Structures and measures in knowledge processing models

Valerian Ivashenko
Intelligent Information Technologies department
Belorusian State University of Informatics and Radioelectronics
Minsk, Belarus
ivashenko@bsuir.by

***Abstract*—This paper contains an overview of approaches to analyzing the structure and quantitative properties of ontological substructures embeddable into the semantic space. Requirements and an approach to the analysis of semantic structures and the synthesis of semantic measures are considered. The results are focused on identifying invariants of ontological structures for knowledge management problems.**

***Keywords*—semantic space, topological space, semantic measure, semantic metrics, dynamic network analysis, denotation semantics, operational semantics, game semantics, graph centralities, meta-operation, knowledge management, model of unified semantic representation of knowledge**

I. Introduction

The semantic space is associated with the process of cognition. Meanings, by their primary nature, are not something given at once, although they can be assimilated that way, but appear in the process of knowledge formation in the forms of signs and values. The becoming of knowledge is described by a meta-model of semantic space, which considers the formation of semantic subspaces [1]. A detailed discussion of approaches to constructing semantic spaces is discussed in [1].

The analysis of structural and quantitative features of semantic subspaces is considered as a process of cognition which can be expressed by a relationship within the framework of the knowledge specification model [2,1]. Analysis of structural properties establishes a correspondence between the analyzed structure and a scale. This correspondence considered as a mapping onto some ontological structure, for example, an ordinal scale.

Let us also consider the measures and metric features of the structures of the semantic space not only for denotational, but also for operational semantics. Such features can be considered as invariants of the structures of the semantic space which can be used in tasks of comparing knowledge bases and equivalent transformations when processing knowledge including not only the tasks of reducing operational semantics to denotational ones but also in the opposite direction. Throughout this discussion we will use both models and methods of network analysis including dynamic network analysis and models of formal concept analysis [3].

II. Brief overview of approaches to the analysis of structures and models of quantitative features

Since the semantic space is considered for the structures of texts in languages of the model of a unified semantic representation of knowledge, the structures under consideration are reducible to graph ones [2,1].

As for the structural features, they are based on the structural features of graphs [4,5,6,7,8]:

global (absolute):

- coherence/incoherence,
- strongly connected\not strongly connected,
- non-recurrence (acyclicity)\recurrence (cyclicity),
- regularity/irregularity,
- trivial automorphism,
- determinism/non-determinism,
- reversibility/irreversibility;

local (relative):

- non-trivial automorphism,
- non-recurrence (acyclicity)\recurrence (cyclicity),
- determinism/non-determinism
- reversibility/irreversibility.

For the purpose of identifying operational semantics, such methods of structure analysis as identifying classes of automorphic elements, which have the global nature of the necessary structure analysis (working in conditions of complete information), are suitable for finite structures, texts, finite state machine models, etc. For potentially infinite structures such as languages, sets of syntactically correct texts, local methods can be applied focused on reducing dynamic structures to static ones and reducing the task of analyzing operational semantics [3] to analyzing denotational ones [9], which can also differ: working only in conditions of complete information, or which can work under conditions of incomplete information. Thus, the identification of operational semantics can be organized based on the analysis of automorphisms, transition to sublanguages and automorphic analysis based on queries.

For a deeper study of structural features, methods of analysis of formal concepts are applicable. The use of methods for the analysis of formal concepts is considered as a special case of the use of the apparatus of meta-operations [2] which includes various types of compositions.

It should be noted that the relationship between operational and denotational semantics and the transition from operational semantics to denotational semantics was considered in [3]. The transition in the opposite direction is closely related to the processes of formation of the sign of a set from its elements and (restoration) of the elements of the set according to its designation (sign).

Due to the fact that sets can be reduced to such processes and operations, operations on sets that transform sets can be considered as meta-operations. Meta-operations are discussed in [2].

Each meta-operation can be specified by a sequence of other meta-operations. For example, closure within the framework of formal concept analysis can be expressed by the following sequence:

$$x' = \left((x \overleftarrow{\circ} A) \overrightarrow{\circ} A\right)$$

An important quality of the studied features of space elements, including structural and quantitative features, is their stability in relation to the identified structure of semantic relationships in the process of knowledge formation.

