

Article

Graphical Classification in Multi-Centrality-Index Diagrams for Complex Chemical Networks

Yasutaka Mizui, Tetsuya Kojima, Shigeyuki Miyagi and Osamu Sakai * 

Electronic Systems Engineering, The University of Shiga Prefecture, 2500 Hassakacho, Hikone, Shiga 522-8533, Japan; oh23ymizui@ec.usp.ac.jp (Y.M.); oh23tkojima@ec.usp.ac.jp (T.K.); miyagi.s@e.usp.ac.jp (S.M.)

* Correspondence: sakai.o@e.usp.ac.jp; Tel.: +81-749-28-8382

Received: 22 October 2017; Accepted: 5 December 2017; Published: 9 December 2017

Abstract: Various sizes of chemical reaction network exist, from small graphs of linear networks with several inorganic species to huge complex networks composed of protein reactions or metabolic systems. Huge complex networks of organic substrates have been well studied using statistical properties such as degree distributions. However, when the size is relatively small, statistical data suffers from significant errors coming from irregular effects by species, and a macroscopic analysis is frequently unsuccessful. In this study, we demonstrate a graphical classification method for chemical networks that contain tens of species. Betweenness and closeness centrality indices of a graph can create a two-dimensional diagram with information of node distribution for a complex chemical network. This diagram successfully reveals systematic sharing of roles among species as a semi-statistical property in chemical reactions, and distinguishes it from the ones in random networks, which has no functional node distributions. This analytical approach is applicable for rapid and approximate understanding of complex chemical network systems such as plasma-enhanced reactions as well as visualization and classification of other graphs.

Keywords: chemical reaction network; centrality index; statistical analysis; random graph

1. Introduction

Graph theory provides for us a graphical approach to a system containing various elements with connections between them [1,2]. Chemical reaction networks are one such system, and small and simple reaction systems are visualized in linear or small graphs with sufficient understanding of reaction procedures [3–5]. On the other hand, the networks of protein and metabolic systems in a biological cell are quite complicated due to their numbers of nodes (at least, more than 1000), so that not only visualization as a graph but also statistical properties such as degree distributions are representative for characterizing their complexity [6,7].

Recently, we performed graph visualization for plasma-enhanced chemical reactions [8,9]. In low-temperature reactive plasma, high-energy electrons trigger a number of simultaneous dissociations of mother molecules, and its chemistry is more complex than other chemical systems in artificial environments for chemical plants [10–12]. After definitions of a node (for one species) and an edge (for each reaction) for a display in a graph, centrality indices of nodes derived from the graph work as representatives of chemical roles in the system, such as agents, intermediates, and products. However, except for such microscopic points of view, approaches have not been accomplished for describing macroscopic properties of graphs for chemical reaction networks that contain several tens of species.

Such relatively small-sized chemical network systems create rich outputs despite limited numbers of nodes and edges in a graph. For instance, information processing in biochemical reactions revealed collective behaviors that can be interpreted using interactions among analogical spins, leading to

similar features to electronic information processing or mechanical systems [13,14]. Another example of graphical network approaches for medium-sized chemical complexity was on numerical calculations of rate equations in plasma-enhanced chemical reactions, and the calculated results were visualized in reaction pathways in a graph to summarize complicated time evolutions of densities of species [15]. In comparison with the previous achievements based on numerical finite-difference methods [10–12,15], direct visualization of reactions based on graph theory is applied here, and we focus on graphical classification of nodes or species on statistical aspects that are missing in our previous studies [8,9].

In this study, for such systems as medium-sized chemical reaction networks, which are neither so small as several numbers of reactions nor too large with more than 1000 species, we demonstrate suitable macroscopic measures by graphical diagrams based on multi-centrality indices. Using such diagrams, semi-statistical properties with confirmation of systematic and global structures in the corresponding system can be deduced, even if the total number of nodes is limited to less than 100, as well as identifications of roles in each species, such as agents, intermediates and products. This method of classification of species in macroscopic points of view provides us with insight and understandings about predictions of global properties of chemical reactions for approximate designs of upcoming chemical reactors and rapid selection of mother chemicals for products when accurate computer aided designs are not available. In Section 2, using reactions in silane and methane plasmas, we show the diagrams with axes of betweenness and closeness centrality indices. Using such a diagram, we can understand both macroscopic and microscopic properties of graphs with tens of nodes. In Section 3, we compare these two examples with random graphs, and discuss the validities of this graphical characterization.

