ISSN: **1992-8645** www.jatit.org E-ISSN: **1817-3195**

COMPREHENSIVE METHOD FOR DISCOVERING AND ASSESSING THE RICH-CLUB ORGANIZATION IN MULTIHYPERGRAPHS

ARTEM POTEBNIA
Independent Investigator, Korosten, Ukraine;
ORCID: 0000-0002-8162-5613;
Website: artem.pp.ua; E-mail: potebnia@mail.ua

ABSTRACT

The topological structure of the graph-based models constructed for a wide range of the real-world complex systems is characterized by the clear presence of the rich-club organization. Conceptually, such organization implies the tendency of the most important (in accordance to the prescribed metric) vertices to be tightly interconnected with each other and form the cohesive communities referred to as the rich-clubs. Recently, the rich-club ordering has attracted a considerable attention of investigators due to its impact on the robustness and performance of the modeled system as well as the regime of its functioning. At the same time, the prior studies in this direction are entirely limited to the case of simple graphs. This, in turn, points to the existence of the fundamental research gap associated with the need to develop the method for detecting the rich-club organization in multihypergraphs (i.e. hypergraphs allowing the presence of the parallel hyperedges). With a view to bridging the identified gap, this work introduces the family of the original metrics providing the formal way for determining whether the submultihypergraph induced by the most important nodes of multihypergraph could be properly regarded as its rich-club. The proposed metrics are designed to exhaustively capture the complex nature of relationships established in the considered submultihypergraph and, accordingly, take into account not only the number of its hyperedges but also their cardinality and role in ensuring the connectivity of vertices. Furthermore, the paper elaborates the scheme of normalizing the introduced metrics with respect to the reference ensemble of random multihypergraphs possessing the same sequences of vertex degrees and hyperedge cardinalities as the multihypergraph under investigation. Such normalization allows discovering the intentionally emerged rich-club ordering in multihypergraphs that does not follow merely from the structural restrictions imposed by the local properties of vertices and hyperedges. Finally, the paper illustrates the descriptive potential of the developed method by constructing the multihypergraph-based representation of the scientific co-authorship hypernetwork extracted from the IEEE Xplore database and performing the rigorous experimental analysis of its rich-club organization.

Keywords: *Rich-Club Organization, Rich-Club Coefficient, Multihypergraph, Simple Hypergraph, Multigraph, Co-Authorship Hypernetwork.*

1. INTRODUCTION

The generalization of notions initially introduced for the simple graphs to the more abstract incidence structures serves as one of the most actual and challenging research directions in the field of combinatorial mathematics. The crucial need for producing the new knowledge in this direction stems from the conceptual inadequacy of applying the simple graphs for handling the structure of the hypernetworks defined as embodying the grouping (i.e. non-pairwise) interactions among the homogeneous objects or many-to-many relationship established between the objects of two fundamentally different types (referred to as the actors and entities) [1]. At the same time, the structural organization underlying a number of real-world complex systems is naturally represented in terms of hypernetworks [2]. For example, the non-stellar bodies in the galaxy could be viewed as the objects of the hypernetwork in which they are grouped into the planetary systems associated with the corresponding host stars or star systems. Moreover, the complexity classes could be interpreted as the groupings of the computational problems and, in this sense, serve as forming the hypernetwork on them [3, 4]. In turn, the hypernetworks realizing the many-to-many relationship between the actors and entities commonly occur at the design of the relational databases and are known for the necessity of introducing the auxiliary table in order to ensure the consistency of data. One concrete example is the scientific co-authorship hypernetwork representing the many-to-many relationship established between the articles and their authors [5, 6]. Meanwhile, the modeling of the hypernetworks with preserving their structural information relies on the concept of multihypergraph formulated in the following way:

Definition 1. An arbitrary *multihypergraph* $\hat{H}=\left(V,\tilde{E},\tilde{m}\right)$ is represented by the vertex set V equipped with the multiset $\left(\tilde{E},\tilde{m}\right)$ composed of the (possibly repeating) non-empty subsets of V referred to as *hyperedges*. Here the *underlying set* $\tilde{E}\subseteq \mathrm{P}\left(V\right)\setminus\{\varnothing\}$, where $\mathrm{P}\left(V\right)$ denotes the power set constructed on V, comprises all distinct (i.e. given by the non-identical subsets of vertices) hyperedges of $\hat{H}$. In turn, $\tilde{m}:\tilde{E}\to\mathbf{N}^+$ is the *multiplicity function* mapping every $\tilde{e}\in\tilde{E}$ into the number of its copies (standing for the *parallel hyperedges*) existing in $\left(\tilde{E},\tilde{m}\right)$. Remark that an arbitrary vertex $v\in V$ and hyperedge $\tilde{e}\in\left(\tilde{E},\tilde{m}\right)$ in $\hat{H}$ are referred to as *incident* if and only if $v\in\tilde{e}$. In this context, any *simple hypergraph* $H=\left(V,\tilde{E}\right)$ is represented by the multihypergraph with the multiplicity function fixed to the constant $\tilde{1}:\tilde{E}\to\{1\}$. For its part, any *multigraph* $\hat{G}=\left(V,E,m\right)$ is defined as the multihypergraph with the multiset of hyperedges $\left(E,m\right)$ having the underlying set $E\subseteq\left\{\{v,w\}\middle|\left(v,w\in V\right)\wedge\left(v\neq w\right)\right\}$ at an arbitrary multiplicity function $m:E\to\mathbf{N}^+$. For clarity, the items comprising $\left(E,m\right)$ are additionally mentioned as *edges*. Finally, any *simple graph* $G=\left(V,E\right)$ is represented by the multigraph with the constant multiplicity function $1:E\to\{1\}$.

With a view to supporting the subsequent manipulations by the introduced mathematical structures, let us adopt the notations $\Gamma_{\hat{H}}$, Γ_H, $\Gamma_{\hat{G}}$, and Γ_G for the classes containing all possible instances of multihypergraphs, simple hypergraphs, multigraphs, and simple graphs, respectively. Notice that these classes meet the following system of relationships: $\Gamma_H\subset\Gamma_{\hat{H}}$, $\Gamma_{\hat{G}}\subset\Gamma_{\hat{H}}$, and $\Gamma_G=\Gamma_H\bigcap\Gamma_{\hat{G}}$.

Conceptually, the representation of the structure underlying any hypernetwork in terms of the multihypergraph instance involves representing every particular grouping of its objects or collection of actors associated with the concrete entity by the corresponding hyperedge. For example, in the multihypergraph-based model of the co-authorship hypernetwork, all authors are putted in one-to-one correspondence with vertices, while every article is reflected by exactly one hyperedge incident to all nodes depicting its authors. Remark that the descriptive power of multihypergraphs in the lossless structural representation of hypernetworks follows from the absence of any restrictions on the cardinality and uniqueness of their hyperedges [7].

At the same time, the complex combinatorial nature of multihypergraphs significantly complicates the extraction of their topological properties (i.e. such properties that are inherent to the overall class of multihypergraphs obtained from the examined one by relabeling its vertices). Thereby, the researchers are forced to use less accurate models of simple graphs when investigating the structure of hypernetworks, which could potentially result in deducing the wrong conclusions [8]. In this light, the further advancement of knowledge regarding the structure of complex systems requires generalizing the existing approaches to mining the topology of simple graphs to the wider class of multihypergraphs in order to provide the opportunity to use the more adequate models in the process of investigation.

Conceptually, the problem of mining the multihypergraph's topology lies in detecting the presence or absence of the special substructures revealing the important properties of the modeled system. Serving as one of such substructures, the rich-class organization implies the tendency of the most important (according to the prescribed metric) vertices to form the tightly interconnected "elite" communities [9, 10]. The prior works were focused on studying such organization only for the particular case of simple graphs based on the formal instruments of the rich-club coefficient given by the density of the subgraph induced by the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

subset of the most important vertices and its normalized version [11, 12]. Remark that the density of any simple graph $G=(V,E)\in\Gamma_G$ is defined as the ratio between the actual number of its edges and their maximum possible number on $|V|$ nodes, i.e. $\rho(G)=2|E|/(|V|(|V|-1))$. At the same time, the rich-club organization in the more general incidence structures of multihypergraphs still lacks the formal instruments for assessment largely due to the ambiguity of determining the tightness of interconnections among their vertices, which clearly points to the presence of the research gap. In turn, ***the objective of this work*** lies in filling the identified gap by introducing the family of coefficients ensuring the detection and evaluation of the rich-club organization in multihypergraphs tailored to their complex combinatorial nature. Moreover, the paper illustrates the application of the proposed coefficients for mining the topology of the multihypergraph-based model constructed for the real-world co-authorship hypernetwork.

2. CONCEPTUAL APPROACHES AND FORMAL METRICS FOR CHARACTERIZING THE TIGHTNESS OF INTERCONNECTIONS AMONG THE VERTICES OF MULTIHYPERGRAPHS

The generalization of the simple graph density $\rho(G)$ to the $\Gamma_{\hat{H}}$ class multihypergraphs is precluded by the absence of the upper bound on the number of their hyperedges at the finiteness of the vertex set. This issue along with an arbitrary cardinality of each hyperedge acts as the fundamental impediment to the detection and proper assessment of the cohesive communities in such incidence structures. In fact, the multigraphs comprising the subclass $\Gamma_{\hat{G}}$ also require formulating the conceptual substitute of density due to their potential capability of accommodating the countably infinite number of edges on any pair of vertices. With a view to addressing this challenge, let us introduce the following topological metric that, by adequately capturing the nature of both multihypergraphs and multigraphs, provides the most straightforward quantitative description for the tightness of interconnections among their nodes:

Definition 2. The *hyperedge-to-vertex ratio* $\sigma(\hat{H})$ of any non-null (i.e. containing at least one node) multihypergraph $\hat{H}=(V,\tilde{E},\tilde{m})$ lies within the range $[0,+\infty)$ and is formally given by $\sigma(\hat{H})=|(\tilde{E},\tilde{m})|/|V|$, where $|(\tilde{E},\tilde{m})|=\sum_{\tilde{e}\in\tilde{E}}\tilde{m}(\tilde{e})$ denotes the cardinality of the multiset $(\tilde{E},\tilde{m})$.

For brevity, the ratio $\sigma(\hat{G})$ associated with the multigraph $\hat{G}$ is also called without the prefix "hyper". At the same time, the loop hyperedges (i.e. having the cardinality of one), by definition, are deprived of any role in linking the multihypergraph's nodes. For example, in the context of the above-discussed model of the scientific co-authorship hypernetwork, all loops serve as the representations of the solely authored articles that do not contribute to the formation of the inter-researcher collaboration scheme. This consideration points to the crucial need for the appropriate correction of the metric $\sigma(\hat{H})$. Accordingly, let us define the loop removal function $\psi:\Gamma_{\hat{H}}\to\tilde{\Gamma}_{\hat{H}}$ mapping an arbitrary input multihypergraph $\hat{H}=(V,\tilde{E},\tilde{m})\in\Gamma_{\hat{H}}$ into the corresponding loopless multihypergraph $\hat{H}'=(V,\tilde{E}',\tilde{m}|_{\tilde{E}'})\in\tilde{\Gamma}_{\hat{H}}$ equipped with the modified multiset of hyperedges $(\tilde{E}',\tilde{m}|_{\tilde{E}'})$ given by the underlying set $\tilde{E}'=\tilde{E}\setminus\{\{v\}|v\in V\}$ and the multiplicity function $\tilde{m}|_{\tilde{E}'}:\tilde{E}'\to\mathbf{N}^+$ obtained by restricting $\tilde{m}$ to the domain $\tilde{E}'\subseteq\tilde{E}$. Remark that while being surjective, the function ψ is non-injective since all multihypergraph instances comprising its image $\tilde{\Gamma}_{\hat{H}}\subset\Gamma_{\hat{H}}$ are required to include exclusively the non-loop hyperedges. This background, in turn, allows introducing the derivative metric $\tilde{\sigma}(\hat{H})=\sigma(\psi(\hat{H}))$ intended to estimate the concentration of hyperedges in the multihypergraph $\hat{H}$ at the discounted influence of loops.