Stability assumes that if a certain feature is considered in a substructure, then in the presence of a superstructure belonging to the same class of structures, for any feature specified in the substructure, there is a feature specified in the superstructure the values of which lie in the same interval as the values of the feature specified in the substructure, for all arguments from its domain of definition (attribute specified in the substructure).

$$\Big((S \subseteq U) \to \big(\gamma(\langle f, S, \lambda(S)\rangle) = \gamma(\langle f, U, \lambda(S)\rangle)\big)\Big)$$

$$\Big((S \subseteq U) \to \big(\gamma(\langle f, S, \lambda(S)\rangle) \subseteq \gamma(\langle f, U, \lambda(U)\rangle)\big)\Big)$$

$$(\gamma(\langle \varphi, \alpha, \beta\rangle) = \{\langle \chi, \varphi(\langle \alpha, \chi\rangle)\rangle | \chi \in \beta\})$$

The basis of such stability is topological relations in the structure of semantic interrelations of space elements. The structure of semantic space can be expressed through the concept of topological closure.

Topological closures are the basis for identifying topological subspaces.

Among the topological spaces considered are topological spaces based on extensional and fully connected closures, which can be finite. Such spaces are stable with respect to NON-factors of knowledge [1].

When analyzing topological properties, the structure of the ordinal scale must have the properties of a lattice, and additional restrictions are imposed on the mapping of the analyzed structure onto this lattice. To determine the sets of elements (topological base) that are mapped into a sublattice (one lattice node), meta-operations are considered that implement (topological) closure operators.

Stable structures of semantic space can be considered as components for constructing knowledge bases. Such structures form a topological semantic space based on the denotational semantics of the represented knowledge. From the point of view of operational semantics, one can also analyze the topological properties of operations based on their transitive closures within the corresponding topological space. Such closures can be irreflexive or recurrent (including reflexive), which is associated with the recurrence property characteristic of reconstruction and selection problems in accordance with the general classification of problems [2].

Since the number of elements in the fragments under study is not limited, the number of features that map the elements of the fragments onto ordinal and metric scales is not limited. The number of possible display options for structures of different sizes is infinite, infinite and uncountable. Requirements for structural stability reduce the number of options, but presumably do not reduce it quantitatively. This assumption is associated with the number of structurally distinguishable fragments (motifs [10]) in graph structures, which are an objective basis for considering a significant number of different topologies on the same structure.

Stability requirements are general for both structural ($\lambda(S) = 2^S$), and for quantitative characteristics ($\lambda(S) = S$), ($\lambda(S) = S^2$), $f \in \left(2^{(S_2)^{S_1}}\right)^{S_1 \times S_2}$, where $\emptyset \subset S_2 \cap \mathbb{R}^n$. From the point of view of the metamodel of the semantic space, the transition in one of the arguments from it to its superset is a transition in time in accordance with the “information arrow of time”.

Synthesis of dynamic models based on the results of structure analysis and transition to the analysis of operational semantics. For this task, many methods can be considered, among which the following types can be distinguished: methods based on (topological) closures, including methods based on the analysis of automorphisms, including conditional morphisms (see dynamic network analysis [7] and game-theoretic centrality [11]), methods based on composition-based closures, unidirectional and bidirectional methods, including formal concept analysis methods. Below are examples of transitions from static structures to dynamic ones (Fig.5 and Fig.6).

As for quantitative characteristics, they are also based on quantitative characteristics of graphs:

The following features are used:

• quasimetric [4] having properties: identity, non-negativity, triangle inequality;

identity: $\rho(x, x) = 0$

non-negativity: $\rho(x, y) \geq 0$

Triangle inequality: $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$

$\rho(x, y) = \bigwedge_z \big(\rho(x, z) \oplus \rho(z, y)\big)$;

$\rho = \left(\left(A \underset{\oplus}{\wedge} \rho\right) \vee A\right)$, where ρ – quasimetric (matrix), A – adjacency matrix;

• farness:

$$F(x) = \oplus_y \rho(x, y)$$

• centralities [11,9,4]:

local centralities:

degree centrality:

$deg(x) = i * A * x$, where x – vertex vector, A – adjacency matrix, $i = \{1\}^{dim(i)}$ – one’s vector;

eigenvalue centrality [7]:

$$x = 1/\lambda_{max} * A * x$$

$A * x = \lambda_{max} * x$, where λ_{max} – maximal eigenvalue of A.

