infinite order
phase transition is also valid for CRNs. Note that,
although original GCC size data differentiates in a
broad range, after normalization we observe a con-
sistent and similar trend for all the reactants which
is discriminated only by percolation thresholds.

Strong change of spreading capacity and navigation
efficiency related to global connectivity occur during
percolation. In the case of positive direction for
system evolution, fine synchronization of spreading
and navigation at percolation provides the best
way of boosting network functionality. However,
for the situations where critical point coherence is
lost, the network is adversely affected by percola-
tion transition. Spreading capacity of network, like
mass flux of reactions or information transport in
our CRNs, is controlled by the largest eigenvalue
\( \lambda_{\text{max}} \) [77]. \( \lambda_{\text{max}} \) undergoes a transition, \( p_{c}^{\xi} \) leading
to rise in spreading capacity that is similar to usual
percolation. In this regard, relative position of \( p_{c}^{\xi} \) is
the critical parameter affecting spreading efficiency.
\( p_{c}^{\xi} < p_c \) is highly unlikely, however if \( p_{c}^{\xi} \gg p_c \) then one can have a GCC but in fact the network is
not capable of spreading. Hence in order to com-
pare spreading efficiency of CRNs we propose the
measure in terms of conversion below:

\[
e_{\lambda} = \frac{\xi_c}{\xi_{c}^{\xi}} \tag{11}
\]

In the case of navigation efficiency \( e_{\lambda} \), we simply
calculate inverse of the average shortest path length:

\[
e_{\ell} = \frac{1}{\langle l \rangle} \tag{12}
\]

Spontaneous increase of \( \langle l \rangle \) in the vicinity of perco-
lation threshold hindering network transport and
navigability has been already reported by Lopez
et al. [78]. Recall that CRNs are abstract grids for
wave function routing on molecules, navigation ef-
ciciency is a vital factor determining the yield and
productivity of chemical reactions [79]. We apply
spreading and navigation efficiency notions to our
CRNs. Almost all the reaction systems studied here
are highly efficient spreaders with \( e_{\lambda} \approx 1 \) (see Fig.
7). On the other hand, average path length does not
have a clearly tractable trend. Navigation efficiency
of ethane as well as of acetylene, ethylene, and
methane, rapidly drops after percolation threshold.
Even worse, navigation efficiency for the majority
of reactants does not recover to its original high
level in contrast to [78] in which navigability falls
only at percolation threshold locally.
Percolation threshold of physical systems e.g. elec-
trically conductive composites [80] or granular
packings [81] is known to have a strong corre-
lation to the geometry of system units. This universal
correlation is
\[
p_c \propto \frac{1}{\phi}
\]  
(13)
where \( \phi \) is the aspect ratio, the size ratio of two
major topological axes of the geometrical objects.
Eq. (13) is practically very important by enabling
estimation of percolation threshold via only sim-
ple geometrical data of the system building-blocks.
Although it is proposed geometrically, in order to
benefit from its utility, we try to extend it to our
CRNs by establishing an analogy. The major feature
governing aspect ratio influence on percolation is
the coordination number or vertex degree in graph
terms. More anisotropic granular particles are able
to build system-wide contact paths using less parti-
cles. Thus, at percolation, there are smaller num-
er of particles (smaller coordination number) giving
reduced percolation threshold. In chemical reac-
tions, system unit is molecules and connectivity
is expected as the function of reactivity. However
one can clearly discern from basic chemistry that
geometrical aspect ratio of molecules or any other
pure geometrical characteristic is not sufficient to
estimate their reactivity. Quantum mechanical elec-
tronic and nuclear structure of molecules is the ac-
tual parameter defining chemical reactivity which
is analyzed here in the framework of QTAIM. The
critical points of the Laplacian of the electron den-
sities (the LCPs) have been successfully used in a
wide variety of chemical reactivity situations [28],
including the correct prediction of potential sites
for nucleophilic attack [82], the structures of hydro-
gen bond complexes [83], the preferred attacking
sites for electrophilic substitution in substituted aro-
matic compounds [84], and the relative susceptibil-
ity of activated double bonds to Michael addition
[85]. The collected evidence of the ability of the
Laplacian of the electron density to correctly pre-
dict chemical interactions lead to the formulation of
the Laplacian complementary principle [86], which
states that "A maximum in the valence shell of an
atom in one molecule combines with a minimum
in \( L(r) \) in the valence shell of another atom in another
molecule". In this work, we use the properties of
the LCPs of the electron density (those for which
\( \nabla L(r) = 0 \)) to gain information about chemical
reactivity. In this context, we suggest that LCPs
for the reacting molecules is the equivalent of ge-
ometrical aspect ratio for the particles in granular
percolation. To be precise, the aspect ratio analogy
is the ratio of local Laplacian maxima to minima:
\[
\frac{|L(r)_{\text{max}}|}{|L(r)_{\text{min}}|} \equiv \phi
\]  
(14)
We illustrate this analogy in Fig. 8 indicating a
correlation highly similar to the literature [80, 87].
Simple inverse proportionality like Eq. (13) can be
clearly seen. Following [80], we attempt to fit the
empirical data using Pade approximation but the
results are not convincing. Nevertheless, it sup-
ports the evidence of Eq. 13 and 14 which, to our
knowledge, is the first demonstration of a relation-
ship between the percolation of abstract CRNs and
real-life ab initio principles. Percolation threshold
for CRNs, like global reactivity measures, could be
used to evaluate activity of overall reaction mech-
nisms. Besides, mathematical approach of this
study is far more unsophisticated than common
reactivity estimation techniques which will help
potential technological applications.
1,3-Butadiene
Acetylene
Benzene
Ethylene
Methane
MIBK
Toluene
Ethane