Intuitively, the higher value of $\tilde{\sigma}(\hat{H})$ indicates that the vertices of $\hat{H}$ share the larger number of distinct hyperedges and, from this viewpoint, are in the more cohesive relationship. Notice that the usefulness of such approach consists in the interpretation of hyperedges as the separate indivisible entities with the preservation of their original conceptual sense, which is necessary in

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

many contexts. For example, in the case of the multihypergraph $\hat{H}$ abstracting the scientific co-authorship hypernetwork, the metric $\tilde{\sigma}(\hat{H})$ characterizes the cohesiveness of the considered group of researchers based on the number of actually existing multi-author works serving as the documentary evidences of collaboration.

However, while being entirely focused on treating the hyperedges as conceptually self-sustainable objects, the quantitative metric $\tilde{\sigma}(\hat{H})$ completely ignores the heterogeneity of their contributions to joining the multihypergraph's nodes. In particular, the number of vertex pairs linked by the non-loop hyperedge grows with the increase in its cardinality $k \geq 2$ as the quadratic function $k(k-1)/2$. These remarks stimulate us to consider the following mathematical structure characterizing the whole multihypergraph by summarizing the local configurations of all its hyperedges:

Definition 3. The *hyperedge cardinality sequence* associated with an arbitrary non-empty (i.e. such that $\tilde{E} \neq \varnothing$) multihypergraph $\hat{H} = (V, \tilde{E}, \tilde{m})$ is represented in the form $\Omega(\hat{H}) = (\omega_1(\hat{H}), \omega_2(\hat{H}), \ldots, \omega_{|V|}(\hat{H}))$, where each item $\omega_i(\hat{H})$ gives the number of hyperedges with the cardinality of i in $(\tilde{E}, \tilde{m})$, i.e.

$$\omega_i(\hat{H}) = \sum_{\tilde{e} \in \tilde{E}(i)} \tilde{m}(\tilde{e}); \quad \tilde{E}(i) = \left\{ \tilde{e} \,\middle|\, (\tilde{e} \in \tilde{E}) \wedge (|\tilde{e}| = i) \right\}.$$

The examination of the sequence $\Omega(\hat{H})$ allows making fruitful conclusions regarding the dominating patterns of interaction among the vertices of $\hat{H}$. In particular, the ratio $\omega_i(\hat{H})/|V|$ calculated for the multihypergraph $\hat{H}$ modeling the co-authorship hypernetwork indicates the likelihood of having i authors listed on the randomly picked article and, in this way, strengthens understanding the collective behavior of researchers. Based on this background, let us introduce the scalar metric $\theta(\hat{H})$ derived from both $|V|$ and $\Omega(\hat{H})$ according to the following expression:

$$\theta(\hat{H}) = \frac{1}{|V|} \sum_{\substack{\tilde{e} \in \tilde{E} \\ |\tilde{e}| \geq 2}} \binom{|\tilde{e}|}{2} \tilde{m}(\tilde{e}) = \frac{1}{|V|} \sum_{i=2}^{|V|} \binom{i}{2} \omega_i(\hat{H}),$$

where the symbol $\binom{a}{b}$ denotes the binomial coefficient defined as $\frac{a!}{b!(a-b)!}$ if $a \geq b \geq 0$ and zero otherwise.

Conceptually, the value of this metric could be interpreted as the cumulative number of the direct (i.e. established through only one hyperedge) pairwise connections among the nodes of $\hat{H}$ divided by the cardinality of the vertex set V. In this regard, $\theta(\hat{H})$ serves as the useful additional criterion for quantifying the cohesiveness of the multihypergraph $\hat{H}$ in terms of the total vertex linking effect produced by all its hyperedges. For example, if $\hat{H}$ describes the structural design of the co-authorship hypernetwork, the value of $\theta(\hat{H})$ reflects the tightness of relationships among the researchers from the viewpoint of the number of times when two of them appear as authors on the same article. Moreover, the meaning enclosed in the metric $\theta(\hat{H})$ could be vividly illustrated based on the transition to the projection of $\hat{H}$ on the class of multigraphs $\Gamma_{\hat{G}}$ driven by the following function:

Definition 4. The *hyperedge splitting function* $\varphi : \Gamma_{\hat{H}} \to \Gamma_{\hat{G}}$ transforms any argument multihypergraph $\hat{H} = (V, \tilde{E}, \tilde{m}) \in \Gamma_{\hat{H}}$ into the output multigraph $\hat{G}[\hat{H}] = (V, E^*, m^*) \in \Gamma_{\hat{G}}$ constructed on the same vertex set V and holding the multiset of edges (E^*, m^*) given by the underlying set $E^* = \left\{ \{v, w\} \,\middle|\, \Phi(\tilde{E}, \{v, w\}) \neq \varnothing \right\}$ along with the multiplicity function defined as $m^*(\{v, w\}) = \sum_{\tilde{e} \in \Phi(\tilde{E}, \{v, w\})} \tilde{m}(\tilde{e})$ for all $\{v, w\} \in E^*$. In these expressions, the auxiliary notation $\Phi(\tilde{E}, \{v, w\}) = \left\{ \tilde{e} \,\middle|\, (\tilde{e} \in \tilde{E}) \wedge (\{v, w\} \subseteq \tilde{e}) \right\}$ stands for the subset of $\tilde{E}$ composed of all its

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

hyperedges $\tilde{e}$ serving as the supersets for the given edge $\{v, w\}$. For clarity, $\hat{G}\left[\hat{H}\right]$ is referred to as the *multigraph approximation* of $\hat{H}$.

In simple terms, the number of edges joining any two vertices of $\hat{G}\left[\hat{H}\right]$ is determined as the number of hyperedges shared by them in $\hat{H}$. Notice that the function φ is non-invertible since, in general, the structure of hyperedges existing in $\hat{H}$ as well as the cardinality of its multiset $\left(\tilde{E}, \tilde{m}\right)$ could not be unambiguously recovered from the information encapsulated in $\hat{G}\left[\hat{H}\right]$. At the same time, the number of edges in $\hat{G}\left[\hat{H}\right]$ serves as the upper bound on the number of non-loop hyperedges in $\hat{H}$. The metric $\theta\left(\hat{H}\right)$, in turn, represents the topological property of $\hat{G}\left[\hat{H}\right]$ and comes down to its edge-to-vertex ratio, i.e. $\theta\left(\hat{H}\right) = \sigma\left(\varphi\left(\hat{H}\right)\right)$.

Conceptually, the non-loop hyperedges comprising the underlying set $\tilde{E}$ associated with the multihypergraph $\hat{H}$ could be viewed as embodying the collective groupings of actors abstracted by the vertices of V that are established at the formation of entities reflected by the items of the multiset $\left(\tilde{E}, \tilde{m}\right)$. For example, revisiting the model of the co-authorship hypernetwork, the hyperedges of $\tilde{E} \setminus \left\{\{v\} \middle| v \in V\right\}$ depict the distinct collectives of researchers each of which includes two or more members and plays a role of the unordered author list for at least one document within the considered corpus. In turn, the metric relying on the number of such collective groupings per vertex as the measure for the tightness of $\hat{H}$ is naturally expressed as $\tilde{\tau}\left(\hat{H}\right) = \tau\left(\psi\left(\hat{H}\right)\right)$, where $\tau\left(\hat{H}\right) = \left|\tilde{E}\right| / |V|$.

Definition 5. The *hyperedge deduplication function* $\chi : \Gamma_{\hat{H}} \to \Gamma_H$ produces the *simple hypergraph approximation* $H\left[\hat{H}\right] = \left(V, \tilde{E}\right) \in \Gamma_H$ for an arbitrary initial multihypergraph $\hat{H} = \left(V, \tilde{E}, \tilde{m}\right) \in \Gamma_{\hat{H}}$ by replacing its multiplicity function $\tilde{m}$ with the constant $\tilde{1} : \tilde{E} \to \{1\}$.

Since $\tau\left(\hat{H}\right) = \sigma\left(\chi\left(\hat{H}\right)\right)$ and $\tilde{\tau}\left(\hat{H}\right) = \tilde{\sigma}\left(\chi\left(\hat{H}\right)\right)$, both $\tau\left(\hat{H}\right)$ and $\tilde{\tau}\left(\hat{H}\right)$ could be viewed as characterizing the topology of $\hat{H}$ through its projection on the class of simple hypergraphs Γ_H in the form of the approximation $H\left[\hat{H}\right]$. At the same time, the number of hyperedges contained in any simple hypergraph $H = \left(V, \tilde{E}\right) \in \Gamma_H$ is bounded above by $\left|\mathrm{P}(V)\right| - 1$, where $\left|\mathrm{P}(V)\right| = 2^{|V|}$. It is worth noticing that the given bound takes into account the forbiddance of the empty hyperedge deprived of the incident vertices. This discussion clearly shows that the simple hypergraphs, in contrast to both multihypergraphs and multigraphs, could be properly characterized by the concept of density adapted in the following original manner:

Definition 6. The *density* $\lambda(H)$ of any non-null simple hypergraph $H = \left(V, \tilde{E}\right)$ is formally given by the expression $\lambda(H) = \left|\tilde{E}\right| / \left(2^{|V|} - 1\right)$ and indicates the actual number of hyperedges existing in H as the fraction of their maximum possible number on $|V|$ nodes in the Γ_H class incidence structures. In turn, the *loopless density* $\tilde{\lambda}(H)$ defined for every H having at least two vertices additionally implies the exclusion of the loop hyperedges from the consideration and is calculated as $\tilde{\lambda}(H) = \left|\tilde{E} \setminus \left\{\{v\} \middle| v \in V\right\}\right| / \left(2^{|V|} - |V| - 1\right)$.

The denominator of the above-stated expression for $\tilde{\lambda}(H)$ stems directly from the ability of H to accommodate up to $|V|$ loops anchored to the distinct nodes. In order to keep the conceptual consistency, let us define the *complete simple hypergraph* KH_n on $n \in \mathbf{N}^+$ vertices as having the density $\lambda(KH_n) = 1$. Notice that the topological organization of KH_n is entirely determined by n, while the number of hyperedges incident to each its node equals $\sum_{i=1}^{n} \binom{n-1}{i-1}$, where every summand represents the number of such hyperedges with the cardinality of i.

By definition, both introduced density metrics $\lambda(H)$ and $\tilde{\lambda}(H)$ take values within the range $[0,1]$. Apart from that, compared to the ratios $\sigma(H)$ and $\tilde{\sigma}(H)$, they provide the conceptually different and non-trivially related viewpoint on the concentration of hyperedges in H by adopting another reference levels (placed in the denominators of their formulas). These considerations eventually prompt us to define the metrics $\mu(\hat{H})=\lambda(\chi(\hat{H}))$ and $\tilde{\mu}(\hat{H})=\tilde{\lambda}(\chi(\hat{H}))$ characterizing the density of the approximation $H[\hat{H}]$ constructed for the multihypergraph $\hat{H}$ and, thereby, providing an additional insight into its topology. Moreover, the value of $\tilde{\mu}(\hat{H})$ could be interpreted as the percentage of all possible collective groupings constructed from the nodes of $\hat{H}$ that are realized in the hyperedges of $\tilde{E}$. In this regard, $\tilde{\mu}(\hat{H})$ represents an important instrument supplementing $\tilde{\tau}(\hat{H})$ in reporting on the cooperative behavior of the multihypergraph's vertices.