PageRank centrality:

$x_i = \alpha * \sum_j a_{ji} * x_i / \sum_i a_{ji} + (1 - \alpha)/n$, where a_{ji} – element of A, $\alpha \in [0; 1]$, n – number of vertexes.

Katz centrality:

$$x_i = \sum_{k=1}^{\infty} \sum_{j=1}^{N} \alpha^k * (A^k)_{ji}$$

$$x_i = \alpha * \sum_{j=1}^{N} a_{ij} * (x_j + 1)$$

communicability centrality:

$$Com(x) = (e^A)_{ii}$$

information centrality [1];

harmonic centrality:

$$F(x) = \oplus_{y \neq x} \frac{1}{\rho(x,y)}$$

closeness:

$$C(x) = 1/F(x)$$

percolation centrality:

$PC^t(v) = \frac{1}{n-2} * \sum_{s \neq v \neq r} \frac{\sigma_{sr}(v)}{\sigma_{sr}} * \frac{x^t{}_s}{\sum[x^t{}_i] - x^t{}_v}$, where σ_{sr} – number of shoterst paths, $\sigma_{sr}(v)$ – number of shoterst paths that pass through v, $x^t{}_s$ – percolation state of the node s at time t;

betweennes centrality:

$$C(v) = \sum_{s \neq v \neq r} \frac{\sigma_{sr}(v)}{\sigma_{sr}}$$

global centralities:

Freeman centralization:

$C_x = \frac{\sum_{i=1}^{n}\left(C_x(p_*) - C_x(p_i)\right)}{\max_G \sum_{i=1}^{n}\left(C_x(p_*) - C_x(p_i)\right)}$. where $C_x\left(p_i\right)$ – centrality of element p_i, $C_x\left(p_*\right)$ is centrality of element p_* centrality with maximal centrality among elements, G – any graph with the same vertexes number.

Considering the listed types of features (global – graph, vertex, edge, etc.), the identification of these features can be considered as the result of an extended cognition process within the framework of the knowledge specification model, moving from mapping elements of the formal ontology model of the original structure to mapping elements of formal ontology models (generalized) strings composed of elements of this structure. Thus, features can be classified by the type of their areas of departure and destination as well as by the type of computational complexity of meta-operations (within closed systems of meta-operations): features expressed by: a polynomial formula [12,13], ratio of polynomials, elementary functions, and features that are not expressed by elementary functions.

$\sum_{k=0}^{\infty}(A_R * \beta)^k$

$\sum_{k=0}^{\infty}\frac{(A_R * \gamma)^k}{k!}$

Let us note the following features: the defining properties of these features can be specified both in recurrent and non-recurrent form. The calculation of features can be done exactly (using a formula) or approximately (using a recurrent formula); in addition, the expression can be linear or nonlinear (power law, etc.). A more detailed classification of features can be given based on the circuit complexity of the operation of their calculation, which is specified by a sequence of meta-operations. Let's consider invariant properties for models that were built on the basis of closures.

We will distinguish information properties based on the number of (reachable, achieved, attainable): states, (internally) distinguishable states, paths (cycles), (internally, potentially) distinguishable cycles.

Quantitative measures, their values and sets of their values will first of all be considered as invariants of (semantically closed) structures of semantic space. In order to identify the closest ones when searching in knowledge management problems.

Below are examples of calculating some quantitative characteristics (Fig.1, Fig.2 and Fig.3).

III. ONTOLOGICAL STRUCTURES AND ITS MEASURES

Within the context of task types (kinds) of general classification of problems [2], reflexive transitive closures of operations are important to be able models of information storage systems with some information capacity.