Combustion:
\[ d_1 = 2, \ d_2 = 4 \]
Pyrolysis:
\[ d_1 = 5, \ d_2 = 7 \]

IV. Discussion

Abstraction of CRNs and analysis of their topology via methods of complex network theory provides rich data and could change the view on chemical kinetics which has been already proven in biochemistry. However, a more complementary discussion to assert the graph theoretical results here needs to evoke classical chemical kinetics. Our reaction mechanisms generated artificially by the RMG calculating also their mass fluxes by Eq. (2) and Arrhenius parameters are given in Supplementary Information. Determination of the key elementary steps of a reaction mechanism is a significant part of kinetics research. Considering practical yield, we assume the chemical reactions of maximum flux are the major pathways which are listed in Table 1 for our results. We compare those reactions with the species found in centrality analysis (see Fig. 5 and Fig. S2). It is observed that central species predicted by graph theory are in excellent agreement with the major pathways. These topologically identified species are same as the ones participating the key elementary reactions which is proof of how the centrality is a powerful indicator. Some of the difference in results can be easily attributed to the further reaction steps for example C\(_2\)H\(_4\) and C\(_4\)H\(_3\)^* of methane pyrolysis can be emanated from CH\(_3\)^* by consecutive reactions. Similarly conversion between C\(_2\)H\(_2\)O and C\(_2\)HO^* of acetylene is possible with a H\(_2\) reaction. In an attempt to support our findings, we also test the Transition State Theory calculations (see Fig. S3, Supplementary Information) however extremely high temperature, unimolecular and free radical dominant reaction mechanisms in our study prevent us to use that technique which could fail for these situations.

In addition to the theoretical importance, high temperature hydrocarbon chemistry has substantial implications for the Earth and humankind. 70% of global electricity generation still relies on thermal decomposition of fossil fuels\(^2\). Furthermore, transportation, commodity chemicals and even many advanced materials are dependent on petroleum. Critical carbon allotropes fullerene [88], carbon nanotube [89] and graphene [90] as well as pyrolytic graphite [91] are produced by the high temperature reactions of hydrocarbons. Unfortunately, those chemical processes pose a great danger to the environment and human health. Atmospheric carbon black (or soot) and greenhouse gases which are the two major causes of the global warming, and PAHs which are highly mutagenic and carcinogenic to biological life originate from anthropogenic hydrocarbon decomposition [92, 93]. In this regard, design and control of hydrocarbon CRNs is a two sided problem: elimination of the hazardous effects require inhibition and disruption of the reaction networks whilst robust and efficient networks are vital for the processes of power generation or material production. Evolution of hydrocarbon reactants to PAHs and soot consists of nearly countless and complex pathways which is still a controversial research topic. The leading models in that field are proposed by Howard [60, 94, 95] and Frenklach [96, 97, 98] independently. According to the general picture developed in those models [60, 96] high temperature hydrocarbon reactions start with the addition of C\(_2\) species, particularly acetylene C\(_2\)H\(_2\),