2. Analysis of Reaction Networks in Plasma Chemistry

As examples, we use two plasma-enhanced chemical networks reported in [10,11]. We performed some analyses of methane plasma [8] and silane plasma [9] using some centrality indices, but they did not include macroscopic measures of complex chemical networks. Here, we proceed to study them using the same reaction systems to obtain their macroscopic and microscopic properties simultaneously. Table A1 in Appendix A shows active species in silane plasma with temporary indicator numbers, where we use reaction sets in [11]. There are a wide variety of species that originate initially from two species: SiH₄ and Ar. The number of species or nodes in the corresponding graph is 58, and that of reactions or edges is 222. The reason why we can observe such rich diversity is based on high-energy electrons whose energy spreads up to 20 eV with electron temperature of 1–5 eV; kinetic energy with 1–20 eV induces most of reactions of decomposition, dissociation and ionization [16]. Almost all reactions are bi-molecule, and we treat all reactions as irreversible ones.

To convert a chemical reaction into a graph, when we handle the following reaction,



and use the names of species in reactions as node indicators, we set directed edges from node *A* to node *D*, from node *A* to node *E*, from node *B* to node *D* and from node *B* to node *E*, as shown in Figure 1a. When one performs graph representation for an underlying system of interest, elements of the system becomes nodes and interactions between elements are displayed as edges. For chemical reactions, as interpreted in [3], species becomes nodes, and not a simple co-existence but an agent-product relation is suitable for a directed edge, leading to a form of representation in Figure 1a. In general chemical reactions for inorganic molecules, since roles of agents, intermediates and products are fairly clear with energy consumption as driving forces, regular graphs or graphs with frequent cycles are less important, and monotone graphs or trees with several nodes are sufficient in many cases [3–5]. However, in reactions enhanced by energetic electrons in plasma, cycles appear with their various sizes with increasing values of clustering coefficients [8], which makes a role of each species more complicated.

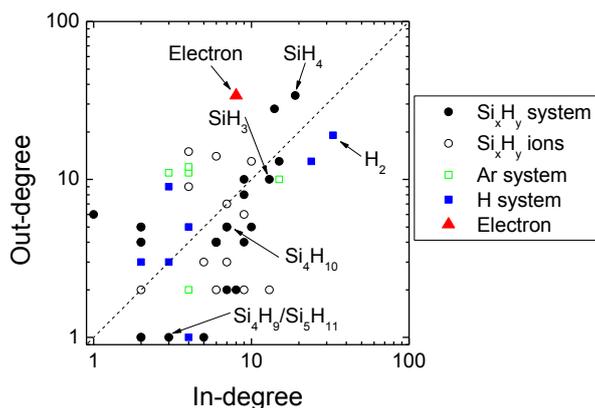


Figure 2. Distributions of in-degree and out-degree of species (or nodes) in silane plasma displayed in Figure 1. Inset dashed line indicates equal cases of in-degrees and out-degrees.

Here, we propose a diagram with C_b and C_c to analyze a macroscopic property of a graph and to obtain microscopic roles of species simultaneously, as shown in Figure 3 for the graph in Figure 1. In general, both C_b and C_c show centrality indices, and a node closer to the center has a higher value, and positive correlation is roughly expected between them. However, this diagram includes various types of information, as described below. For instance, nodes Si_4H_9 and Si_5H_{11} are located in the periphery region of Figure 1. Both of them are in the low- C_b area in Figure 3, but their positions in the diagram are different. Si_4H_9 has three directed paths from other species and another to Si_4H_{10} . The degree, the total number of edges, is limited to 4 as shown in Figure 2, and we can recognize the edges and the nodes around Si_4H_9 in Figure 1. Consequently, Si_4H_9 is not a significant species that affects many reactions, and it is located in the low- C_b area in Figure 3, while C_c of Si_4H_9 is in the middle range. Si_5H_{11} also has three directed paths from other species and one directed path to another species, Si_5H_{12} . It is in the low- C_b area with very low C_c . The difference of locations between Si_4H_9 and Si_5H_{11} is destination of out-degree edges; Si_4H_{10} is more active in reactions than Si_5H_{12} in our model, and we can clarify such a point using this diagram. Note that both of them have the same statistical values of degrees, and they are in the same position in the degree distribution that is a conventional classification in theory of complex network [2].