Another fundamental approach for assessing the cohesiveness of the multihypergraph $\hat{H}$ implies mining its projection on the class of simple graphs Γ_G representing the result of applying the function composition $\chi\circ\varphi:\Gamma_{\hat{H}}\to\Gamma_G$ (or, equivalently, $\chi\circ\varphi\circ\chi:\Gamma_{\hat{H}}\to\Gamma_G$) to $\hat{H}$ and defined in the following way:

Definition 7. The *simple graph approximation* $G[\hat{H}]=(V,E^*)\in\Gamma_G$ of an arbitrary multihypergraph $\hat{H}=(V,\tilde{E},\tilde{m})\in\Gamma_{\hat{H}}$ is constructed by replacing its multiset $(\tilde{E},\tilde{m})$ with the set of edges $E^*=\{\{v,w\}|\exists\tilde{e}\in\tilde{E}[\{v,w\}\subseteq\tilde{e}]\}$ (i.e. two nodes are adjacent in $G[\hat{H}]$ only if they have at least one common hyperedge in $\hat{H}$).

Despite having more reduced descriptive potential compared to both $\hat{G}[\hat{H}]$ and $H[\hat{H}]$, the approximation $G[\hat{H}]$ encapsulates in the easy-to-process form the basic properties of the hyperpath structure underlying the interaction of vertices in $\hat{H}$. With a view to establishing the formal background needed for understanding these properties, let us define the *hyperwalk* in $\hat{H}$ as the ordered sequence $(v_b,\tilde{e}_1,v_1,\tilde{e}_2...,v_{k-1},\tilde{e}_k,v_e)$ composed of the alternating nodes and hyperedges satisfying the conditions $v_b\in\tilde{e}_1$, $v_e\in\tilde{e}_k$, and $v_i\in\tilde{e}_i\bigcap\tilde{e}_{i+1}$ for every $i\in\{1,2,...,k-1\}$. Here v_b and v_e represent the terminating vertices of hyperwalk, while its length k is determined as the number of visited hyperedges. The *hyperpath*, for its part, is given by the hyperwalk traversing the pairwise distinct nodes $v_b,v_1,...,v_{k-1},v_e$ (and, thereby, passing exclusively through the non-loop hyperedges). Remark that under such definition, the hyperpaths in the incidence structures belonging to $\Gamma_{\hat{H}}\setminus\Gamma_{\hat{G}}$ could visit the duplicating hyperedges. Moreover, for clarity of discussion, the hyperwalks and hyperpaths in the $\Gamma_{\hat{G}}$ class multigraphs are referred to simply as walks and paths. These preliminary concepts allow defining the *topological distance* $d_{\hat{H}}(v,w)$ between the distinct nodes v and w in $\hat{H}$ as the length of the shortest hyperpath terminating in them or infinity if such hyperpath does not exist. Additionally, to ensure the identity of indiscernibles, let the topological distance $d_{\hat{H}}(v,v)$ from any vertex v in $\hat{H}$ to itself be equal to zero. With respect to the metric $d_{\hat{H}}:V\times V\to\mathbf{N}$ defined in such way, the multihypergraph $\hat{H}$ could be viewed as inducing the metric space $(V,d_{\hat{H}})$ on its vertex set V. In turn, let us introduce the *distance matrix* $\mathbf{D}(\hat{H})$ collecting the topological distances between all pairs of nodes in $\hat{H}$ and, thereby, providing the full specification of its space $(V,d_{\hat{H}})$. Taking into account that the vertices of $\hat{H}$ are labeled as $V=\{v_1,v_2,...,v_n\}$, $\mathbf{D}(\hat{H})$ is given in the form of the symmetric $n\times n$ array filled with the entries $[\mathbf{D}(\hat{H})]_{ij}=d_{\hat{H}}(v_i,v_j)$ and zeros on the main diagonal.

At the same time, for every hyperpath $(v_b,\tilde{e}_1,v_1,\tilde{e}_2...,v_{k-1},\tilde{e}_k,v_e)$ in $\hat{H}$, $G[\hat{H}]$ should contain the corresponding path

$(v_b, e_1', v_1, e_2' ..., v_{k-1}, e_k', v_e)$ passing through the same nodes in the identical order and including the edges $e_i' \subseteq \tilde{e}_i$ for each $i \in \{1,2,...,k\}$, and vice versa. Finally, these considerations along with the fact that the presence of the parallel hyperedges does not affect the length of the shortest hyperpaths lead to the conclusion that the distance matrix $\mathbf{D}(\hat{H})$ serves as the invariant under the reductions to the approximations $G[\hat{H}]$, $\hat{G}[\hat{H}]$, and $H[\hat{H}]$, i.e.

$$\mathbf{D}(\hat{H}) = \mathbf{D}(G[\hat{H}]) = \mathbf{D}(\hat{G}[\hat{H}]) = \mathbf{D}(H[\hat{H}]).$$

Meanwhile, the restrictions imposed on the structure of the simple graphs allow deducing the following fundamental statement regarding the sensitivity of their distance matrices to the elementary modification of the edge set:

Theorem 1. Considering an arbitrary non-empty simple graph $G = (V, E) \in \Gamma_G$ on n vertices, let the operation of removing any single edge $\overline{e} \in E$ from G result in increasing the topological distance between $f(G, \overline{e})$ pairs of its nodes. Then $f(G, \overline{e})$ is sharply bounded (i.e. in such way that there does not exist any greater lower bound or smaller upper bound) as follows:

$$1 \le f(G, \overline{e}) \le \begin{cases} n^2/4 & \text{if} \quad n \text{ is even;} \\ (n^2 - 1)/4 & \text{if} \quad n \text{ is odd.} \end{cases}$$

▲ For convenience, let us adopt the auxiliary notations $\overline{e} = \{v_s, v_t\}$ for the endpoints of the deleted edge and $G - \overline{e} = (V, E \setminus \{\overline{e}\})$ for the resulting graph containing the remaining edges. With a view to providing the formal proof for the upper bound on $f(G, \overline{e})$, let us without losing the generality associate the edge $\overline{e}$ with the induced bisection of the vertex set V of the initial graph G into the disjoint non-empty subsets defined as follows:

$$S = \{v_s\} \bigcup \{v \,|\, (v \in V \setminus \{v_s, v_t\}) \wedge ((d_G(v, v_s) \le d_G(v, v_t)) \vee (d_G(v, v_t) = \infty))\}; \quad T = V \setminus S.$$

In more plain words, $v_s \in S$ and $v_t \in T$, while the division of all other nodes is based on their global topological position with respect to the endpoints of $\overline{e}$. In particular, $S \setminus \{v_s\}$ contains the vertices that are located closer to v_s than to v_t or are equidistant or unreachable from them, while the nodes included in $T \setminus \{v_t\}$ have strictly smaller distance to v_t than to v_s. Remark that for any bisection constructed in such manner, the subgraph $G_T = (T, E_T)$ equipped with the edge set $E_T = E \bigcap \{\{v, w\} \,|\, (v \in T) \wedge (w \in T)\}$ is connected (i.e. its distance matrix is deprived of the infinite entries). Furthermore, let us define the bisection of V induced by $\overline{e}$ as balanced if $|S| = |T|$ for even n and $|S| = |T| \pm 1$ for odd n.

Conceptually, the removal of $\overline{e}$ from G affects the topological distance between two its distinct nodes only if $\overline{e}$ lies on the shortest path between them (necessary condition) and there does not exist any alternative shortest path that joins these nodes bypassing $\overline{e}$ (sufficient condition). In order to derive the restriction on the maximum possible number of vertex pairs complying with these conditions, let us consider an arbitrary shortest path $P = (v, ..., v_s, \overline{e}, v_t, ..., w)$ linking the nodes $v, w \in V$ in G with the involvement of $\overline{e}$. Due to the property of optimal substructure, any consecutive subsequence of P terminating in vertices represents the nested shortest path in G, which implies that $d_G(v, v_t) = d_G(v, v_s) + 1$ and $d_G(w, v_s) = d_G(w, v_t) + 1$. These observations indicate that the shortest path including $\overline{e}$ could exist exclusively between a pair of nodes belonging to $\Theta = \{\{v, w\} \,|\, (v \in S) \wedge (w \in T)\}$. Moreover, $\overline{e}$ lies on the shortest paths between all $|\Theta| = |S| \times |T|$ such vertex pairs only if both topological configuration of G and position of $\overline{e}$ satisfy two mandatory requirements. Firstly, the subgraph $G_S = (S, E_S)$ having the edge set $E_S = E \bigcap \{\{v, w\} \,|\, (v \in S) \wedge (w \in S)\}$ should be connected (which entails the connectedness of the whole G). Secondly, playing a role of bridge, $\overline{e}$ should be the only edge between the vertices of S

Journal of Theoretical and Applied Information Technology
15th December 2018. Vol.96. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and T (since the presence of any additional edge $e'=\{v'_s,v'_t\}$ such that $v'_s\in S$ and $v'_t\in T$ would inevitably exclude $\overline{e}$ from the participation in the shortest path at least between v'_s and v'_t). Remark that the last requirement directly precludes the existence of at least one walk linking the nodes of S and T without the traversal of $\overline{e}$. Eventually, this discussion allows concluding that if all vertex pairs $\{v,w\}\in\Theta$ meet the necessary condition for having the mutual distance $d_G(v,w)<d_{G-\overline{e}}(v,w)$, then they also satisfy the sufficient one.

Accordingly, the proof of the upper bound on $f(G,\overline{e})$ comes down to the maximization of $|S|\times|T|$ at an arbitrary fixed n. Let us confine the consideration area to even n and express $|S|$ and $|T|$ as $n/2+x$ and $n/2-x$, where $x\in\{0,\pm1,...,\pm(n/2-1)\}$. Under such representation, the product $|S|\times|T|$ depends on x as $(n/2+x)(n/2-x)=n^2/4-x^2$ and reaches the peak value of $n^2/4$ only at the balanced bisection corresponding to $x=0$. By analogy, the case of odd n could be handled by writing $|S|$ and $|T|$ as $(n+1)/2+y$ and $(n-1)/2-y$ for $y\in\{-(n-1)/2,...,-1,0,1,...,(n-3)/2\}$. In this light, $|S|\times|T|$ takes the form of $((n+1)/2+y)((n-1)/2-y)=(n^2-1)/4-y^2-y$ and maximizes to $(n^2-1)/4$ at the balanced bisections associated with $y=0$ and $y=-1$.

The positivity of the lower bound on $f(G,\overline{e})$ could be proved in the trivial way by taking into account that the exclusion of $\overline{e}$ from G inevitably leads to the increase in the topological distance between the nodes v_s and v_t at least by one edge due to the inevitable loss of their adjacency, i.e. $d_{G-\overline{e}}(v_s,v_t)\ge d_G(v_s,v_t)+1$. Remark that, by definition, $d_G(v_s,v_t)=1$. Moreover, taking into account the optimal substructure of the shortest paths, $f(G,\overline{e})=1$ if and only if, for each vertex $u\in V\setminus\{v_s,v_t\}$, $\{u,v_s\}\in E$ implies $\{u,v_t\}\in E$ or the existence of such $l\in V\setminus\{v_s,v_t,u\}$ that $\{u,l\},\{l,v_t\}\in E$, while $\{u,v_t\}\in E$, for its part, implies $\{u,v_s\}\in E$ or the existence of such $l'\in V\setminus\{v_s,v_t,v\}$ that $\{u,l'\},\{l',v_s\}\in E$. On the other hand, $f(G,\overline{e})\ge2$ if and only if there exists such node that, being adjacent to only one endpoint of $\overline{e}$, is linked to its another endpoint by only one two-length path (traversing $\overline{e}$).