\(^2\)The World Bank and U.S. Energy Information Administration data.
Table 1: Chemical reactions with maximal flux generated in this study. If fluxes of the 1st and 2nd maximum flux reactions for a reactant are nearly equal, we present both of them. Pure H-OH reactions without carbon species are omitted. $A$, $\beta$ and $E_a$ are respectively the pre-exponential factor, the temperature exponent and the activation energy (kcal/mol) of the modified Arrhenius equation: $k = A T^{\beta} \exp \left( \frac{-E_a}{R T} \right)$ with the gas constant $R$ and the reaction temperature $T_{gas}$.

<table>
<thead>
<tr>
<th>Reactants</th>
<th>Reactions</th>
<th>$A$</th>
<th>$\beta$</th>
<th>$E_a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C_4H_6$</td>
<td>$C_2H_5^* + CH_3^* = C_4H_6$</td>
<td>$7.230 \times 10^{13}$</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$C_4H_6$</td>
<td>$C_2H_4^* + OH = C_4H_6O$</td>
<td>$1.520 \times 10^{8}$</td>
<td>1.70</td>
<td>1.00</td>
</tr>
<tr>
<td>$C_6H_6$</td>
<td>$C_2H_5^* + O_2 = C_2H_5O_2^*$</td>
<td>$6.030 \times 10^{12}$</td>
<td>0.00</td>
<td>0.32</td>
</tr>
<tr>
<td>$C_2H_4$</td>
<td>$C_2H_5 + H = C_2H_4^*$</td>
<td>$2.200 \times 10^{14}$</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$C_2H_4$</td>
<td>$C_2H_4 + H = C_2H_3^*$</td>
<td>$5.500 \times 10^{12}$</td>
<td>0.00</td>
<td>2.42</td>
</tr>
<tr>
<td>$CH_4$</td>
<td>$CH_4 + H = CH_2^* + H_2$</td>
<td>$1.224 \times 10^{8}$</td>
<td>1.87</td>
<td>10.59</td>
</tr>
<tr>
<td>$C_2H_4$</td>
<td>$C_2H_4 + H = C_2H_3^*$</td>
<td>$5.500 \times 10^{12}$</td>
<td>0.00</td>
<td>2.42</td>
</tr>
<tr>
<td>$C_4H_6O_2$</td>
<td>$CO** + CH_3^* = C_4H_3O^*$</td>
<td>$3.060 \times 10^{6}$</td>
<td>1.89</td>
<td>4.82</td>
</tr>
<tr>
<td>$C_7H_8$</td>
<td>$C_6H_5^* + O_2 = C_2H_5O_2^*$</td>
<td>$6.030 \times 10^{12}$</td>
<td>0.00</td>
<td>0.32</td>
</tr>
<tr>
<td>$C_4H_6$</td>
<td>$C_2H_5^* + CH_3^* = C_4H_6$</td>
<td>$7.230 \times 10^{13}$</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$C_6H_6$</td>
<td>$C_2H_4^* = C_4H_6$</td>
<td>$1.132 \times 10^{11}$</td>
<td>1.00</td>
<td>9.50</td>
</tr>
<tr>
<td>$C_2H_2$</td>
<td>$C_2H^* + H_2 = C_2H_2 + H$</td>
<td>$1.080 \times 10^{13}$</td>
<td>0.00</td>
<td>2.17</td>
</tr>
<tr>
<td>$C_2H_2$</td>
<td>$C_2H_2 + C_2H^* = C_2H_4^*$</td>
<td>$1.000 \times 10^{13}$</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$C_6H_6$</td>
<td>$C_2H_4^* + H = C_2H_2 + H$</td>
<td>$2.200 \times 10^{14}$</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$C_2H_4$</td>
<td>$C_2H_4 + C_2H_2^* = C_4H_2H_10$</td>
<td>$5.700 \times 10^{12}$</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$CH_4$</td>
<td>$CH_4 + H = C_2H_3^*$</td>
<td>$5.500 \times 10^{12}$</td>
<td>0.00</td>
<td>2.42</td>
</tr>
<tr>
<td>$C_2H_4$</td>
<td>$C_2H_4 + H = C_2H_3^*$</td>
<td>$5.500 \times 10^{12}$</td>
<td>0.00</td>
<td>2.42</td>
</tr>
<tr>
<td>$C_2H_4$</td>
<td>$C_2H_4 + C_2H_2^* = C_4H_3O^*$</td>
<td>$3.060 \times 10^{6}$</td>
<td>1.89</td>
<td>4.82</td>
</tr>
<tr>
<td>$C_7H_8$</td>
<td>$C_6H_5^* + CH_3^* = C_7H_5^*$</td>
<td>$1.380 \times 10^{13}$</td>
<td>0.00</td>
<td>0.05</td>
</tr>
</tbody>
</table>