A. State capacity

The state capacity can be expressed with the following features:

• (active information volume) number of states of operational closure (for finite irreflexive closured this measure is reducing);

• (reactive information volume) number of states of reverse-operational closure;

• (real information volume) number of states of maximal reflexive operational subclosure;

• (imaginary information volume) number of states of complement of maximal reflexive operational subclosure to the union of operational and reverse-operational closures.

Reachable (future) vertexes (accessible (active))

$C(x) = \bigvee_z ((A_R)^z \wedge \beta)$, where x is the set of vertexes, A_R – the adjacency matrix of the structure R and β is the indicator vector of x.

Leaved (past) vertexes (released (reactive))

$R(x) = \bigvee_z (((A_R)^T)^z \wedge \beta)$

Selected vertexes (real)

$R(x) \wedge C(x)$

Selected vertexes (imaginary)

$V(x)/\left(R(x) \wedge C(x)\right)$

B. Transition capacity

Let‘s consider the open (acyclic) or closed path’s (cyclic) flows for corresponding open or closed structures and take the following requirements for its flow c_{ij}.

Each edge is associated with a flow (energy) c_{ij}. Each vertex s is associated with a flow $c_s = \sum_{j=1}^{n} c_{sj}$. In addition to the forward flow, the backward flow $c^{-1}{}_{ji}$ is also calculated, $c^{-1}{}_s = \sum_{j=1}^{n} c^{-1}{}_{sj}$. Their differences are equal to

$d_{ij} = c_{ij} - \mathrm{c}^{-1}{}_{ji}$, $d_s = c_s - \mathrm{c}^{-1}{}_s$. Forward (local) amplitude is calculated $p_{ij} = \frac{d_{ij}}{c_i + \mathrm{c}^{-1}{}_i} + \frac{1}{\sum_{j=1}^{V} a_{ij}}$.

$$\sum_{j=1}^{N} c_{ij} = \sum_{j=1}^{N} c_{ji}$$

$$c_{ij} = \frac{\sum_{j=1}^{N} c_{ij}}{\sum_{j=1}^{N} a_{ij}} * a_{ij} \ \ ; c_{ij} * \sum_{j=1}^{N} a_{ij} = a_{ij} * \sum_{j=1}^{N} c_{ij}$$

We have also in the matrix form:

$$A^T * C = (A * 1) \cdot C$$

For the structure on Fig. 2 and its forward flow we have:

$$\begin{cases} c_{11} = c_{12} \\ c_{56} = c_{57} \\ c_{89} = c_{814} \\ c_{1112} = c_{1113} \end{cases}$$

We will find the minimal natural solution.

As the result of the meeting these requirements we obtain the following table of results (see Fig.1 (right), Fig.2, Fig.3).

TABLE I. TABLE OF D YNAMIC STRUCTURE CHARACTERISTICS

Edge number	Flow difference	Forward amplitude	Backward amplitude
0	3	19/32	35/35=1
1	-3	13/32	29/29=1
2	3	35/35=1	35/35=1
3	-3	29/29=1	29/29=1
4	3	35/35=1	19/34
5	-1	15/29	15/34
6	-2	14/29	30/30=1
7	2	34/34=1	34/34=1
8	-2	30/30=1	30/30=1
9	0	16/34=8/17	16/30=8/15
10	2	18/34=9/17	18/35
11	-2	30/30=1	14/30=7/15
12	-2	30/30=1	30/30=1
13	1	17/30	33/33=1
14	-3	13/30	29/29=1
15	1	33/33=1	17/35
16	-3	29/29=1	29/29=1
17	3	35/35=1	35/35=1
18	-3	29/29=1	13/32
19	3	35/35=1	19/32

Analogically, we can get the result for open (unclosed) structure (Fig. 1 (left)).