These observations allow claiming that $f(G,\overline{e})=1$ at the removal of every $\overline{e}$ from any such G that contains at least $g(n)=n(n-1)/2-(n-3)$ edges. The formal proof for this claim could be elegantly constructed by contradiction. Let us start by assuming the opposite, i.e. that G with $g(n)$ or more edges could have at least one $\overline{e}$ giving $f(G,\overline{e})\ge2$. Then any such $\overline{e}$ should imply the non-emptiness of the union $P\bigcup Q$, where:

$$P=\{p|(p\in V\setminus\{v_s,v_t\})\wedge(\{p,v_s\}\in E)\wedge(d_{G-\overline{e}}(p,v_t)>2)\};$$

$$Q=\{q|(q\in V\setminus\{v_s,v_t\})\wedge(\{q,v_t\}\in E)\wedge(d_{G-\overline{e}}(q,v_s)>2)\}.$$

Obviously, the condition $d_{G-\overline{e}}(p,v_t)>2$ could be reduced to the system of $n-2$ constraints given by $\{p,v_t\}\notin E$ and $E\bigcap E(p,r,v_t)\neq E(p,r,v_t)$, where $E(p,r,v_t)=\{\{p,r\},\{r,v_t\}\}$, for every $r\in V\setminus\{v_s,v_t,p\}$. Here each pair of edges $E(p,r,v_t)$ represents the two-length path between p and v_t that avoids $\overline{e}$ and involves r as the intermediate vertex. Due to the need for satisfying all these constraints, the inclusion of at least one node in P requires G to have $n-2$ or more missed edges out of $n(n-1)/2$ maximum possible ones. By analogy, the condition $d_{G-\overline{e}}(q,v_s)>2$ could be transformed into the similar system of $n-2$ constraints imposed on the edge set, which allows deducing that the non-emptiness of Q also requires the non-adjacency of at least $n-2$ vertex

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

pairs in G. Consequently, the presence of $g(n)$ or more edges in G inevitably leads to $P = Q = \varnothing$ for every $\overline{e}$. This conclusion, eventually, reduces to absurdum the assumption regarding the satisfaction of the condition $f(G,\overline{e}) \geq 2$ by at least one $\overline{e}$ in any G having not less than $g(n)$ edges and, thereby, proves the statement claimed at the beginning of the previous paragraph. ▼

In the context of the most general class $\Gamma_{\hat{H}}$, the operation of removing any single hyperedge $\overline{\tilde{e}} \in \tilde{E}$ from an arbitrary non-empty multihypergraph $\hat{H} = (V,\tilde{E},\tilde{m}) \in \Gamma_{\hat{H}}$ implies the exclusion of one item from its multiset $(\tilde{E},\tilde{m})$ and is formally defined as the transformation of $\hat{H}$ into

$$\hat{H} - \overline{\tilde{e}} = \begin{cases} \left(V,\tilde{E} \setminus \{\overline{\tilde{e}}\},\tilde{m}\big|_{\tilde{E}\setminus\{\overline{\tilde{e}}\}}\right) & \text{if} \quad \tilde{m}(\overline{\tilde{e}}) = 1; \\ \left(V,\tilde{E},\tilde{m}^*_{\overline{\tilde{e}}}\right) & \text{if} \quad \tilde{m}(\overline{\tilde{e}}) > 1. \end{cases}$$

Here $\tilde{m}\big|_{\tilde{E}\setminus\{\overline{\tilde{e}}\}}$ stands for the restriction of $\tilde{m}$ to the domain $\tilde{E} \setminus \{\overline{\tilde{e}}\}$, while $\tilde{m}^*_{\overline{\tilde{e}}} : \tilde{E} \to \mathbf{N}^+$ is given by $\tilde{m}^*_{\overline{\tilde{e}}}(\tilde{e}) = \tilde{m}(\tilde{e}) - 1$ for $\tilde{e} = \overline{\tilde{e}}$ and $\tilde{m}^*_{\overline{\tilde{e}}}(\tilde{e}) = \tilde{m}(\tilde{e})$ for every $\tilde{e} \in \tilde{E} \setminus \{\overline{\tilde{e}}\}$. Notice that such definition is consistent with the interpretation of the parallel hyperedges as reflecting the independent entities and serves as the formal background needed for extending Theorem 1 to the wider classes of incidence structures.

Theorem 2. Let the removal of any single edge $\overline{e} \in E$ from an arbitrary non-empty multigraph $\hat{G} = (V,E,m) \in \Gamma_{\hat{G}}$ on n vertices lead to the increase in the topological distance between $f(\hat{G},\overline{e})$ pairs of its nodes. Then $f(\hat{G},\overline{e})$ is sharply bounded as follows:

$$0 \leq f(\hat{G},\overline{e}) \leq \begin{cases} n^2/4 & \text{if} \quad n \text{ is even;} \\ (n^2-1)/4 & \text{if} \quad n \text{ is odd.} \end{cases}$$

▲ The distance matrix of any $\Gamma_{\hat{G}}$ class multigraph $\hat{G}$ does not depend on its multiplicity function m due to the homogeneity of all parallel edges following from the absence of the weighted coefficients associated with them. Therefore, in contrast to $f(G,\overline{e})$, $f(\hat{G},\overline{e})$ has zero lower bound, which is achieved if and only if the single edge $\overline{e}$ removed from $\hat{G}$ has the multiplicity of $m(\overline{e}) > 1$. Revisiting the proof of Theorem 1, the upper bound on $f(G,\overline{e})$ originates from the fundamental limitation on the number of such vertex pairs $\{v,w\} \subseteq V$ that $\overline{e}$ occurs in the shortest path between v and w (i.e. satisfying the necessary condition). Since the allowance of the parallel edges does not affect such limitation, the upper bound on $f(G,\overline{e})$ holds also for $f(\hat{G},\overline{e})$. ▼

Theorem 3. Let the removal of any single hyperedge $\overline{\tilde{e}} \in \tilde{E}$ from an arbitrary non-empty simple hypergraph $H = (V,\tilde{E}) \in \Gamma_H$ on n nodes cause the increase in the topological distance between $f(H,\overline{\tilde{e}})$ pairs of its vertices. Then $f(H,\overline{\tilde{e}})$ is sharply bounded as follows:

$$0 \leq f(H,\overline{\tilde{e}}) \leq \frac{n(n-1)}{2}.$$

▲ Obviously, zero lower bound of $f(H,\overline{\tilde{e}})$ is tight if the removed hyperedge $\overline{\tilde{e}}$ constitutes a loop or if, for every pair of its vertices $\{v,w\} \subseteq \overline{\tilde{e}}$, the set of remaining hyperedges $\tilde{E} \setminus \{\overline{\tilde{e}}\}$ contains such $\tilde{e}$ that $\{v,w\} \subseteq \tilde{e}$. At the same time, in the complete simple hypergraph KH_n, each pair of nodes serves as the subset of $\sum_{k=2}^{n}\binom{n-2}{k-2}$ hyperedges. Accordingly, if H has at least $h(n) = 2^n + 1 - \sum_{k=2}^{n}\binom{n-2}{k-2}$ hyperedges, then the removal of any single one among them produces no influence on the entries of its distance matrix $\mathbf{D}(H)$. This observation, in turn, allows introducing the critical density $\lambda_c(n) = h(n)/(2^n - 1)$ such that $f(H,\overline{\tilde{e}}) = 0$ for every $\overline{\tilde{e}}$ in any H meeting the condition $\lambda(H) \geq \lambda_c(n)$ (i.e. having the density that equals to or exceeds the critical level at the corresponding

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

n). The in-depth analysis shows that $\lambda_c(n)$ monotonically decreases with the increase in n and asymptotically approaches 0.75 as n goes to infinity.

On the other hand, the upper bound on $f(H,\bar{e})$ equals the overall number of above-diagonal entries in the distance matrix of H (i.e. the number of possible ways to arrange the vertex set V into pairs). Furthermore, this bound is attained at only one configuration characterized by $\bar{e}$ having the maximum possible cardinality of n and such H that does not contain any another non-loop hyperedge (i.e. equipped with the set $\tilde{E}$ satisfying the condition $\tilde{E}\setminus\left(\{\bar{e}\}\bigcup\{\{v\}|v\in V\}\right)=\varnothing$). ▼

Remark that the sharp bounds on $f(H,\bar{e})$, in contrast to ones on $f(G,\bar{e})$ and $f(\hat{G},\bar{e})$, leave the largest possible variability gap of $n(n-1)/2$ vertex pairs and, in this sense, are not affected by the restriction on the form of the inter-vertex relationship underlying the formal definition of the Γ_H class incidence structures. Thereby, as a corollary of Theorem 3, in the general case of an arbitrary non-empty multihypergraph $\hat{H}=(V,\tilde{E},\tilde{m})\in\Gamma_{\hat{H}}$ with $|V|=n$, the number $f(\hat{H},\bar{e})$ of such pairs of nodes $\{v,w\}\subseteq V$ that the topological distance between v and w increases after the removal of any single hyperedge $\bar{e}\in\tilde{E}$ is also sharply bounded as $0\le f(\hat{H},\bar{e})\le n(n-1)/2$.

At the same time, the expression sharply bounding both $f(G,\bar{e})$ and $f(\hat{G},\bar{e})$ above approaches half the number of all possible combinations of n vertices into pairs as n tends to infinity and emanates from the limitation allowing any edge in the multigraphs to join only two nodes. In turn, the positive lower bound on $f(G,\bar{e})$ represents the distinctive property of the simple graphs that is lost at the transition to the wider classes of incidence structures $\Gamma_{\hat{G}}$, Γ_H, and $\Gamma_{\hat{H}}$. Conceptually, this property points to the unavoidable expansion and contraction of the metric space (V,d_G) induced by every $G\in\Gamma_G$ in result of, respectively, removing and adding any single edge. In view of these considerations, the closeness of vertices in the space $(V,d_{\hat{H}})$ associated with an arbitrary multihypergraph $\hat{H}$ is more strongly related with the cardinality of the edge set in its simple graph approximation $G[\hat{H}]$ than with the number of edges in $\hat{G}[\hat{H}]$ or with the number of hyperedges in $\hat{H}$ or $H[\hat{H}]$. This conclusion shows the reasonableness for introducing the metrics $\eta(\hat{H})=\sigma((\chi\circ\varphi)(\hat{H}))$ and $\zeta(\hat{H})=\rho((\chi\circ\varphi)(\hat{H}))$ describing the cohesiveness of $\hat{H}$ in terms of, respectively, the edge-to-vertex ratio and density of the corresponding approximation $G[\hat{H}]$. For example, every edge in the projection $G[\hat{H}]$ constructed for the multihypergraph $\hat{H}$ representing the co-authorship hypernetwork joins the pair of researchers that share the credit for at least one article as its joint authors and, thereby, are likely to be aware of each other. In this context, $\zeta(\hat{H})$ and $\eta(\hat{H})$ report on the number of such pairs normalized by their maximum possible number and the overall quantity of investigators within the considered group, respectively.