From a macroscopic point of view, this diagram displays semi-statistical properties of the graph as a visual classification tool. The overall distribution of data points displayed here is fairly uniform along the $\log(C_b)$ axis, and it covers the range in 4 orders of magnitude of C_b . The Si_xH_y system, which includes main species coming from the mother gas, SiH_4 , is distributed throughout the entire range. The H system, which also originates from SiH_4 , scatters in all ranges. The Ar system is in a less important area as C_b values. In the area with high-value C_b , stable species such as SiH_4 and H_2 exist, and electrons are also one of the highest-value species both in C_b and C_c . SiH_3 , which is the most important precursor for Si thin film deposition using silane plasma, is also in this high-value region. In the case of the linear C_b axis shown in the inset of Figure 3, data points are concentrated around 0, and a few of them scatter in the area of high C_b , which is not a sufficient visualization of the graph. A logarithmic plot for C_b is a key manner for graphs for reaction paths and visual classifications of chemical reactions, in particular, in plasma-enhanced chemistry.

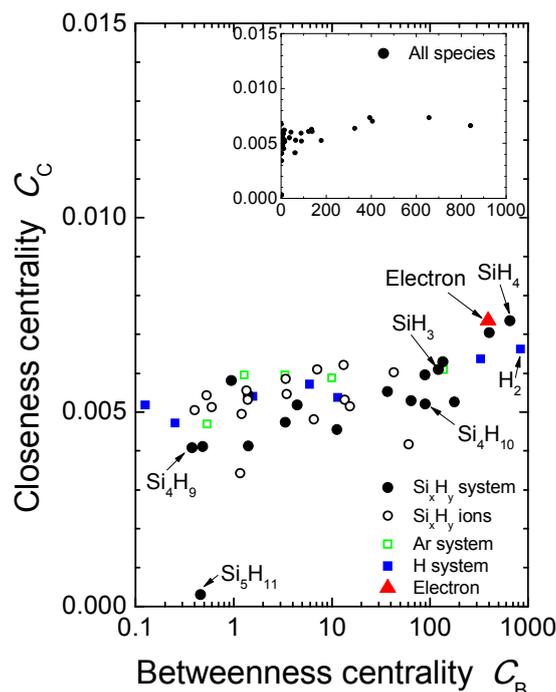


Figure 3. Diagram of betweenness-closeness centrality indices of reaction network in silane plasma displayed in Figure 1. Inset is diagram in linear scales.

Another fact that we can deduce from this diagram is rough estimation on differences coming from roles of species. Values of C_c , which are shown along the vertical axis, become high when the out-degree of a given species is larger than its in-degree (see Figure 2 to confirm these correspondences). For instance, C_c of SiH_4 is larger than that of H_2 : the ratio, in-degree/out-degree of SiH_4 , is 19/34, while that of H_2 is 33/19. Electrons have also unbalanced values: the ratio is 8/34. These facts indicate that species mainly working as agents (on the left-hand side of chemical reactions, such as A and B in Reaction (1), having larger out-degrees) are in the upper area of C_c . Products in chemical reactions tend to be in the lower area.

Figure 4 shows another example of chemical reactions, CH_4 system in methane plasma. In [8], we treat species except ions, but here we include all species listed in Ref. [10], shown in Table A2 in Appendix A. The $\log(C_b) - C_c$ diagram for the graph in Figure 4 is shown in Figure 5, which is quite similar to that in Figure 3; the area of $\log(C_b)$ is quite wide in comparison with the width of the C_c range. Also, most of the species that have been so far pointed out on their importance of roles are located in the higher- C_b range. In a microscopic point of view, we can also find similar points to the case of silane plasma. For instance, CH_4 and CH_3 are in the similar locations to SiH_4 and SiH_3 in Figure 3, respectively. Our previous report [8] in which the centrality indices similar to PageRank [17] indicate importance of CH_3 , and the result here is consistent with the one in Ref. [8]. There are a few points that are different from the case of silane plasma, such as the location of electrons between Figures 3 and 5, although such a feature may be a factor coming from each specific system.

Two results shown here indicate that complex reactions in plasma chemistry can be visualized in a $\log(C_b) - C_c$ diagram, and both macroscopic and microscopic properties are derived from it. In particular, since plasma chemistry includes various levels of roles as well as wide range of contribution frequencies to reactions, the range of node distributions along the $\log(C_b)$ axis is quite wide. C_b , given by Equation (2), includes information as reaction connection between species, while C_c indicates rather simple information about a location in a network.

