Chemistry International 8(4) (2022) 130-135

International Scientific Organization
http://iscientific.org/
Chemistry International
www.bosaljournals.com/chemint/

A network study of chemical relations: From elements to polynary compounds

Guoyong Mao[1]*, Runzhan Liu[1] and Ning Zhang[2]

[1]School of Electrical and Information Engineering, Changzhou Institute of Technology, Changzhou 213032, China
[2]Business School, University of Shanghai for Science and Technology, Shanghai 200093, China
*Corresponding author's E. mail: maogy@czust.edu.cn

A R T I C L E I N F O

Article type:
Research article
Article history:
Received August 2022
Accepted September 2022
October 2022 Issue
Keywords:
Network study
Core network
Chemical relationships
Elements
Polynary compounds

A B S T R A C T

The chemical compounds are composed of elements of different numbers. So far, more than 100 million compounds made up of 103 elements have been discovered. The aim of this study is to investigate the relation between compounds that have different number of elements. Using compounds extracted from a famous website (ignoring the stoichiometric of elements), we found that for compounds with *n* elements, *n-1* elements in this compound can form at least one compound that has *n-1* elements, suggesting that compounds with lower number of elements are foundations of compounds with a greater number of elements and compounds with more elements are partly originated from compounds with lower number of elements. Based on this relation, we build an evolution network that contains 100 elements and 99% of all existing polynary compounds. The analysis result indicates that we can use this network to predict some potential chemical compounds, to find cores of network and to get the evolution mechanism of elements, thus provides a new angle to study chemistry.

Capsule Summary: This study was designed to investigate the relationship between compounds that have different number of elements and this network study provides new insight to study chemical compounds for their network cores and mechanism analysis.

Cite This Article As: G. Mao, R. Liu and N. Zhang. A network study of chemical relations: From elements to polynary compounds. Chemistry International 8(4) (2022) 130-135.
https://doi.org/10.5281/zenodo.7115552

INTRODUCTION

Mendeleev's periodic table of elements revealed the law of elements, indicating that all elements actually constitute a system, in which there are various relations between elements, no element can exist in isolation from the system (Mendeleev, 1989; Scerri, 2007). As chemical reactions are the core of chemical research and reactions are relations between compounds (Schummer, 1998; Goutsias and Jenkinson. 2012; Sole and Munteanu, 2004), people begin to study these relations from the network angle (Estrada, 2008; Leal et al., 2012; Liu et al., 2019; Suarez, 2019). However, all these networks are built only from elements and binary compounds. To the best of our knowledge, these is no research on network that are constructed using compounds that have at least 3 elements, no relevant reports have been found to study the relations between all compounds that contain different number of elements from a systematic perspective, like Mendeleev's periodic table.

With the development of network science, researchers can collect various compounds into databases and make them public, making it possible to study the relationships between compounds that have different

number of elements. To this end, this paper will be based on the data of more than 100 million compounds collected in the network to discover the relations between them.

MATERIAL AND METHODS

Construction of network

Throughout this paper the stoichiometric of elements in compounds is ignored, as we only care about the ingredient of compounds, or how many different elements a compound has. Therefore, if V is the set of elements for a given compound, V_{b1} for binary compound A_1B_2 is identical to V_{b2} for binary compound A_2B_3, as V_{b1} = (*A, B*) and V_{b2} = (*A, B*).

All data used in this paper are from *ChemSpider*, a free chemical database owned by the Royal Society of Chemistry (RCS, 2022). It provides fast access to over 100 million structures, properties and associated information. With the online searching APIs provided by *ChemSpider*, we get 17010 element sets, ranging in length from 1 to 10, as listed in the second row of Table 1. All data are collected before the last day of 2021 and compounds with isotopic elements are ignored.