3. FUNDAMENTAL FORMS OF THE RICH-CLUB ORDERING IN MULTIHYPERGRAPHS AND PROPOSED METHOD FOR THEIR DETECTION AND ANALYSIS

Conceptually, the vertices contained in the multihypergraph $\hat{H}$ could be ranked based on specifying the richness (or importance) metric in the form of function $r_{\hat{H}}:V\to\mathbf{R}$. In particular, the degree $deg_{\hat{H}}:V\to\mathbf{N}$ plays a role of the canonical vertex richness metric, which is expressed as the mapping of every node $v\in V$ in $\hat{H}$ into the sum of the multiplicity scores associated with all hyperedges $\tilde{e}\in\tilde{E}$ including v, i.e. $deg_{\hat{H}}(v)=\sum_{v\in\tilde{e}\in\tilde{E}}\tilde{m}(\tilde{e})$. For example, the degree of any vertex in the multihypergraph model constructed for the co-authorship hypernetwork reflects the number of the separate scientific documents authored by the corresponding researcher. Let us formally define the ranking function $rank_{r_{\hat{H}}}: V \to \mathbf{N}^{+}$ induced by the metric $r_{\hat{H}}$ as matching each node $v \in V$ with its (possibly non-unique) ranking position defined as the increased by one number of vertices having the strictly larger richness than v according to $r_{\hat{H}}$, i.e. $rank_{r_{\hat{H}}}(v) = 1 + \left|\left\{v' \middle| (v' \in V) \wedge \left(r_{\hat{H}}(v') > r_{\hat{H}}(v)\right)\right\}\right|$. It is worth noting that such approach fully reproduces the strategy of the standard competition (or "1224") ranking widely adopted in the existing literature and extensively used in the practical applications. In order to more vividly illustrate the ideas underlying the ranking produced by $rank_{r_{\hat{H}}}$, let us associate the pair $(V, r_{\hat{H}})$ with the ordered sequence of the non-empty subsets $\left(V_1(r_{\hat{H}}), V_2(r_{\hat{H}}), ..., V_w(r_{\hat{H}})\right)$ meeting the requirements $V = V_1(r_{\hat{H}}) \bigcup V_2(r_{\hat{H}}) \bigcup ... \bigcup V_w(r_{\hat{H}})$, $r_{\hat{H}}(v) = r_{\hat{H}}(v')$ for every $v, v' \in V_i(r_{\hat{H}})$ at any $i \in \{1, 2, ..., w\}$, and $r_{\hat{H}}(v) > r_{\hat{H}}(v'')$ for each $v \in V_i(r_{\hat{H}})$ and $v'' \in V_j(r_{\hat{H}})$ such that $i < j$. In the context of such representation, all vertices assigned to the subset $V_i(r_{\hat{H}})$ occupy the same position $1 + \sum_{k=1}^{i-1} \left|V_k(r_{\hat{H}})\right|$ in the entire ranging outputted by $rank_{r_{\hat{H}}}$.

Due to the constraints on the structure of the ranking positions, the number of all possible non-identical standard competition rankings of $n \geq 2$ items (or, equivalently, the number of the distinct functions $rank_{r_{\hat{H}}}$ defined over the set V containing n vertices) is strictly lower than n^n, and, moreover, is determined by the following fundamental combinatorial law originally derived in this work:

$$\vartheta(n) = \sum_{u_1=0}^{\Upsilon(n,0)} \sum_{u_2=0}^{\Upsilon(n,1,u_1)} \sum_{u_3=0}^{\Upsilon(n,2,u_1,u_2)} \cdots \sum_{u_k=0}^{\Upsilon(n,k-1,u_1,u_2,...,u_{k-1})} \cdots \sum_{u_{n-1}=0}^{\Upsilon(n,n-2,u_1,u_2,...,u_{n-2})} X(n, u_1, u_2, ..., u_{n-1});$$

$$X(n, u_1, u_2, ..., u_{n-1}) = \left(\prod_{i=1}^{n-1} Y\left(i+1, u_i, n - \sum_{j=1}^{i-1}\left(u_j(j+1)\right)\right)\right)\left(n - \sum_{i=1}^{n-1}(u_i \cdot i)\right)!;$$

$$\Upsilon(n, h, u_1, u_2, ..., u_h) = \left\lfloor \frac{n - \sum_{i=1}^{h}\left(u_i(i+1)\right)}{h+1} \right\rfloor; \quad Y(k, u, p) = \begin{cases} \prod_{i=0}^{u-1} \binom{p - i \cdot k}{k} & \text{if} \quad u \geq 1 \\ 1 & \text{if} \quad u = 0 \end{cases}.$$

Notice that here $\lfloor \bullet \rfloor$ stands for the floor function. The allowance of the equally ranked items serves as the primary reason underlying the significant complication of the above analytical expression. For clarity, $X(n, u_1, u_2, ..., u_{n-1})$ gives the number of such non-identical rankings of n items that satisfy the parameterized limitation formulated as follows: for every $k \in \{1, 2, ..., n-1\}$, there exist exactly (i.e. not less and not more than) u_k ranking positions each of which is occupied exactly by $k+1$ items. In turn, $\vartheta(n)$ is calculated by summarizing the values of $X(n, u_1, u_2, ..., u_{n-1})$ for all possible limitations $u_1, u_2, ..., u_{n-1}$. Remark that such summarization is organized in the form of $n-1$ nested sums whose upper bounds ensure considering only such $u_1, u_2, ..., u_{n-1}$ that $X(n, u_1, u_2, ..., u_{n-1}) > 0$.

With a view to explaining the manner of deriving the formula for $X(n, u_1, u_2, ..., u_{n-1})$, let us, without the loss of generality, assume that the ranked items are explicitly labeled as $L = \{l_1, l_2, ..., l_n\}$. In addition, let us use the auxiliary notation $LR(L, u_1, u_2, ..., u_{n-1})$ for the set containing all such rankings $lr: L \to \mathbf{N}$ of the items in L that, for every $k \in \{1, 2, ..., n-1\}$, there are exactly u_k values (representing the ranking positions) each of which serves as the image of exactly $u_k + 1$ items under lr. In the context of such representation, the value of $X(n,u_1,u_2,...,u_{n-1})$ could be viewed as the cardinality of $LR(L,u_1,u_2,...,u_{n-1})$. Let us consider the set $LC(L,u_1,u_2,...,u_{n-1})$ containing all such combinations of $\sum_{i=1}^{n-1} u_i$ disjoint subsets of L that, for each $k \in \{1,2,...,n-1\}$, u_k of them have the cardinality of $k+1$. Note that there exists the surjective mapping from $LR(L,u_1,u_2,...,u_{n-1})$ to $LC(L,u_1,u_2,...,u_{n-1})$ associating every ranking $lr \in LR(L,u_1,u_2,...,u_{n-1})$ with one and only one combination in $LC(L,u_1,u_2,...,u_{n-1})$ whose every subset is composed of such items that have the same image under lr (i.e. are equally ranked). In turn, the left multiplier (represented by the product of $n-1$ terms) in the formula of $X(n,u_1,u_2,...,u_{n-1})$ gives the cardinality of $LC(L,u_1,u_2,...,u_{n-1})$. At the same time, any ranking $lr \in LR(L,u_1,u_2,...,u_{n-1})$ should have $\sum_{i=1}^{n-1} u_i$ non-unique (i.e. occupied by two or more items) ranking positions. Since all these positions are assigned to $\sum_{i=1}^{n-1}(u_i(i+1))$ items in total, only $n-\sum_{i=1}^{n-1}(u_i(i+1))$ remaining items hold the unique positions in ranking. Therefore, the overall number of the distinct ranking positions in any $lr \in LR(L,u_1,u_2,...,u_{n-1})$ (or the number of values in the image of L under lr) is given by $n-\sum_{i=1}^{n-1}(u_i \cdot i)$. In this light, there are $\left(n-\sum_{i=1}^{n-1}(u_i \cdot i)\right)!$ times more rankings in $LR(L,u_1,u_2,...,u_{n-1})$ than combinations in $LC(L,u_1,u_2,...,u_{n-1})$, which is reflected in the right multiplier in the formula of $X(n,u_1,u_2,...,u_{n-1})$. Notice that the calculation of $X(n,u_1,u_2,...,u_{n-1})$ relies on the auxiliary expression $Y(k,u,p)$ giving the number of the combinations of u disjoint subsets containing by k items selected from the set containing p items if $u \geq 1$ and one otherwise.

For clarity, let us illustrate the application of the proposed expression for calculating the number of all possible distinct standard competition rankings of four items:

$$\vartheta(4) = X(4,0,0,0)+X(4,0,0,1)+X(4,0,1,0)+X(4,1,0,0)+X(4,2,0,0)=81;$$
$$X(4,0,0,0)=Y(2,0,4)\times Y(3,0,4)\times Y(4,0,4)\times 4!=24;$$
$$X(4,0,0,1)=Y(2,0,4)\times Y(3,0,4)\times Y(4,1,4)\times(4-3)!=1;$$
$$X(4,0,1,0)=Y(2,0,4)\times Y(3,1,4)\times Y(4,0,1)\times(4-2)!=8;$$
$$X(4,1,0,0)=Y(2,1,4)\times Y(3,0,2)\times Y(4,0,2)\times(4-1)!=36;$$
$$X(4,2,0,0)=Y(2,2,4)\times Y(3,0,0)\times Y(4,0,0)\times(4-2)!=12;$$
$$Y(2,0,4)=Y(3,0,4)=Y(4,0,4)=Y(4,1,4)=1;$$
$$Y(3,1,4)=4; \quad Y(2,1,4)=Y(2,2,4)=6.$$

Conceptually, the analysis of the rich-club organization in $\hat{H}$ with respect to the richness metric $r_{\hat{H}}$ lies in determining whether the vertices ranked higher (i.e. having lower ranking positions) according to $rank_{r_{\hat{H}}}$ tend to be more tightly interconnected with each other. At the same time, the multihypergraphs, due to their complex combinatorial nature, allow introducing two following non-equivalent notions extending the concept of the vertex-induced subgraph defined for the simple graphs:

Definition 8. Let us consider an arbitrary multihypergraph $\hat{H}=(V,\tilde{E},\tilde{m}) \in \Gamma_{\hat{H}}$ and the subset of its vertices $A \subseteq V$. The *weak submultihypergraph of* $\hat{H}$ *induced by* A is given by $\hat{H}_A=\left(A,\tilde{E}_A,\tilde{m}\big|_{\tilde{P}_A}\right) \in \Gamma_{\hat{H}}$, where

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$\tilde{E}_A = \left\{ \left(\tilde{e} \bigcap A \right) \middle| \tilde{e} \in \tilde{P}_A \right\};$$

$$\tilde{P}_A = \left\{ \tilde{e} \middle| \left(\tilde{e} \in \tilde{E} \right) \wedge \left(\tilde{e} \bigcap A \neq \varnothing \right) \right\}.$$

Conversely, the *strong submultihypergraph of* $\hat{H}$ *induced by* A is represented in the form of $\hat{H}^*_A = \left(A, \tilde{E}^*_A, \tilde{m}\big|_{\tilde{E}^*_A} \right) \in \Gamma_{\hat{H}}$, where

$$\tilde{E}^*_A = \left\{ \tilde{e} \middle| \left(\tilde{e} \in \tilde{E} \right) \wedge \left(\tilde{e} \subseteq A \right) \right\}.$$

Note that $\tilde{m}\big|_{\tilde{P}_A}$ and $\tilde{m}\big|_{\tilde{E}^*_A}$ stand in the above expressions for the restrictions of the multiplicity function $\tilde{m}$ to the domains $\tilde{P}_A \subseteq \tilde{E}$ and $\tilde{E}^*_A \subseteq \tilde{P}_A$, respectively.

Remark that the substructures $\hat{H}_A$ and $\hat{H}^*_A$ provide the conceptually different and complementary viewpoints on the relationships established among the nodes of A via the hyperedges of the overall multihypergraph $\hat{H}$. In particular, the strong submultihypergraph $\hat{H}^*_A$, in contrast to the weak one $\hat{H}_A$, is formed without the modification of the individual hyperedges contained in $\tilde{E}$ and encapsulates only such relationships among the nodes of A in $\hat{H}$ that do not involve the participation of the vertices belonging to $V \setminus A$. The natural consequence following from these considerations lies in the need for distinguishing the weak and strong forms of the rich-club organization in multihypergraphs depending on the type of submultihypergraphs used for representing the relationships among the richest nodes.