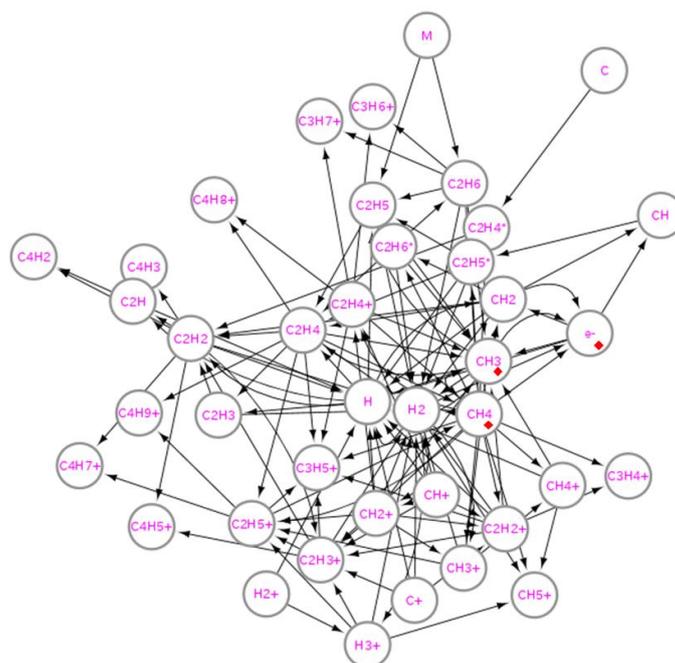


Figure 4. Graph for chemical reaction network in methane plasma. Nodes represent species in chemical reactions, and edges start from agents and ends at products of each reaction. Chemical reactions are listed in Ref. [10]. Closed red diamonds indicate species displayed in Figure 4.

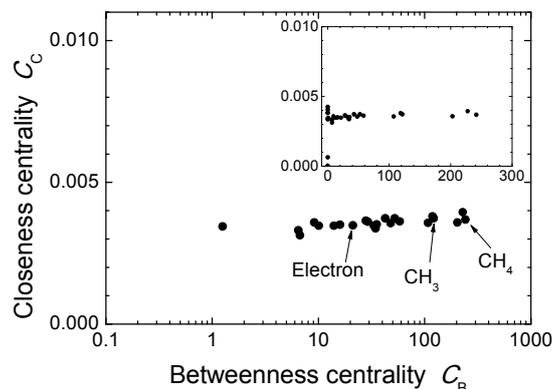


Figure 5. Diagram of betweenness-closeness centrality indices of reaction network in methane plasma displayed in Figure 3. Inset is diagram in linear scales.

3. Discussion

Figures 3 and 5 show wide-range distributions of nodes on the $\log(C_b)$ axis, but it might arise from simple randomness that also exists in random graphs. Here we compare such tendencies to those in arbitrary random graphs.

We fixed the numbers of nodes and edges to the ones in Figure 1, and created random graphs with directed edges in computation. Figure 6 shows three examples, and we cannot see any common points in allocation of roles for specific species. In all $\log(C_b) - C_c$ diagrams shown in Figure 7, the values C_b and C_c for the nodes are around a certain area. In particular, the values of $\log(C_b)$ is in a range with approximately one order of magnitude except for a few nodes. This is attributed to the fact that degrees of random graphs are in the Poisson distributions [18].

To compare spectra in the $\log(C_b)$ scale, the data in Figures 3, 5 and 7 are summarized as histograms of cumulative probabilities or relative densities P in Figure 8. Values of the nodes in random graphs are localized around 10^2 , and almost no changes among these three plots. This is attributed to the fact that degrees of random graphs are in the Poisson distribution [18], and these graphs have certain averaged profiles of parameters with some deviations. On the other hand, the spectrum of silane plasma chemistry with the same numbers of the nodes and the edges is quite broad. The case of methane plasma chemistry shows similar tendencies. This comparison implies that networks of complex plasma chemistry include self-arranged systematic roles in the constituent species. The roles of agents and products are partially distinguished from scattering along the C_c axis in the diagram. A wide range of $\log(C_b)$ indicates a number of levels of intermediate roles from one species to another along successive sequential reactions, where the levels are from inevitable functions in reactions to less-frequent contributions for sub-products creations, etc.