From Table 1 we can see that compared with the total number of compounds of above 100 million, the number of elements set of compounds is hugely reduced. The reason for this can be exemplified from element set of V_{t1} = (*H, C, O*), as there are over 2150000 ternary compounds that contain element *C*, *H* and *O*, yet these compounds constitute just one set. In the second row of Table 1, 103 is the total number of the single-element sets, which equals to the total number of elements. 1946 is the total number of sets that have two elements, which means that not any two elements can form binary compounds, otherwise, the total number of binary sets would be 5253 (103*102/2). The same phenomenon occurs in sets that have three to ten elements. There are only 7 sets that contain 10 elements and the maximum set length is only 10.

From set theory we know that if V_b is the element set for binary compound A_iB_j ($i\neq0$, $j\neq0$) and V_t is the set of elements for ternary compound $C_lA_mB_n$($l\neq0$, $m\neq0$, $n\neq0$), then V_b is a subset of V_t, as V_b = (*A, B*) and V_t = (*C, A, B*). Also from set theory, we know that in a set, each element has the same status and there is no order between the elements. Therefore, (*A, B*) = (*B, A*) and (*C, A, B*) = (*B, A, C*).

For 1946 sets of binary compounds, each of them contains 2 elements. Obviously, all 1946 binary set covers 2 single-element sets, namely, set (*A, B*) covers set (*A*) and set (*B*). For 2744 sets of ternary compounds, we found that 2742 of them can cover at least one subset within 1946 binary sets. For example, Set V_{t1} = (*H, C, O*) can cover 3 binary sets, which are set (*H, C*), (*C, O*) and (*H, O*). Set V_t = (*Ge, Rb, S*) can cover only one binary set *(Ge, S*), as (*Ge, Rb*) and (*Rb, S*) cannot be found in 1946 binary sets. Only 2 ternary sets, which are V_{t2} = (*Re, Rn, Sn*) and V_{t3} = (*B, Ir, Na*), cannot cover sets within 1946 binary sets. The cover relation between element set of length *n* and length *n-1* is listed in the third row of Table 1. Out of 16907(17010-103) sets that have at least 2 elements, only 149 of them, or 0.88% of the total, cannot establish the cover relation.

We build a directed network from this cover relation. For example, ternary set (*H, C, O*) covers (*H, C*), (*C, O*) and (*H, O*), then 3 directed edges are added to this network, which are edges from (*H, C*) to (*H, C, O*), from (*C, O*) to (*H, C, O*) and from (*H, O*) to (*H, C, O*). This network is constructed using the edge list format and node is represented in set format, as listed in Table 2.

From Table 2 we can see that node (*H, C, O*) is not only the target for (*H, O*), (*H, C*) and (*C, O*), but also the source node for node (*H, C, O, P*). Therefore, this network can be viewed as the evolution network as it connects together single element and compounds that have different number of elements, and gives the evolution path among them.

This network contains 16883 nodes and 51009 directed edges, among them 100 out of 103 elements appears in single-element node. The reason for this is, in *ChemSpider* database, there is no compound that contains element *Fr*, only one five-elements compound that contains element *Fm*, 3 five-elements compounds and 2 four-elements compounds contain element *No*. Therefore, there is no binary compound that contains these 3 elements such that the cover relation between binary sets and single-element sets cannot be established. For the 149 element sets that cannot establish the cover relation with sets of less elements, 25 of them can be covered by sets of more elements. For example, quaternary element set (*Ce, F, La, O*) cannot find subset in 2744 ternary element sets, but it is a subset of five-elements set (*C, Ce, F, La, O*). Therefore, only 124 compounds are not in the network, making the total number of nodes to be 16883 (17010-124-3).

Analysis of network

We analyze this network from three aspects: connectivity, centrality of nodes and edges, and core subnet. Because the connectivity of the network can reflect the connectivity of the nodes in the network, if the connectivity is very low, it indicates that most nodes are isolated, and such a network is not of great research significance. The centrality of nodes and edges can reflect which nodes and edges are in an important position in the network, and the core subnet is of great significance to study the origin and evolution of the network.