For convenience of the subsequent discussion, let us introduce the auxiliary notation $V\left(r_{\hat{H}}, t\right) = \left\{ v \middle| \left(v \in V \right) \wedge \left(r_{\hat{H}}\left(v \right) \geq t \right) \right\} \subseteq V$ for the subset comprising such vertices of $\hat{H}$ whose richness score according to the generic metric $r_{\hat{H}}$ equals to or exceeds the threshold value t. Additionally, let us denote by $\hat{H}\left(r_{\hat{H}}, t\right) = \hat{H}_{V\left(r_{\hat{H}}, t\right)}$ and $\hat{H}^*\left(r_{\hat{H}}, t\right) = \hat{H}^*_{V\left(r_{\hat{H}}, t\right)}$, respectively, the weak and strong submultihypergraphs of $\hat{H}$ induced by $V\left(r_{\hat{H}}, t\right)$. In terms of such notation, the problem of discovering and analyzing the weak and strong forms of the rich-club organization in $\hat{H}$ (with respect to the prescribed richness metric $r_{\hat{H}}$) is reduced to examining the corresponding families of the weak and strong submultihypergraphs $F\left(\hat{H}, r_{\hat{H}}, T\right) = \left\{ \hat{H}\left(r_{\hat{H}}, t_1\right), \ldots, \hat{H}\left(r_{\hat{H}}, t_\kappa\right) \right\}$ and $F^*\left(\hat{H}, r_{\hat{H}}, T\right) = \left\{ \hat{H}^*\left(r_{\hat{H}}, t_1\right), \ldots, \hat{H}^*\left(r_{\hat{H}}, t_\kappa\right) \right\}$. Here $T = \left\{ t_1, \ldots, t_\kappa \right\}$ represents the indexing set containing the richness thresholds, while its cardinality κ determines the granularity of the analysis.

At the same time, due to the normalization by the number of vertices or within the range $\left[0,1\right]$, all metrics $\tilde{\sigma}$, θ, $\tilde{\tau}$, η, $\tilde{\mu}$, and ζ introduced in the previous section are suitable for tracing and comparing the cohesiveness of the submultihypergraphs comprising the families $F\left(\hat{H}, r_{\hat{H}}, T^r_{\hat{H}}\right)$ and $F^*\left(\hat{H}, r_{\hat{H}}, T^r_{\hat{H}}\right)$. Moreover, since these metrics report on the tightness of inter-vertex relationships from the conceptually different viewpoints, the multihypergraphs, as opposed to the simple graphs, allow introducing multiple natural notions of the rich-club coefficient for both weak and strong forms of the rich-club organization. These considerations, in turn, lead to the formulation of the following formal definitions:

Definition 9. The *ordinary* α*-,* β*-,* γ*-,* δ*-,* $\tilde{\gamma}$*-, and* $\tilde{\delta}$*-coefficients characterizing the weak form of the rich-club ordering in any multihypergraph* $\hat{H}$ are expressed respectively as

$$\alpha_w\left(\hat{H}, r_{\hat{H}}, t\right) = \tilde{\sigma}\left(\hat{H}\left(r_{\hat{H}}, t\right)\right);$$

$$\beta_w\left(\hat{H}, r_{\hat{H}}, t\right) = \theta\left(\hat{H}\left(r_{\hat{H}}, t\right)\right);$$

$$\gamma_w\left(\hat{H}, r_{\hat{H}}, t\right) = \tilde{\tau}\left(\hat{H}\left(r_{\hat{H}}, t\right)\right);$$

$$\delta_w\left(\hat{H}, r_{\hat{H}}, t\right) = \eta\left(\hat{H}\left(r_{\hat{H}}, t\right)\right);$$

$$\tilde{\gamma}_w\left(\hat{H}, r_{\hat{H}}, t\right) = \tilde{\mu}\left(\hat{H}\left(r_{\hat{H}}, t\right)\right);$$

$$\tilde{\delta}_w\left(\hat{H}, r_{\hat{H}}, t\right) = \zeta\left(\hat{H}\left(r_{\hat{H}}, t\right)\right).$$

Definition 10. The *ordinary* α*-,* β*-,* γ*-,* δ*-,* $\tilde{\gamma}$*-, and* $\tilde{\delta}$*-coefficients characterizing the strong*

Journal of Theoretical and Applied Information Technology
15th December 2018. Vol.96. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

form of the rich-club ordering in any multihypergraph $\hat{H}$ are defined respectively as

$$\alpha_s\left(\hat{H}, r_{\hat{H}}, t\right) = \tilde{\sigma}\left(\hat{H}^*\left(r_{\hat{H}}, t\right)\right);$$

$$\beta_s\left(\hat{H}, r_{\hat{H}}, t\right) = \theta\left(\hat{H}^*\left(r_{\hat{H}}, t\right)\right);$$

$$\gamma_s\left(\hat{H}, r_{\hat{H}}, t\right) = \tilde{\tau}\left(\hat{H}^*\left(r_{\hat{H}}, t\right)\right);$$

$$\delta_s\left(\hat{H}, r_{\hat{H}}, t\right) = \eta\left(\hat{H}^*\left(r_{\hat{H}}, t\right)\right);$$

$$\tilde{\gamma}_s\left(\hat{H}, r_{\hat{H}}, t\right) = \tilde{\mu}\left(\hat{H}^*\left(r_{\hat{H}}, t\right)\right);$$

$$\tilde{\delta}_s\left(\hat{H}, r_{\hat{H}}, t\right) = \zeta\left(\hat{H}^*\left(r_{\hat{H}}, t\right)\right).$$

Since every hyperedge of the strong submultihypergraph $\hat{H}^*\left(r_{\hat{H}}, t\right)$ also exists in the weak one $\hat{H}\left(r_{\hat{H}}, t\right)$ (while the reverse is not always true), the introduced coefficients should meet the following system of constraints:

$$\alpha_w\left(\hat{H}, r_{\hat{H}}, t\right) \geq \alpha_s\left(\hat{H}, r_{\hat{H}}, t\right);$$

$$\beta_w\left(\hat{H}, r_{\hat{H}}, t\right) \geq \beta_s\left(\hat{H}, r_{\hat{H}}, t\right);$$

$$\gamma_w\left(\hat{H}, r_{\hat{H}}, t\right) \geq \gamma_s\left(\hat{H}, r_{\hat{H}}, t\right);$$

$$\delta_w\left(\hat{H}, r_{\hat{H}}, t\right) \geq \delta_s\left(\hat{H}, r_{\hat{H}}, t\right);$$

$$\tilde{\gamma}_w\left(\hat{H}, r_{\hat{H}}, t\right) \geq \tilde{\gamma}_s\left(\hat{H}, r_{\hat{H}}, t\right);$$

$$\tilde{\delta}_w\left(\hat{H}, r_{\hat{H}}, t\right) \geq \tilde{\delta}_s\left(\hat{H}, r_{\hat{H}}, t\right).$$

Moreover, in view of the relationships $\tilde{\tau}\left(\hat{H}\right) \leq \tilde{\sigma}\left(\hat{H}\right) \leq \theta\left(\hat{H}\right)$ and $\eta\left(\hat{H}\right) \leq \theta\left(\hat{H}\right)$ established between the multihypergraph cohesiveness metrics underlying the formal definition of the proposed coefficients, we could deduce the additional limitation formulated as follows:

$$\gamma_w\left(\hat{H}, r_{\hat{H}}, t\right) \leq \alpha_w\left(\hat{H}, r_{\hat{H}}, t\right) \leq \beta_w\left(\hat{H}, r_{\hat{H}}, t\right);$$

$$\delta_w\left(\hat{H}, r_{\hat{H}}, t\right) \leq \beta_w\left(\hat{H}, r_{\hat{H}}, t\right);$$

$$\gamma_s\left(\hat{H}, r_{\hat{H}}, t\right) \leq \alpha_s\left(\hat{H}, r_{\hat{H}}, t\right) \leq \beta_s\left(\hat{H}, r_{\hat{H}}, t\right);$$

$$\delta_s\left(\hat{H}, r_{\hat{H}}, t\right) \leq \beta_s\left(\hat{H}, r_{\hat{H}}, t\right).$$

At the specified metric $r_{\hat{H}}$, any individual ordinary rich-club coefficient is considered as attesting to the presence of the ordinary rich-club phenomenon in $\hat{H}$ if it demonstrates the growing tendency with the increase in t and reaches the peak value at the threshold t exceeding the richness score for the vast majority of vertices in $\hat{H}$. Moreover, if all six coefficients given in Definition 9 (Definition 10) are characterized by such behavior, $\hat{H}$ is referred to as exhibiting the *complete weak (strong) ordinary rich-club phenomenon* with respect to $r_{\hat{H}}$. On the other hand, if only from one to four of the coefficients formulated in Definition 9 (Definition 10) excepting $\tilde{\gamma}_w\left(\hat{H}, r_{\hat{H}}, t\right)$ ($\tilde{\gamma}_s\left(\hat{H}, r_{\hat{H}}, t\right)$) demonstrate the above-described behavior, $\hat{H}$ is referred to as expressing the *partial weak (strong) ordinary rich-club phenomenon* with respect to $r_{\hat{H}}$.

Remark that due to the presence of the exponential functions in their denominators, the coefficients $\tilde{\gamma}_w\left(\hat{H}, r_{\hat{H}}, t\right)$ and $\tilde{\gamma}_s\left(\hat{H}, r_{\hat{H}}, t\right)$ demonstrate the catastrophic growth with the increase in t at the sparsity (i.e. close to zero loopless density) of the simple hypergraph approximation $H\left[\hat{H}\right]$ constructed for $\hat{H}$. Thereby, these coefficients are ignored in the formal definition of the partial ordinary rich-club phenomenon given above.

Considering such two thresholds t' and t'' such that $t' < t''$, $\left|V\left(r_{\hat{H}}, t''\right)\right| < \left|V\left(r_{\hat{H}}, t'\right)\right|$, and $\left|V\left(r_{\hat{H}}, t''\right)\right| \geq 2$, we could impose the following restrictions on the behavior of the γ-, δ-, $\tilde{\gamma}$-, and $\tilde{\delta}$-coefficients characterizing the weak and strong forms of the rich-club ordering in $\hat{H}$:

1. $\tilde{\delta}_w\left(\hat{H}, r_{\hat{H}}, t''\right) \leq \tilde{\delta}_w\left(\hat{H}, r_{\hat{H}}, t'\right) \Rightarrow \delta_w\left(\hat{H}, r_{\hat{H}}, t''\right) < \delta_w\left(\hat{H}, r_{\hat{H}}, t'\right);$
2. $\tilde{\delta}_s\left(\hat{H}, r_{\hat{H}}, t''\right) \leq \tilde{\delta}_s\left(\hat{H}, r_{\hat{H}}, t'\right) \Rightarrow \delta_s\left(\hat{H}, r_{\hat{H}}, t''\right) < \delta_s\left(\hat{H}, r_{\hat{H}}, t'\right);$
3. $\tilde{\gamma}_w\left(\hat{H}, r_{\hat{H}}, t''\right) \leq \tilde{\gamma}_w\left(\hat{H}, r_{\hat{H}}, t'\right) \Rightarrow \gamma_w\left(\hat{H}, r_{\hat{H}}, t''\right) < \gamma_w\left(\hat{H}, r_{\hat{H}}, t'\right);$

4. $\tilde{\gamma}_s\left(\hat{H},r_{\hat{H}},t''\right)\leq\tilde{\gamma}_s\left(\hat{H},r_{\hat{H}},t'\right)\Rightarrow\gamma_s\left(\hat{H},r_{\hat{H}},t''\right)<\gamma_s\left(\hat{H},r_{\hat{H}},t'\right)$.