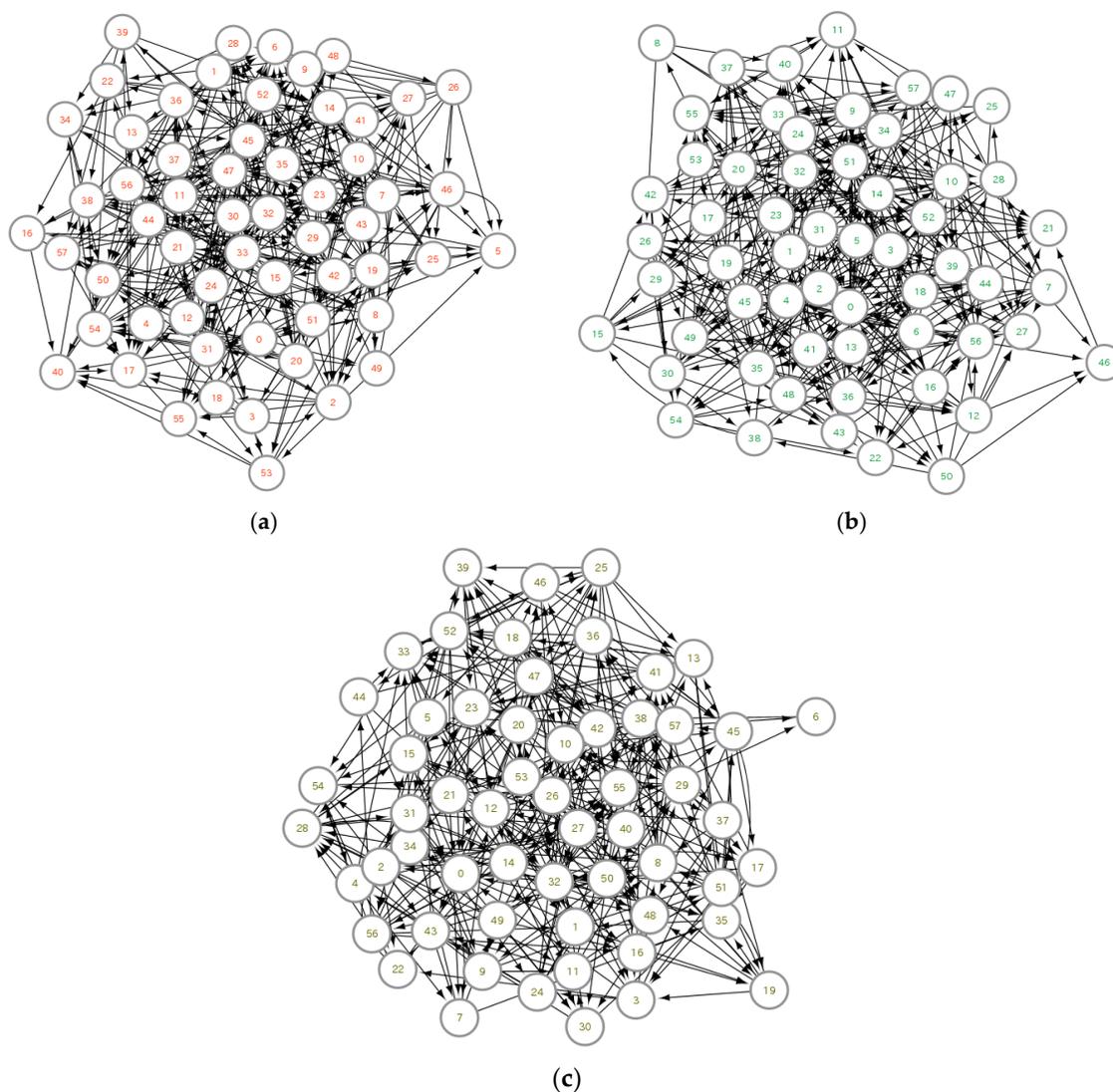


Figure 6. Random graphs created randomly in computation (a–c).