Each strongly connected structure has a (own) period [14] T which is the GCD of all periods (lengths of simple cycles) in this structure and also has a partition into levels of wave fronts corresponding to this period. The number of these levels will be called the length of the structure $L = T$. The length L of an acyclic structure is the maximum length of the shortest path for two connected vertices. Each (acyclic) structure has a mapping W of the set of numbers of moments of time onto the set of subsets of vertices according to the levels of wave fronts at given moments of time the number of which does not exceed the length and diameter of the structure. Each wave front has energy $E(t) = \sum_{s \in W(t)} c_s$. The amplitude at the top of the wavefront $p_s^t = \frac{c_s}{E(t)}$ is in the interval $[0; 1]$. The average amplitude is inversely proportional to the number of wavefront elements $\frac{E(t)}{|W(t)|}$. The wavefront entropy at time t is expressed by:

$$-\sum_{s \in W(t)} \left(\frac{\left|U_s^{(t)}\right|}{p_s^t} * \ln \frac{\left|U_s^{(t)}\right|}{p_s^t} \right)$$

where $U_s^{(t)}$ is the set of undistinguishable (automorphic) vertices of the vertex s in wavefront t

$$\{s\} \subseteq U_s^{(t)} \subseteq W(t)$$

The set of indistinguishable moments of time is $I_{(t)}$.

$$\{t\} \subseteq I_{(t)} \subseteq Dom(W)$$

Average (arithmetic) entropy of structure:

$$-\frac{1}{T} * \sum_{t=1}^{T} \sum_{s \in W(t)} \left(\frac{\left|U_s^{(t)}\right|}{p_s^t} * \ln \frac{\left|I_{(t)}\right| * \left|U_s^{(t)}\right|}{T * p_s^t} \right)$$

We will call entropy (neg-information) in a strongly connected structure real (elliptic) and also call entropy in an acyclic structure imaginary (hyperbolic).

IV. CONCLUSION

The paper provides an overview of existing approaches to the analysis of semantic space structures and measures on graph structures. The approach to the analysis of semantic space structures based on a knowledge specification model and the use of meta-operations is proposed. Requirements for the characteristics of these structures identifiable as a result of such analysis are formulated. In accordance with them, it is proposed to consider topologically closed structures of semantic space as the basic structures to be analyzed. The transition from static structures to dynamic ones based on the use of meta-operations and closure operators is considered. Quantitative features for the dynamic structures of the semantic space are proposed. These features based on the general classification of problems in order to solve knowledge management problems.

REFERENCES

[1] V. Ivashenko, “Semantic space integration of logical knowledge representation and knowledge processing models” in Open Semantic Technologies for Intelligent Systems (OSTIS-2023), vol. 7, V.V. Golenkov et al, Eds. Minsk: BSUIR, 2023, pp. 95–114.

[2] V. P. Ivashenko, Models for solving problems in intelligent systems. In 2 parts, Part 1: Formal models of information processing and parallel models for solving problems: a tutorial, Minsk: BGUIR, 2020. 79 p. [Modeli resheniya zadach v intellektual’nykh sistemakh. V 2 ch. Ch. 1 : Formal’nye modeli obrabotki informatsii i parallel’nye modeli resheniya zadach : ucheb.-metod. posobie (in Russian)].

[3] V. P. Ivashenko, “Operational semantics of multi-agent knowledge processing systems” in Information Technologies and Systems, Minsk: BGUIR, 2020, pp. 78–79 [Operatsionnaya semantika mnogoagentnykh sistem obrabotki znanii (in Russian)]

[4] M. E. J. Newman, “The mathematics of networks” , The New Palgrave Encyclopedia of Economics. Basingstoke, Palgrave Macmillan, 2007.

[5] C. Aggarwal, K. Subbian, “Evolutionary Network Analysis: A Survey” in ACM Computing Surveys, vol. 47, no.1, 2014.

[6] P. Holme, J. Saramäki, Temporal networks. https://arxiv.org/abs/1108.1780, 2011.

[7] K. M. Carley, 2003, "Dynamic Network Analysis" in Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers, R. Breiger, K. Carley, P. Pattison, Eds. Washington, DC: Committee on Human Factors, National Research Council, National Research Council, pp. 133–145,

[8] H. S. Mortveit, Graph Dynamical Systems – A Mathematical Framework for Interaction-Based Systems, Their Analysis and Simulations,

https://web.archive.org/web/20140903062024/http://www.samsi.info/sites/default/files/samsi-05-dec-08.pdf, 24.09.2023.