Remark that here the symbol $\Rightarrow$ should be read as "implies that". For clarity, let us provide the proof for the first statement by using the auxiliary notations x' and x'' for $\left|V\left(r_{\hat{H}},t'\right)\right|$ and $\left|V\left(r_{\hat{H}},t''\right)\right|$, respectively. In addition, let us denote the number of edges contained in the simple graph approximations constructed for the weak submultihypergraphs $\hat{H}\left(r_{\hat{H}},t'\right)$ and $\hat{H}\left(r_{\hat{H}},t''\right)$ respectively by y' and y''.

This background allows conveniently expressing $\tilde{\delta}_w\left(\hat{H},r_{\hat{H}},t''\right)\leq\tilde{\delta}_w\left(\hat{H},r_{\hat{H}},t'\right)$ in the form $2y''/\left(x''\left(x''-1\right)\right)\leq 2y'/\left(x'\left(x'-1\right)\right)$. By multiplying both sides of this inequality by $\left(x''\left(x''-1\right)\right)/2$ (which is positive over the whole considered domain bounded by $2\leq x''<x'$), we could get $y''\leq y'\left(x''\left(x''-1\right)\right)/\left(x'\left(x'-1\right)\right)$.

Since $\left(x''\left(x''-1\right)\right)/\left(x'\left(x'-1\right)\right)<x''/x'$ for $2\leq x''<x'$, we could deduce that $y''<y'\cdot x''/x'$. Moreover, by dividing both sides of the obtained inequality by any positive x'', we could rewrite it as $y''/x''<y'/x'$, which reproduces the expression $\delta_w\left(\hat{H},r_{\hat{H}},t''\right)<\delta_w\left(\hat{H},r_{\hat{H}},t'\right)$ in terms of the adopted notation and eventually completes the proof. Notice that the second statement could be proved by analogy.

When dealing with the third statement, let us introduce the additional notations z' and z'' standing for the number of the non-loop hyperedges in the simple hypergraph approximations of $\hat{H}\left(r_{\hat{H}},t'\right)$ and $\hat{H}\left(r_{\hat{H}},t''\right)$, respectively. In this light, $\tilde{\gamma}_w\left(\hat{H},r_{\hat{H}},t''\right)\leq\tilde{\gamma}_w\left(\hat{H},r_{\hat{H}},t'\right)$ is given by the inequality $z''/\left(2^{x''}-x''-1\right)\leq z'/\left(2^{x'}-x'-1\right)$. In turn, the multiplication of both its sides by any positive $2^{x''}-x''-1$ allows obtaining $z''\leq z'\left(2^{x''}-x''-1\right)/\left(2^{x'}-x'-1\right)$.

Since $\left(2^{x''}-x''-1\right)/\left(2^{x'}-x'-1\right)<x''/x'$ for $2\leq x''<x'$, we could conclude that, over the considered domain, $z''<z'\cdot x''/x'$ and $z''/x''<z'/x'$. The last inequality reproduces the expression $\gamma_w\left(\hat{H},r_{\hat{H}},t''\right)<\gamma_w\left(\hat{H},r_{\hat{H}},t'\right)$, which clearly proves the validity of the implication. The proof of the fourth statement could be given based on the same principles. Conceptually, all formulated statements allow viewing the ordinary γ- and δ-coefficients as embodying the more strict (i.e. fulfilled at the fewer topological configurations) criteria compared to the corresponding ordinary $\tilde{\gamma}$- and $\tilde{\delta}$-coefficients when assessing the presence of the ordinary rich-club phenomenon in $\hat{H}$.

In sum, under the proposed approach, the $\Gamma_{\hat{H}}$ class multihypergraphs are assessed by twelve non-equivalent ordinary rich-club coefficients. At the same time, with the restriction of the consideration area to the subclasses Γ_H, $\Gamma_{\hat{G}}$, and Γ_G, some of these coefficients collapse into the identical notions. In particular, for any simple hypergraph H belonging to Γ_H, we would have $\tilde{\sigma}(H)=\tilde{\tau}(H)$, which implies that

$$\alpha_w(H,r_H,t)=\gamma_w(H,r_H,t);$$
$$\alpha_s(H,r_H,t)=\gamma_s(H,r_H,t).$$

Therefore, the simple hypergraphs are described by only ten non-equivalent ordinary rich-club coefficients. For its part, the case of an arbitrary multigraph $\hat{G}\in\Gamma_{\hat{G}}$ is characterized by the equalities $\tilde{\sigma}\left(\hat{G}\right)=\theta\left(\hat{G}\right)$, $\tilde{\tau}\left(\hat{G}\right)=\eta\left(\hat{G}\right)$, and $\tilde{\mu}\left(\hat{G}\right)=\zeta\left(\hat{G}\right)$. Additionally, let us consider the pair of the weak $\hat{G}_A$ and strong $\hat{G}_A^*$ submultihypergraphs of $\hat{G}$ induced by any subset of its vertices A. Obviously, $\hat{G}_A\in\Gamma_{\hat{H}}$ and $\hat{G}_A^*\in\Gamma_{\hat{G}}$, while $\hat{G}_A$ could differ from $\hat{G}_A^*$ only by the presence of the loop hyperedges, which are ignored by all metrics $\tilde{\sigma}$, θ, $\tilde{\tau}$, η, $\tilde{\mu}$, and ζ. Thereby, in the particular case of the $\Gamma_{\hat{G}}$ class multigraphs, the weak and strong forms of the rich-club organization could be viewed as collapsing into the single one, which allows writing the next equalities:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$\alpha_w\left(\hat{G},r_{\hat{G}},t\right)=\beta_w\left(\hat{G},r_{\hat{G}},t\right)=$$
$$=\alpha_s\left(\hat{G},r_{\hat{G}},t\right)=\beta_s\left(\hat{G},r_{\hat{G}},t\right);$$
$$\gamma_w\left(\hat{G},r_{\hat{G}},t\right)=\delta_w\left(\hat{G},r_{\hat{G}},t\right)=$$
$$=\gamma_s\left(\hat{G},r_{\hat{G}},t\right)=\delta_s\left(\hat{G},r_{\hat{G}},t\right);$$
$$\tilde{\gamma}_w\left(\hat{G},r_{\hat{G}},t\right)=\tilde{\delta}_w\left(\hat{G},r_{\hat{G}},t\right)=$$
$$=\tilde{\gamma}_s\left(\hat{G},r_{\hat{G}},t\right)=\tilde{\delta}_s\left(\hat{G},r_{\hat{G}},t\right).$$

In this light, the multigraphs are characterized by only three non-equivalent ordinary rich-club coefficients. Furthermore, the restrictions imposed on the structure of any simple graph $G\in\Gamma_G$ produce the equalities $\tilde{\sigma}(G)=\theta(G)=\tilde{\tau}(G)=\eta(G)$ and $\tilde{\mu}(G)=\zeta(G)$ underlying the most significant collapse of the ordinary rich-club coefficients given as follows:

$$\alpha_w\left(G,r_G,t\right)=\beta_w\left(G,r_G,t\right)=\gamma_w\left(G,r_G,t\right)=$$
$$\delta_w\left(G,r_G,t\right)=\alpha_s\left(G,r_G,t\right)=\beta_s\left(G,r_G,t\right)=$$
$$=\gamma_s\left(G,r_G,t\right)=\delta_s\left(G,r_G,t\right);$$
$$\tilde{\gamma}_w\left(G,r_G,t\right)=\tilde{\delta}_w\left(G,r_G,t\right)=$$
$$=\tilde{\gamma}_s\left(G,r_G,t\right)=\tilde{\delta}_s\left(G,r_G,t\right).$$

Therefore, any simple graph G is described by only two non-equivalent ordinary rich-club coefficients reduced to the edge-to-vertex ratio $\sigma\left(G^*\left(r_G,t\right)\right)$ and density $\rho\left(G^*\left(r_G,t\right)\right)$ of its strong submultihypergraph $G^*\left(r_G,t\right)\in\Gamma_G$ induced by the subset of vertices having the richness score of t or more according to the specified metric r_G.

Meanwhile, the origin of the ordinary rich-club phenomenon attested by the ordinary coefficients given in Definitions 9 and 10 could lie not entirely in the self-organizing behavior of the hypernetwork modeled by $\hat{H}$ but also in the local properties of its constituent actors and entities reflected in the degrees and cardinalities of the corresponding nodes and hyperedges. In turn, the analysis of the rich-club organization at the discounted influence of such effects requires formulating the appropriately corrected coefficients. Above all, by analogy to the hyperedge cardinality sequence, let us introduce the following integrative characteristic summarizing the degrees of all multihypergraph's vertices:

Definition 11. The *vertex degree sequence* of an arbitrary non-null multihypergraph $\hat{H}=\left(V,\tilde{E},\tilde{m}\right)$ is represented in the form $\Xi\left(\hat{H}\right)=\left(\xi_1\left(\hat{H}\right),\xi_2\left(\hat{H}\right),\ldots,\xi_{\upsilon\left(\hat{H}\right)+1}\left(\hat{H}\right)\right)$. Here $\upsilon\left(\hat{H}\right)=\left|\left(\tilde{E},\tilde{m}\right)\right|$ stands for the largest possible degree of vertex in the multihypergraph with $\left|\left(\tilde{E},\tilde{m}\right)\right|$ hyperedges, while the size of $\Xi\left(\hat{H}\right)$ is determined as $\upsilon\left(\hat{H}\right)+1$. In turn, every item $\xi_i\left(\hat{H}\right)$ for $i\in\left\{1,2,\ldots,\upsilon\left(\hat{H}\right)+1\right\}$ represents the number of nodes with the degree of $i-1$ in $\hat{H}$, i.e.

$$\xi_i\left(\hat{H}\right)=\left|\left\{v\middle|\left(v\in V\right)\wedge\left(deg_{\hat{H}}\left(v\right)=i-1\right)\right\}\right|.$$

Based on this background, let us associate the examined multihypergraph $\hat{H}$ with the reference ensemble $\varepsilon\left(\hat{H}\right)$ represented by the set composed of the multihypergraphs following strictly the same sequences of vertex degrees $\Xi\left(\hat{H}\right)$ and hyperedge cardinalities $\Omega\left(\hat{H}\right)$ as $\hat{H}$. At the same time, the construction of such ensembles needs designing the algorithm of generator producing the random multihypergraph with the prescribed sequences of vertex degrees and hyperedge cardinalities, while ensuring the proper level of stochasticity (i.e. uniformity in the probabilities of forming all allowed configurations). Remark that, in the prior findings, the problem of generating the random simple graphs and multigraphs having the given sequence of vertex degrees was extensively researched due their popularity as the reference models in analyzing the topological organization of the complex networks [13]. In particular, the article [14] deals with the overview of main ideas underlying the existing generators of such structures. However, the above-stated problem of generating the random multihypergraph with the fixed sequences of vertex degrees and hyperedge cardinalities still constitutes the gap in knowledge and is the much more challenging since the produced structures should exactly fulfill the requirements to not one but two sequences.