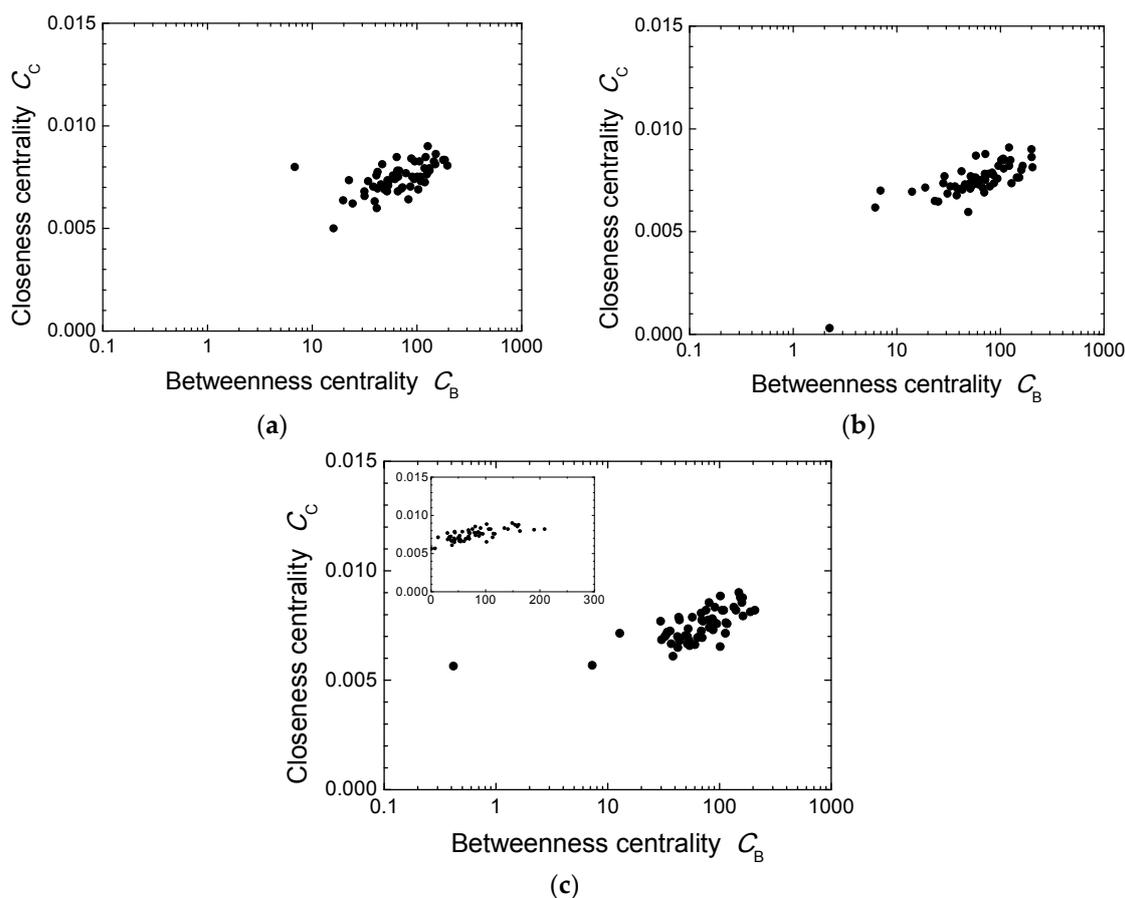


Figure 7. Diagrams of betweenness-closeness centrality indices. (a–c) corresponds to (a–c) in Figure 6, respectively. Inset is diagram in linear scales.

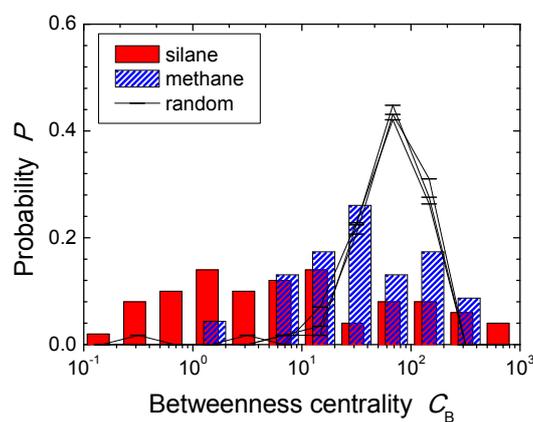


Figure 8. Histograms of cumulative probabilities or relative densities as function of betweenness centrality index. Data points are from Figures 3, 5 and 7.

The above descriptions are based on the data set on silane and methane plasma chemistry. If another sufficient set of reactions is available, one can perform similar data analysis on complex chemical networks. As shown here, approximate estimations of roles on species and global balance in the reaction system are beneficial for various purposes such as rapid analysis of robustness of a reaction system against impurity mixing. Another feasible and interesting contribution is selection of species for monitoring an industrial chemical reactor; behaviors of the limited number of the species detectable

on the edge of reaction space can predict the ongoing entire system working on by comparing values of indices with the detected signals.

This kind of graph analysis using diagrams composed of multi-centrality indices is applicable for analyses of other networks. In particular, networks that include less than 100 nodes are suitable for this method to investigate both macroscopic topologies and microscopic identifications of nodes. This is also applicable for larger networks to clarify roles of nodes by specifying their positions in a rough distribution of node data in a diagram. For instance, other physical and chemical processes can be analyzed by this approach when they are described in rate equations, since chemical reactions are given in one type of rate equation and successfully interpreted in this study.