Table 1: Number of elements sets of different length

Set Length	1	2	3	4	5	6	7	8	9	10	Total
Set Number	103	1946	2744	3095	3358	3208	1806	625	118	7	17010
Covered Set Number	|	1946	2742	3082	3318	3174	1790	599	105	5	16761

Table 2: The edge list format

Source node	Target node
(H)	(H, O)
(O)	(H, O)
(H, O)	(H, C, O)
(H, C)	(H, C, O)
(C, O)	(H, C, O)
(H, C, O)	(H, C, O, P)
(H, C, O, P)	(H, C, O, P, U)

Python and *R* are used as the program language. *NetworkX* (Hagberg et al., 2012) and *igraph* (Csardi and Nepusz, 2006), two publicly available free packages for network analysis, are used to do all the analysis job, and to draw figures in this paper.

Connected components

This network is not connected, as it contains 16 connected components, the biggest component contains 16850 nodes and 50991 edges, accounting for 99.8% of the total nodes and 99.98% of the total edges, indicating that most nodes are connected and most compounds can be formed from compounds with less elements. The other 15 components contain only 33 nodes and 18 edges. For the 15 components, 3 of them have 3 nodes and 2 edges, they are placed in the first 6 rows of column 1 and column 2 in Table 3. Other 12 components have 2 nodes and 1 edge, they are placed in the rest part of Table 3.

Connectivity of network

There are two possible reasons why network connectivity is 99.8% rather than 100%. The first reason is that some compounds are missing, that is, some compounds are not yet collected by *ChemSpider* database. The second reason is that some compounds related to nodes listed in Table 3 are not regular compounds. For example, according to *ChemSpider*, the name of compounds *FeMnOSnSr* in Table 3 is Stannane, oxo-, iron manganese strontium salt (1:1:1:1), it is a mixture rather than a compound. There are 11 nodes constructed by such compounds in 15 components. If these irregular compounds are removed from the network, the connectivity of network can be improved.

Important nodes and edges

Betweenness centrality (*BC*) is one of the most important measurement indexes of network centrality (Brandes and Pich, 2007), *BC* of a node *i* is the sum of the fraction of all-pairs shortest paths that pass through *i*, *BC* of an edge *e* is the sum of the fraction of all-pairs shortest paths that pass through *e*. Nodes or edges with high *BC* can be regarded as the efficient mediators to connect different nodes, or in other word, to form more compounds. We use this index to

Table 3: 18 edges in 15 components

Source node	Target node	Source node	Target node
(Ag, Ba, Cu, Gd, O)	(Ag, Ba, Cu, Gd, O, Sm)	(Ge, K, O, Ti, W)	(Ge, H, K, O, Ti, W)
(Ag, Ba, Cu, O, Sm)	(Ag, Ba, Cu, Gd, O, Sm)	(Ba, Ca, Cu, Hg, O)	(Ba, Ca, Cu, H, Hg, O)
(K, O, Sb, V, W)	(H, K, O, Sb, V, W)	(B, Cr, Mo, Ni, Si)	(B, Cr, H, Mo, Ni, Si)
(K, O, Sb, V, W)	(K, Na, O, Sb, V, W)	(C, H, Os, P, Re, Se)	(C, Cl, H, Os, P, Re, Se)
(C, H, Re, Ru, S, Si)	(C, H, Ir, Re, Ru, S, Si)	(C, H, Mn, N, O, Re, Se)	(C, Cl, H, Mn, N, O, Re, Se)
(C, H, Ir, Re, S, Si)	(C, H, Ir, Re, Ru, S, Si)	(Ba, Co, Cu, Fe, O, Pb, W)	(Ba, Co, Cu, Fe, H, O, Pb, W)
(Fe, Mn, O, Sn, Sr)	(Fe, Mg, Mn, O, Sn, Sr)	(Al, Co, Cr, Mo, Ni, Ti, W)	(Al, Co, Cr, H, Mo, Ni, Ti, W)
(Bi, Ca, Cu, O, Sr)	(Bi, Ca, Cu, O, Pb, Sr)	(C, H, Li, Mg, N, Na, O, Si)	(C, F, H, Li, Mg, N, Na, O, Si)
(Bi, K, O, V, W)	(Bi, H, K, O, V, W)	(C, H, N, Ni, O, P, V, W)	(C, H, N, Na, Ni, O, P, V, W)