[9] V. Ivashenko, “Attributes, scales and measures for knowledge representation and processing models” in Open Semantic Technologies for Intelligent Systems (OSTIS-2019), vol. 4, V. V. Golenkov et al, Eds. Minsk: BSUIR, 2019, pp. 247–250.

[10] R. Milo, S. S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, “Network motifs: simple building blocks of complex networks” in Science, vol. 5594, no. 298, 2002, pp. 824–827.

[11] T. P. Michalak, K. V. Aadithya, P. L. Szczepański, B. Ravindran, N. R. Jennings. “Efficient Computation of the Shapley Valuefor Game-Theoretic Network Centrality” in Journal of Artificial Intelligence Research, vol. 46, 2013, pp. 607–650.

[12] A. J. Alvarez-Socorro, G. C. Herrera-Almarza, L. A. González-Díaz, “Eigencentrality based on dissimilarity measures reveals central nodes in complex networks” in Scientific Reports, vol. 5, no.1, 2015.

[13] M. Benzi, C. Klymko, “A matrix analysis of different centrality measures” in SIAM Journal on Matrix Analysis and Applications, vol. 36, no. 2, 2013. pp. 686–706.

[14] J. Aledo, L. Diaz, S. Martinez, J. Valverde, “On the Periods of Parallel Dynamical Systems” in Complexity, 2017, pp. 1-6.

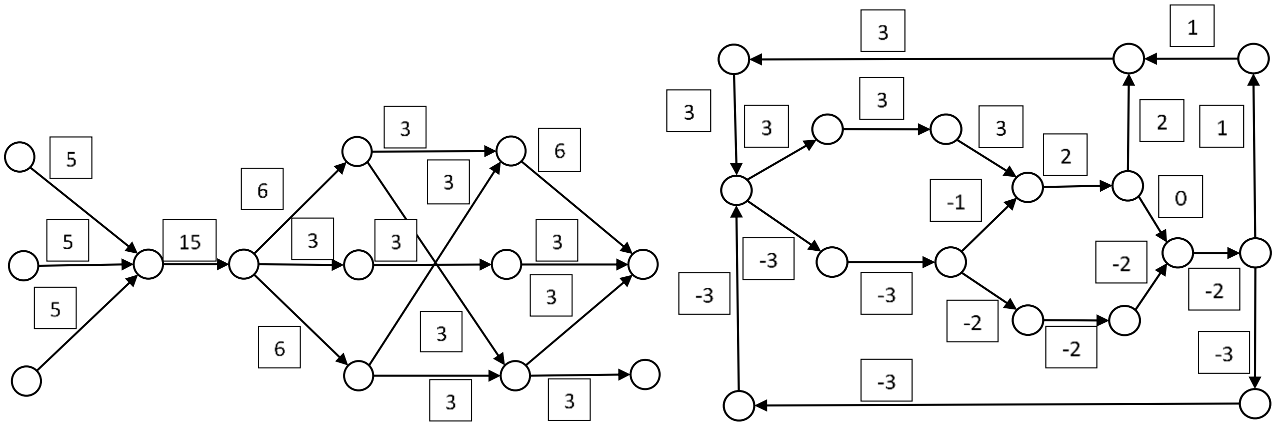

Fig. 1. Examples of open (acyclic) dynamic structure and its flows (left) and closed (strongly connected) dynamic structure and its flow differencies (right)