4. Conclusions

A diagram with axes of multi-centrality indices works to clarify both macroscopic structures of the graph and microscopic features such as roles of a species. Plasma-enhanced chemical reactions, in which tens of chemical reactions take place in parallel, are analyzed using a diagram based on C_b and C_c indices. The derived diagram indicates, from its macroscopic distribution of node values, that nodes have a wide variety of roles that are quite different from random classification. It also shows some roles of specific nodes by their locations, such as roles of products, agents, and intermediates. For plasma chemistry, this study provides us a measure for statistical properties for tens or hundreds of species in reaction systems, and for graph analysis, this analysis will become a practical and applicable method for understanding such medium-sized graphs.

Acknowledgments: One of the authors (Osamu Sakai) thanks Professor Tomoyuki Murakami at Seikei University, Professor Mark J. Kushner at University of Michigan, and Professor Tsuyoshi Akiyama at The University of Shiga Prefecture for their useful comments in this study. This study was partly supported by Grant-in-Aid for Scientific Research from the Japanese Ministry of Education, Culture, Sports, Science and Technology, Japan (JSPS KAKENHI Grant Number JP16K13711).

Author Contributions: Yasutaka Mizui performed main analysis shown here and wrote the paper by drawing figures. Tetsuya Kojima performed numerical calculation for random graphs. Shigeyuki Miyagi prepared for calculation environments and gave advices on analysis procedures including error corrections. Osamu Sakai conceived and designed the research activities, and wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

The chemical reactions we analyze in this study are in two sets: silane (SiH_4) and methane (CH_4) plasma chemical reactions. All species treated in Figures 1–3 are shown in Table A1, and those in Figures 4 and 5 are in Table A2; they are listed in [10,11], respectively.

Table A1. List of species in silane plasma chemical reactions [11]. Except charge indicators “+” and “−”, other symbols like “*” and “v” mean various electronic and mechanical excitation levels of species. M indicates arbitrary species except ions.

Number	Species	Number	Species
0	Ar	31	$\text{Si}_2\text{H}_6^{**}$
1	Ar *	32	Si_2H_6^+
2	Ar **	33	Si_2H_7^+
3	Ar^+	34	Si_3H_4^+
4	ArH^+	35	Si_3H_6^+
5	e^-	36	Si_3H_7
6	H	37	Si_3H_8
7	H *	38	Si_4H_{10}
8	H^+	39	Si_4H_2^+
9	H_2	40	Si_4H_6^+
10	$\text{H}_2(v(1))$	41	Si_4H_8^+
11	$\text{H}_2(v(2))$	42	Si_4H_9

Table A1. Cont.

Number	Species	Number	Species
12	H ₂ *	43	Si ₅ H ₁₀ ⁺
13	H ₂ ⁺	44	Si ₅ H ₁₁
14	H ₃ ⁺	45	Si ₅ H ₁₂
15	M	46	Si ₅ H ₄ ⁺
16	Si	47	SiH
17	Si ⁺	48	SiH ⁺
18	Si ₂ H ⁺	49	SiH ₂
19	Si ₂ H ₂	50	SiH ₂ ⁻
20	Si ₂ H ₂ ⁺	51	SiH ₂ ⁺
21	Si ₂ H ₃	52	SiH ₃
22	Si ₂ H ₃ ⁺	53	SiH ₃ ⁻
23	Si ₂ H ₄	54	SiH ₃ ⁺
24	Si ₂ H ₄ ⁺	55	SiH ₄
25	Si ₂ H ₅	56	SiH ₄ (v(1,3))
26	Si ₂ H ₅ ⁺	57	SiH ₄ (v(2,4))
27	Si ₂ H ₆		
28	Si ₂ H ₆ (v(1,3))		
29	Si ₂ H ₆ (v(2,4))		
30	Si ₂ H ₆ *		

Table A2. List of species in methane plasma chemical reactions [10]. Except charge indicator “+”, other symbols like “*” mean various electronic and mechanical excitation levels of species. M indicates arbitrary species except ions.