Table 4: Top 10 nodes and edges with the highest BC

rank	Node	Edge
1	(C, H, N, O)	(C)! (H, C)
2	(H, C, O)	(C, O)! (H, C, O)
3	(H, C, N)	(H)! (H, C)
4	(H, C, N, O, S)	(H, O)! (H, C, O)
5	(H, C)	(C, N, O)! (C, H, N, O)
6	(H, C, N, O, Cl)	(C, N)! (H, C, N)
7	(H, C, N, O, P)	(H, N)! (H, C, N)
8	(H, C, O, P)	(H, C)! (H, C, O)
9	(H, C, O, S)	(H, N, O)! (C, H, N, O)
10	(H, C, N, O, F)	(C, P)! (H, C, P)

measure the importance of nodes and edges. The top 10 nodes and edges for *BC* is listed in Table 4.

We can see from Table 4 that 10 nodes with the highest *BC* are all composed of organ elements. We checked the top 100 nodes with the highest *BC*, 95 nodes are all composed of organ elements, 5 nodes contain element *Na*, which are (*H, C, N, O, Na*), (*H, C, N, O, S, Na*), (*H, C, N, O, S, Cl, Na*), (*H, C, O, Na*) and (*H, C, O, S, Na*). *BC* of these 5 nodes are higher than many nodes that are all composed of organ elements like (*H, O, F*) and (*C, N, O, Cl*), suggesting that element *Na* is active in constructing important nodes.

Core network

The *K* shell algorithm is usually used to find the subnetwork that conforms to the specified degree *k* in a network (Carmi et al., 2007). The subnetwork assumes the position of the core in the network. The higher the core degree is, the smaller the subnetwork and the larger the corresponding core degree of the subnetwork is. In a sense, the subnetwork of core degree division plays an important role in the original network, such as the origin and evolution trend tracing of the network. We use this algorithm to get the core of the network, the specified degree *k* is 7 as this network cannot be decomposed when *k* = 7. The core network contains 192 nodes and 704 edges. All nodes are made up of at least one element, which is element *C*, and at most 9 elements, which are element *H, C, N, O, S, F, Cl, Br* and *I*.

Results and discussion

From the analysis of network, we know that almost all nodes are connected and covers almost all elements and compounds, making the network suitable to represent existing elements and compounds, and to study element and compounds from the network angle.

Evolution trend of core network

For the core network that has 192 nodes, there is 1 single-element node, 7 two-elements nodes, 22 three-elements nodes, 41 four-elements nodes, 50 five-elements nodes, 41 six-elements nodes, 22 seven-elements nodes, 7 eight-elements nodes and 1 nine-element nodes. The evolution trend from single-element node *C* to nodes with more elements is illustrated in Fig 1 (To display nodes and edges clearly, we only give the first 3 steps of evolution). Among 192 nodes, all nodes contain element *C*, 128 nodes contain element *H*, 96 nodes contain element *S, N, O, F, Cl* and *Br*, 64 nodes contain element *I*, indicating that element *C* is the origin of the core network, which is consistent with what we generally agree.