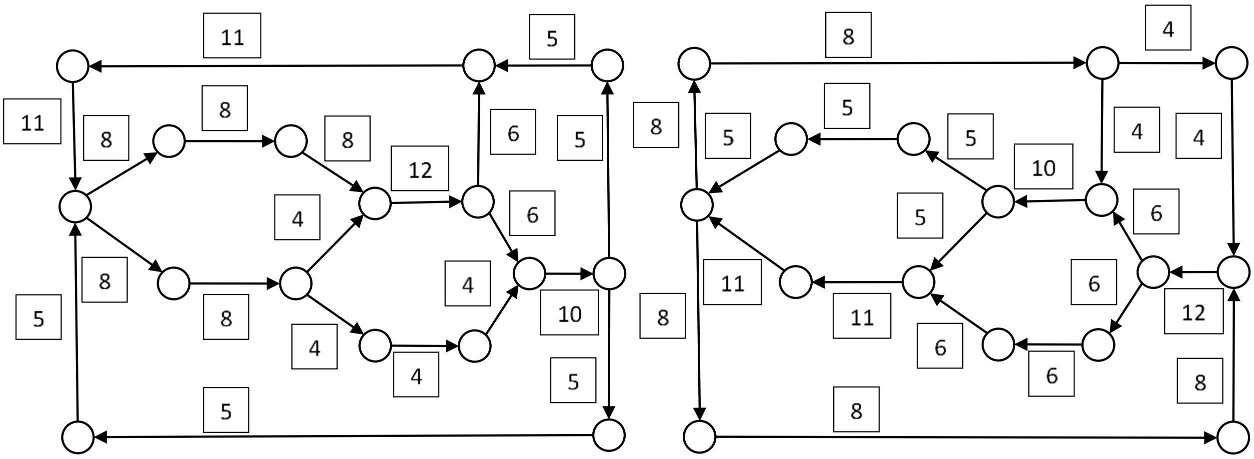

Fig. 2. Example of closed (strongly connected) dynamic structure and its forward and backward flows

$$
A = \begin{matrix}
0 & 3 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0 & 0 & 0 & 0 & 2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -3 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 \\
-3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{matrix}
$$