To the small molecules ($C_1$-$C_4$ species). First of the critical points in that reaction scheme is the formation of initial aromatic rings which is followed by the growth to PAHs via aliphatic growth by addition of $C_2$ species and/or direct aromatic growth. The final stage is the physical agglomeration of the giant PAHs and evolution of these PAH clusters to larger soot particles. Although there are various views on the formation of first aromatic rings, the following reactions are accepted as the key pathways [60, 96]:

$$C_2H_2 + C_2H_3^* := 1,3-C_4H_5^* \quad (1)$$
$$1,3-C_4H_5^* + C_2H_2 = C_6H_6 + H \quad (2)$$
$$n-C_4H_5^* + C_2H_2 = C_6H_5^* \quad (3)$$
$$C_3H_3^* + C_2H_2 = c-C_3H_5 \quad (4)$$

Since $C_4H_5^*$ and $C_2H_2$ are abundant species in our mechanisms it is thought that reaction (3) could be a crucial pathway in our conditions. Along with $C_2H_2$ we detect many $C_2$ species supporting $C_2$-based evolution theory. $C_5$ also play an essential role in PAH growth formulated by the HACA (H-abstraction-$C_2$H-addition) reactions [96]:

$$A_i + H = A_i^* + H_2 \quad (5)$$
$$A_i^* + C_2H_2 = PAHs \quad (6)$$

where $A_i$ is an aromatic molecule and $A_i^*$ is the aromatic radical of $A_i$. Prominence of this type of reactions prompts us to investigate the topological routes involving $C_2$ species in our CRNs. We measure the average shortest path between the main reactant and $C_2$ molecules in species graphs (see Fig. 9). Profiles and average levels of $d_{s1} \rightarrow C_{2j}$ are highly consistent and similar for combustion and pyrolysis. $d_{s1} \rightarrow C_{2j}$ levels of each reactant surprisingly follow a rule of hydrocarbon chemistry: for the order of reactants of alkanes, mono-olefins, di-olefins, benzenes and naphthalenes soot yielding capacity...
increases [60]. This refers to a correlation, the CRNs with higher \(d_{S_1 \rightarrow C_2,i} \) tend to produce more soot. The smallest \(d_{S_1 \rightarrow C_2,i} \) for MIBK might be related to its molecular structure which is the only ketone in this work.

Theory of CRNs in combination with chemical kinetics knowledge could help the design of reaction systems satisfying both of the targets mentioned above. Quenching techniques like carbon sequestration is believed to be used in reduction of the global warming. However enormous amount of the atmospheric matter to be stored is a serious technological barrier. At that point, the centrality analysis of graph theory might determine the low concentration key species and reactions which can be more easily handled. In the case of positive targets like chemical synthesis, CRNs can be optimized to improve the production efficiency and enhance the robustness against disturbances in the reaction conditions which is also the future research goal of this study.

**Acknowledgement**

We are grateful to Prof. W.H. Green and his research team from MIT, USA for their essential support on the RMG software. The authors also wish to thank the referees for their invaluable comments and recommendations enabling pretty refined final form of this study.

**Supplementary Content**

An introductory text on the foundations of graph theory, detailed description of computational techniques and methods as well as complete list of reaction mechanisms are available in Supplementary Information.

**References**


[89] Vincent Jourdain and Christophe Bichara. Current understanding of the growth of carbon


Bipartite Chemical Reaction Graph

CH₄ - Combustion
50% Conversion