Compounds prediction

As mentioned in the previous section, 124 compounds cannot establish the cover relation. For example, ternary element set (*Re, Rn, Sn*) and (*B, Ir, Na*) cannot cover sets within 1946 binary sets. We can assume that at least one of the binary sets (*Re, Sn*) or (*Re, Rn*) or (*Rn, Sn*) is missing for (*Re, Rn, Sn*), and

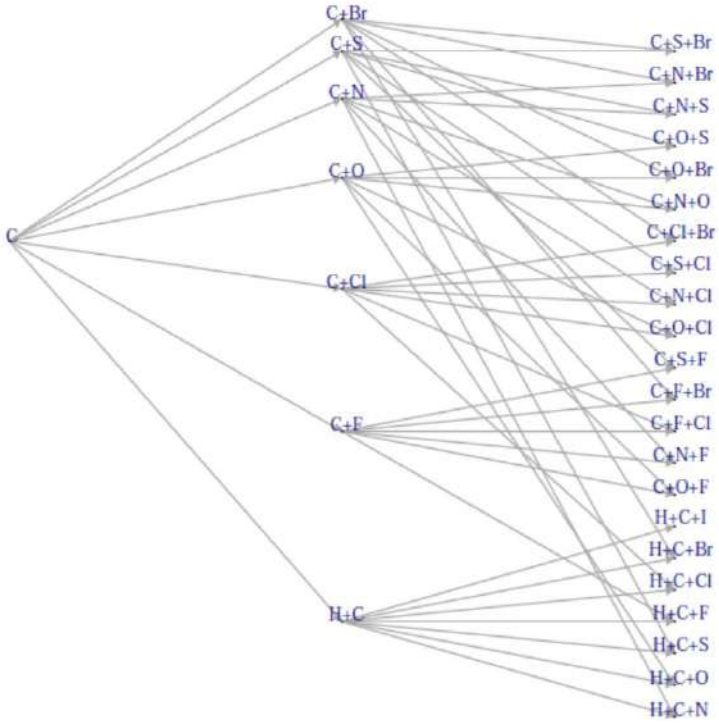

Fig. 1: Evolution trend of core network

at least one of the binary sets (*B, Ir*) or (*B, Na*) or (*Ir, Na*) for (*B, Ir, Na*) is missing. The later assumption is verified by literature (Zeng et al., 2014) that proves the existence of element set (*B, Ir*)—it gives the mechanical properties of compound *IrB* and IrB_2. Similar assumption can be made for other compounds for prediction purpose.

Reachable depth of elements

For 16907 nodes that have at least 2 elements, the number of nodes that contain various elements differs greatly. For example, 11457 nodes contain element *H*, while the number of nodes contains element *Md* is only 2. If we neglect