Number	Species	Number	Species
1	CH ₄	21	CH ₃ ⁺
2	H	22	CH ₄ ⁺
3	CH ₃	23	C ₂ H ₂ ⁺
4	CH ₂	24	C ₂ H ₃ ⁺
5	C ₂ H ₆ *	25	C ₂ H ₄ ⁺
6	M	26	C ₂ H ₅ ⁺
7	CH	27	H ₂ ⁺
8	C ₂ H ₅ *	28	H ₃ ⁺
9	C	29	e ⁻
10	C ₂ H ₄ *	30	C ₄ H ₃
11	C ₂ H ₆	31	C ₄ H ₂
12	C ₂ H ₅	32	CH ₅ ⁺
13	C ₂ H ₄	33	C ₃ H ₄ ⁺
14	C ₂ H ₃	34	C ₃ H ₅ ⁺
15	C ₂ H ₂	35	C ₄ H ₅ ⁺
16	C ₂ H	36	C ₄ H ₈ ⁺
17	C ⁺	37	C ₃ H ₆ ⁺
18	CH ⁺	38	C ₃ H ₇ ⁺
19	H ₂	39	C ₄ H ₇ ⁺
20	CH ₂ ⁺	40	C ₄ H ₉ ⁺

References

1. Fox, J. *Proceedings of the Symposium on Generalized Networks*; Polytechnic Press: New York, NY, USA, 1966.
2. Kolaczyk, E.D. *Statistical Analysis of Network Data: Methods and Models*; Springer: Berlin, Germany, 2009.
3. Temkin, O.N.; Zeigarnik, A.V.; Bonchev, D. *Chemical Reaction Networks*; CRC Press: Boca Raton, FL, USA, 1996.
4. Leenheer, P.D.; Angeli, D.; Sontag, E.D. Monotone chemical reaction networks. *J. Math. Chem.* **2007**, *41*, 295–314. [[CrossRef](#)]
5. Gorban, A.N.; Yablonsky, G.S. Extended detailed balance for systems with irreversible reactions. *Chem. Eng. Sci.* **2011**, *66*, 5388–5399. [[CrossRef](#)]

6. Jeong, H.; Tombor, B.; Albert, R.; Oltvai, Z.N.; Barabasi, A.-L. The large-scale organization of metabolic networks. *Nature* **2000**, *407*, 651–654. [[PubMed](#)]
7. Albert, R.; Jeong, H.; Barabasi, A.-L. Error and attack tolerance of complex networks. *Nature* **2000**, *406*, 378–382. [[CrossRef](#)] [[PubMed](#)]
8. Sakai, O.; Nobuto, K.; Miyagi, S.; Tachibana, K. Analysis of weblike network structures of directed graphs for chemical reactions in methane plasmas. *AIP Adv.* **2015**, *5*, 107140. [[CrossRef](#)]
9. Mizui, Y.; Nobuto, K.; Miyagi, S.; Sakai, O. Complex reaction network in Silane Plasma chemistry. In *Complex Networks VIII*; Springer International Publishing: Cham, Switzerland, 2017; pp. 135–140.
10. Tachibana, K.; Nishida, M.; Harima, H.; Urano, Y. Diagnostics and modelling of a methane plasma used in the chemical vapour deposition of amorphous carbon films. *J. Phys. D* **1984**, *17*, 1727–1742. [[CrossRef](#)]
11. Kushner, M.J. A model for the discharge kinetics and plasma chemistry during plasma enhanced chemical vapor deposition of amorphous silicon. *J. Appl. Phys.* **1988**, *63*, 2532–2551. [[CrossRef](#)]
12. Murakami, T.; Niemi, K.; Gans, T.; O’Connell, D.; Graham, W.G. Chemical kinetics and reactive species in atmospheric pressure helium-oxygen plasmas with humid-air impurities. *Plasma Sources Sci. Technol.* **2013**, *22*, 015003. [[CrossRef](#)]
13. Agliari, E.; Barra, A.; Bartolucci, S.; Galluzzi, A.; Guerra, F.; Moauro, F. Parallel processing in immune networks. *Phys. Rev. E* **2013**, *87*, 42701. [[CrossRef](#)] [[PubMed](#)]
14. Agliari, E.; Barra, A.; Schiavo, L.D.; Moro, A. Complete integrability of information processing by biochemical reactions. *Sci. Rep.* **2016**, *6*, 36314. [[CrossRef](#)] [[PubMed](#)]
15. Bie, C.D.; Dijk, J.; Bogaerts, A. The dominant pathways for the conversion of methane into oxygenates and syngas in an atmospheric pressure dielectric barrier discharge. *J. Phys. Chem. C* **2015**, *119*, 22331–22350. [[CrossRef](#)]
16. Lieberman, M.A.; Lichtenberg, A.J. *Principles of Plasma Discharges and Material Processing*; John Wiley and Sons: New York, NY, USA, 1994.
17. Brin, S.; Page, L. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.* **1998**, *30*, 107–117. [[CrossRef](#)]
18. Albert, R.; Barabasi, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **2002**, *74*, 47–97. [[CrossRef](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).