Fig. 3. The matrix of flow differences

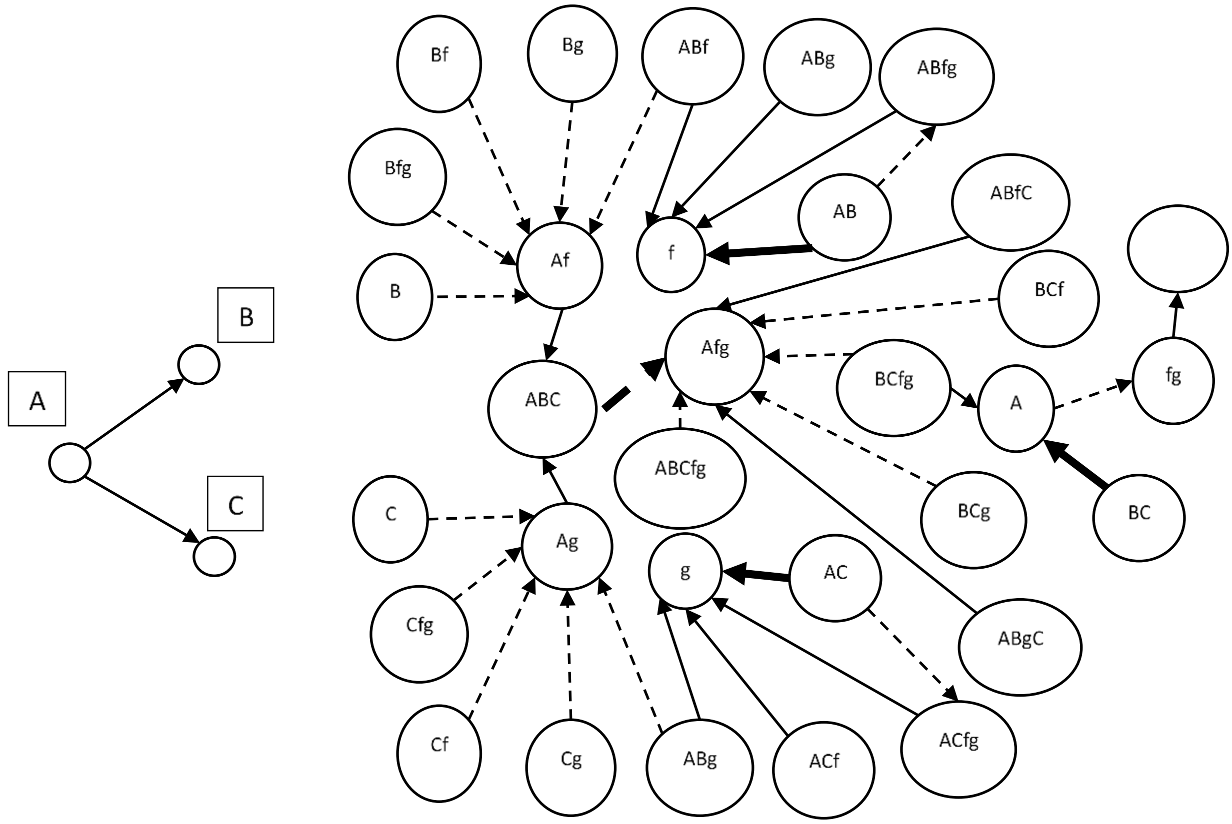

Fig. 4. Example of the symmetric ontological structure (left) and the corresponding sythesized dynamic strucrure of its found elements (right); different arcs denote different operations

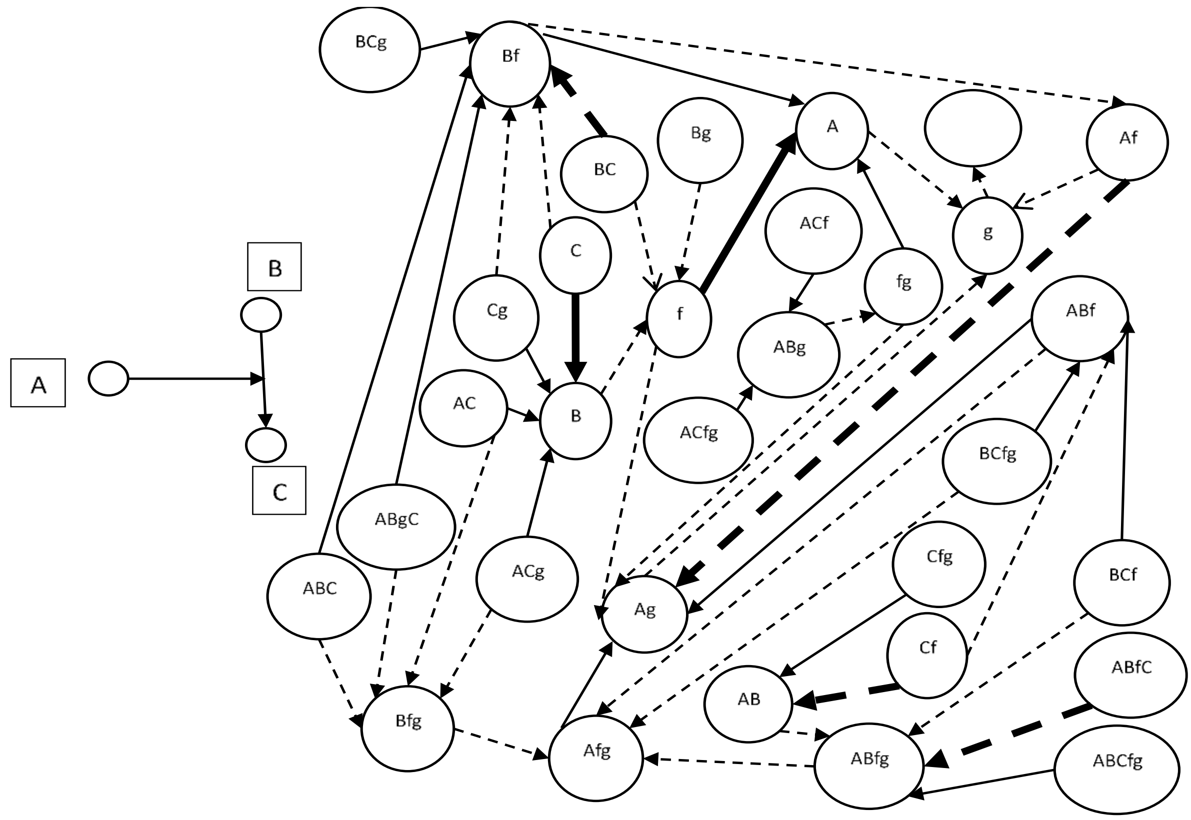

Fig. 5. Example of the asymmetric ontological structure (left) and the corresponding sythesized dynamic strucrure of its found elements (right); different arcs denote different operations