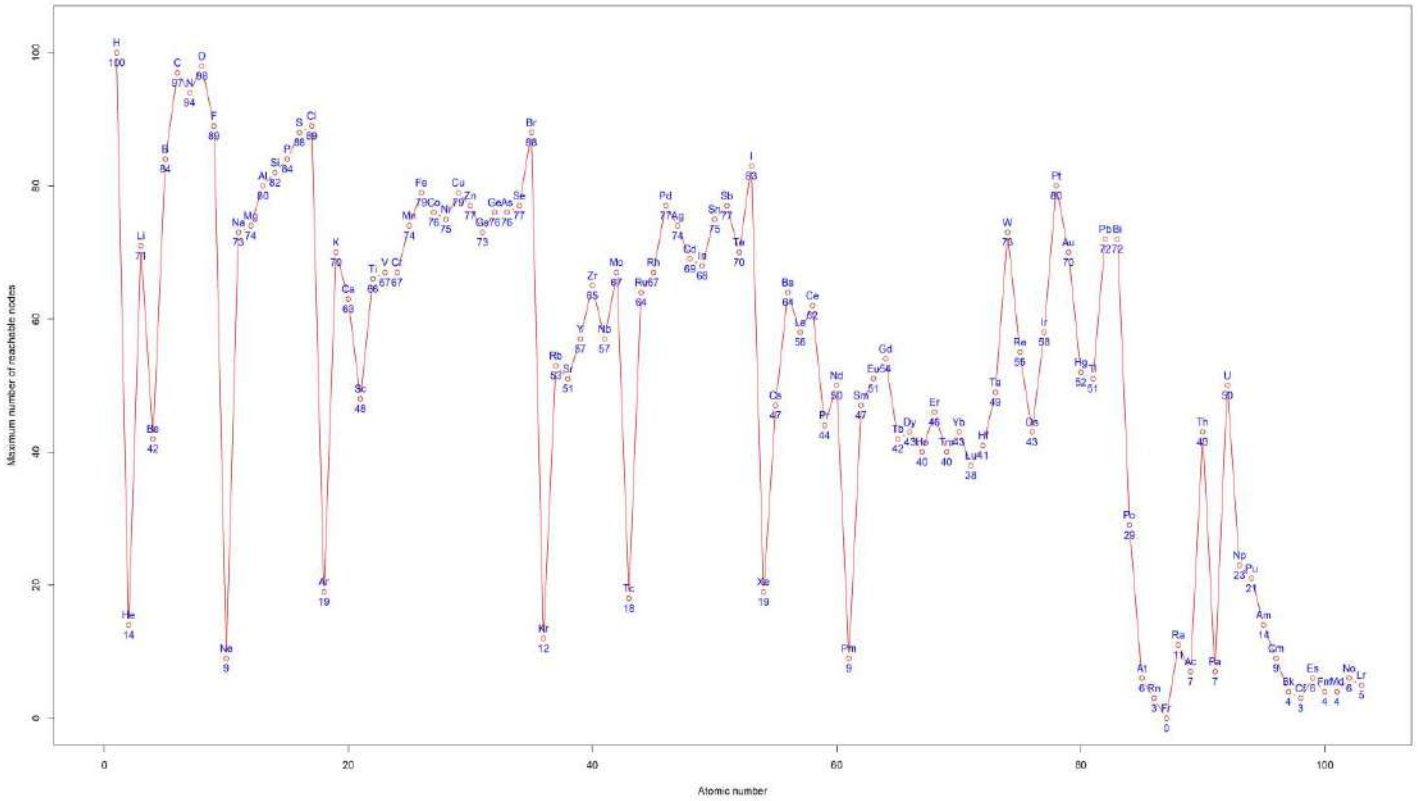

Fig. 2. Maximum reachable number of elements

Reachable number of elements

Reachable number of elements *R (E, S)* is defined as the number of elements that element *E* can reach in *S* steps. For example, element *Li* can reach 33 elements in one step, as there are 33 binary compounds contain element *Li*. It can reach 53 elements in two steps, as there are 20 ternary compounds that contain element *Li*. It can reach a maximum of 72 elements in 6 steps. For all elements, the maximum number of reached elements is periodically changed, as illustrated in Fig. 2.

We also found that some elements can reach all reachable elements in short steps. For example, element *H* can reach all 100 reachable elements in 4 steps. While some elements take more steps to achieve this goal, for example, 6 steps for *Li* to reach all 72 elements. we found that 6 is the maximum step for all elements to get to the maximum reachable elements. In other words, if element *A* cannot reach element *B* in 6 steps, then we cannot expect *A* to reach *B* in more steps and *B* is unreachable for *A*.

elements that the number of occurrences in all nodes is less than 10, which are element *Ac, Cm, Ne, At, Pa, Pm, Bk, Es, No, Cf, Rn, Md, Lr, Fm* and *Fr*, 15 elements that are very hard to find their compounds, then for other 88 elements, 86 of them bear the same interesting rule as that if element *A* appears in a compound with *n* elements, then this element can be found in compounds with *k* elements (k = 2, 3, 4, ..., n-1). If element *A* cannot be found in compounds with *i* elements (i = 2, 3, ...,9), then this element cannot be found in compounds with *j* elements (j = i+1, i+2, ...,10). For example, element *In* appears in a compound with 7 elements, then it can be found in compounds with elements number ranging from 2 to 6. Also, element *In* cannot be found in a compound with 8 elements, then it cannot be found in compounds with 9 or 10 elements. Therefore, the reachable depth of element *In* is 7. Obviously, the reachable depth for most non-metallic element is 10. Two elements that do not follow this rule are element *Th* and element *Sc*, as *Sc* appears in compound with 10 elements but does not appear in compounds with 8 elements and 9 elements, Th appears in compound with 8 elements but does

not appear in compounds with 7 elements. This may be due to a lack of data and can be used for prediction purpose—we can assume that 7-elements compound contains element *Th* exist, 8-elements compound and 9-elements compound contain element Sc exist.

CONCLUSIONS

The main contribution of this paper is the discovery of cover relation between compounds with *n* elements and compounds with *n-1* elements, and the building of the network that can integrate almost all existing elements and compounds using this relation. Almost all nodes in this network are connected, not isolated, making the study of properties like network core, node centrality and node reachability meaningful, and making network tools like node prediction applicable, thus provide a new way to study chemical relations. This paper only gives several applications of this network now, we are expecting more outcomes in the future, for example, the meaning of network topology, the meaning of shortest path between nodes, and the similarity between nodes from network point of view.

ACKNOWLEDGEMENTS

The author would like to thank anonymous reviewers for their valuable suggestions and comments. This work was supported by the National Natural Science Foundation of China under Grant no. 70971089.

REFERENCES

Brandes, U., Pich, C., 2007. Centrality estimation in large networks, International Journal of Bifurcation and Chaos 17, 2303-2318.

Carmi, S., Havlin, S., Kirkpatrick, S., Shavitt, Y., Shir, E., 2007. A model of Internet topology using k-shell decomposition. Proceedings of the National Academy of Sciences 104, 11150-11154.

Csardi, G., Nepusz, T., 2006. The igraph software package for complex network research, InterJournal Complex Systems 1695,1-9.

Estrada, E., 2008. The complex networks of earth minerals and chemical elements, MATCH Communications in Mathematical and in Computer Chemistry 59, 605-624.

Goutsias, J., Jenkinson, G., 2012. Markovian dynamics on complex reaction networks, Physics Reports 529, 199-264.

Hagberg, A., Schult, D., Swart, P., 2008. Exploring Network Structure, Dynamics, and Function using NetworkX, Proceedings of the 7th Python in Science Conference, Pasadena, CA, USA, 11-15.

Leal, W., Restrepo, G., Bernal, A., 2012. A network study of chemical elements: From binary compounds to chemical trends, MATCH Communications in Mathematical and in Computer Chemistry 68, 417-442.

Liu, R., Mao, G., Zhang, N., 2019. Research of chemical elements and chemical bonds from the view of complex network, Foundations of Chemistry 21,193-206.

Mendeleev, D.I., 1889. The periodic law of the chemical elements, Journal of the Chemistry Society Transactions 55, 634-656.

RSC (Royal Society of Chemistry), Chemistry Database [DB/OL], http://www.chemspider.com, 2001-2022.

Scerri, E., 2007. The Periodic Table: Its Story and Its Significance, Oxford University Press, New York.

Schummer, J., 1889. The chemical core of chemistry I: A conceptual approach, Hyle International Journal for Philosophy of Chemistry. 1998, 4, 129-162.

Sole, R.V., Munteanu, A., 2004. The large-scale organization of chemical reaction networks in astrophysics, Europhysics Letters 68, 170-176.

Suarez, R., 2019. The network theory: A new language for speaking about chemical elements relations through stoichiometric binary compounds, Foundations of Chemistry 21, 207-220.

Zeng, X., Zhu, X., Li, D., Chen, Z., Ai, Y., 2014. First-principles calculations of the mechanical properties of IrB and IrB_2, Acta Physica Sinica 63, 153101.