Conceptually, such generator could be elaborated in accordance to either one of two fundamental paradigms. Looking into detail, the first paradigm consists in performing the series of such elementary random modifications of the initial multihypergraph $\hat{H}$ that, while keeping its sequences of vertex degrees and hyperedge cardinalities unaffected, allow obtaining the new multihypergraph equipped with the randomized multiset of hyperedges. The most naive example of such modification lies in choosing two hyperedges $\tilde{e}$ and $\tilde{e}'$ in $\hat{H}$ such that $\tilde{e} \setminus \tilde{e}' \neq \varnothing$ and $\tilde{e}' \setminus \tilde{e} \neq \varnothing$ with the subsequent swapping the pair of vertices $v \in \tilde{e} \setminus \tilde{e}'$ and $v' \in \tilde{e}' \setminus \tilde{e}$ such that $v \notin \tilde{e}' \setminus \tilde{e}$ and $v' \notin \tilde{e} \setminus \tilde{e}'$ between the single copies of these hyperedges (i.e. without influencing the parallel hyperedges). In simple terms, v' is placed instead of v in the single instance of $\tilde{e}$, while v occupies the place of v' in the single instance of $\tilde{e}'$. Conversely, the second paradigm implies the direct construction of the random multihypergraph possessing the desired characteristics from the set of isolated nodes by the successive addition of hyperedges. Notice that the second paradigm, in contrast to the first one, precludes the dependence of the output produced by the generator on the particular hypergraph used as initial and, thereby, could potentially lead to the formation of the more qualitative reference ensembles $\varepsilon(\hat{H})$. Accordingly, the algorithm for generating the random multihypergraph with the prescribed sequences of vertex degrees and hyperedge cardinalities designed in this work implements the second paradigm.

As input, the proposed algorithm takes two sequences $\Xi(\hat{H})$ and $\Omega(\hat{H})$. For brevity, let us denote the number of all their items respectively as $k+1$ and n. In turn, the description of the algorithm could be arranged in the following way:

Step 1: Initialize two empty sequences $VS = ()$ and $ES = ()$. Traverse all items of $\Xi(\hat{H})$ in the decreasing order of their index (i.e. starting from the tail of $\Xi(\hat{H})$) and, for every considered $\xi_i(\hat{H}) > 0$, append $\xi_i(\hat{H})$ elements with the value of $i-1$ to the tail of VS. Similarly, traverse all items of $\Omega(\hat{H})$ in the order of decreasing their index and, for each considered $\omega_i(\hat{H}) > 0$, update ES by appending $\omega_i(\hat{H})$ elements with the value of i to its tail. Remark that, in result of such operations, both VS and ES should represent the non-increasing sequences. Additionally, initialize the zero-filled sequence VS' whose size (i.e. number of items) is equal to n. Apart from that, construct the sets $TS = \{1, 2, ..., k\}$ and $FS = \varnothing$. Finally, initialize the sequence $C = (C_1, C_2, ..., C_k)$ containing k empty sets, i.e. $C_i = \varnothing$ for $i \in \{1, 2, ..., k\}$ and set $p = 1$.

Step 2: If $p \leq n$, go to the step 3. Otherwise, the construction of the desired random multihypergraph is complete.

Step 3: If $|TS| \geq vs_p - vs'_p$, proceed to the step 4. Otherwise, go to the step 8.

Step 4: If $vs'_p < vs_p$, go to the step 5. Otherwise, set $p = p+1$ and return to the step 2.

Step 5: Generate the random natural number er lying within the range from 1 to k.

Step 6: If $p \notin C_{er}$ and $er \in TS$, proceed to the step 7. Otherwise, return to the step 5 and generate another er.

Step 7: Perform the operation $C_{er} = C_{er} \bigcup \{p\}$ and set $vs'_p = vs'_p + 1$. If $|C_{er}| = es_{er}$, $TS = TS \setminus \{er\}$, and $FS = FS \bigcup \{er\}$. Return to the step 4.

Step 8: Randomly select the pair of numbers $u \in TS$ and $w \in FS$. Proceed to the step 9.

Step 9: If $C_w \setminus C_u = \varnothing$, return to the step 8 and generate another u and w. Otherwise, proceed to the step 10.

Step 10: Pick randomly $q \in C_w \setminus C_u$ and update the sets C_u and C_w by performing the operations $C_u = C_u \bigcup \{q\}$ and $C_w = (C_w \setminus \{q\}) \bigcup \{p\}$. Set $vs'_p = vs'_p + 1$. If $|C_u| = es_u$, $TS = TS \setminus \{u\}$ and $FS = FS \bigcup \{u\}$. Go to the step 3.

Notice that the random multihypergraph outputted by this generator has n nodes labeled by the natural numbers within the range from 1 to n. In turn, its hyperedges (including possibly the parallel ones) are represented by the sets comprising the sequence C.

Let us provide the brief commentary of the main ideas underlying the proposed algorithm. The generating process is driven by the non-increasing sequence VS whose each item vs_i gives the degree of the vertex i in the multihypergraph to be constructed (i.e. the nodes are labeled in the non-increasing order of their degrees). For p (meaning the currently considered vertex) going from 1 to k (i.e. the nodes with larger degrees are considered earlier), the generator adds p to vs_p distinct sets in C (which means the inclusion of the vertex in the hyperedges). At the same time, the cardinality of every set C_i for $i \in \{1,2,...,k\}$ in C could not exceed the value of es_i, which allows reproducing the specified sequence of the hyperedge cardinalities. For convenience, the set FS contains the indexes i of all such sets C_i in C that are fully filled to their cardinality limit es_i. Conversely, TS includes the indexes i of such sets C_i in C that have the vacant places for accommodating the new numbers. Remark that the steps 3 and 8 – 10 are required for escaping the generating process from the deadlock situation in which the number of the remaining non-filled sets in C (i.e. the cardinality of TS) is lower than the number of sets to which the currently considered vertex p should be added in order to meet the degree requirement.

Let us consider the metric $r_{\hat{H}}$ explicitly defined by the underlying strategy (i.e. the fundamental system of rules) of mapping every vertex in $\hat{H}$ into its richness score based on the topology of $\hat{H}$ (such as the degree $deg_{\hat{H}}$). In this case, for every multihypergraph $\hat{R}$ belonging to the ensemble $\varepsilon(\hat{H})$, there exists such metric $r_{\hat{R}}$ that realizes exactly the same topology-based strategy in assigning the richness scores to its vertices, which allows introducing the following families of the normalized rich-club coefficients:

Definition 12. The *normalized* α-, β-, γ-, *and* δ-*coefficients characterizing the weak form of the rich-club ordering in any multihypergraph* $\hat{H}$ are calculated respectively as

$$\alpha_w^n\left(\hat{H}, r_{\hat{H}}, t\right) = \frac{\alpha_w\left(\hat{H}, r_{\hat{H}}, t\right)}{\frac{1}{\left|\varepsilon\left(\hat{H}\right)\right|}\sum_{\hat{R}\in\varepsilon\left(\hat{H}\right)} \alpha_w\left(\hat{R}, r_{\hat{R}}, t\right)};$$

$$\beta_w^n\left(\hat{H}, r_{\hat{H}}, t\right) = \frac{\beta_w\left(\hat{H}, r_{\hat{H}}, t\right)}{\frac{1}{\left|\varepsilon\left(\hat{H}\right)\right|}\sum_{\hat{R}\in\varepsilon\left(\hat{H}\right)} \beta_w\left(\hat{R}, r_{\hat{R}}, t\right)};$$

$$\gamma_w^n\left(\hat{H}, r_{\hat{H}}, t\right) = \frac{\gamma_w\left(\hat{H}, r_{\hat{H}}, t\right)}{\frac{1}{\left|\varepsilon\left(\hat{H}\right)\right|}\sum_{\hat{R}\in\varepsilon\left(\hat{H}\right)} \gamma_w\left(\hat{R}, r_{\hat{R}}, t\right)};$$

$$\delta_w^n\left(\hat{H}, r_{\hat{H}}, t\right) = \frac{\delta_w\left(\hat{H}, r_{\hat{H}}, t\right)}{\frac{1}{\left|\varepsilon\left(\hat{H}\right)\right|}\sum_{\hat{R}\in\varepsilon\left(\hat{H}\right)} \delta_w\left(\hat{R}, r_{\hat{R}}, t\right)}.$$

Definition 13. The *normalized* α-, β-, γ-, *and* δ-*coefficients characterizing the strong form of the rich-club ordering in any multihypergraph* $\hat{H}$ are calculated respectively as

$$\alpha_s^n\left(\hat{H}, r_{\hat{H}}, t\right) = \frac{\alpha_s\left(\hat{H}, r_{\hat{H}}, t\right)}{\frac{1}{\left|\varepsilon\left(\hat{H}\right)\right|}\sum_{\hat{R}\in\varepsilon\left(\hat{H}\right)} \alpha_s\left(\hat{R}, r_{\hat{R}}, t\right)};$$

$$\beta_s^n\left(\hat{H}, r_{\hat{H}}, t\right) = \frac{\beta_s\left(\hat{H}, r_{\hat{H}}, t\right)}{\frac{1}{\left|\varepsilon\left(\hat{H}\right)\right|}\sum_{\hat{R}\in\varepsilon\left(\hat{H}\right)} \beta_s\left(\hat{R}, r_{\hat{R}}, t\right)};$$

$$\gamma_s^n\left(\hat{H}, r_{\hat{H}}, t\right) = \frac{\gamma_s\left(\hat{H}, r_{\hat{H}}, t\right)}{\frac{1}{\left|\varepsilon\left(\hat{H}\right)\right|}\sum_{\hat{R}\in\varepsilon\left(\hat{H}\right)} \gamma_s\left(\hat{R}, r_{\hat{R}}, t\right)};$$

$$\delta_s^n\left(\hat{H}, r_{\hat{H}}, t\right) = \frac{\delta_s\left(\hat{H}, r_{\hat{H}}, t\right)}{\frac{1}{\left|\varepsilon\left(\hat{H}\right)\right|}\sum_{\hat{R}\in\varepsilon\left(\hat{H}\right)} \delta_s\left(\hat{R}, r_{\hat{R}}, t\right)}.$$

Remark that the numerator of the expression for every normalized rich-club coefficient is represented by the corresponding ordinary coefficient, while its denominator gives the average

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

value of such ordinary coefficient over the multihypergraphs $\hat{R}$ of the ensemble $\varepsilon\left(\hat{H}\right)$ at the respective richness metrics $r_{\hat{R}}$. Conceptually, any individual normalized rich-club coefficient could be interpreted as attesting to the presence of the normalized rich-club phenomenon in $\hat{H}$ at the specified richness metric $r_{\hat{H}}$ and the threshold t if its value at all these parameters exceeds one.

In turn, if all four (only from one to three) conditions $\alpha_w^n\left(\hat{H},r_{\hat{H}},t\right)>1$, $\beta_w^n\left(\hat{H},r_{\hat{H}},t\right)>1$, $\gamma_w^n\left(\hat{H},r_{\hat{H}},t\right)>1$, and $\delta_w^n\left(\hat{H},r_{\hat{H}},t\right)>1$ are satisfied, $\hat{H}$ is referred to as demonstrating the *complete (partial) weak normalized rich-club phenomenon* at the richness metric $r_{\hat{H}}$ and its threshold value t, while the corresponding weak submultihypergraph $\hat{H}\left(r_{\hat{H}},t\right)$ is called the *complete (partial) weak normalized rich-club* of $\hat{H}$. Conversely, if all four (only from one to three) conditions $\alpha_s^n\left(\hat{H},r_{\hat{H}},t\right)>1$, $\beta_s^n\left(\hat{H},r_{\hat{H}},t\right)>1$, $\gamma_s^n\left(\hat{H},r_{\hat{H}},t\right)>1$, and $\delta_s^n\left(\hat{H},r_{\hat{H}},t\right)>1$ are fulfilled, $\hat{H}$ is referred to as exhibiting the *complete (partial) strong normalized rich-club phenomenon* at the richness metric $r_{\hat{H}}$ and its threshold value t, while the corresponding strong submultihypergraph $\hat{H}^*\left(r_{\hat{H}},t\right)$ represents the *complete (partial) strong normalized rich-club* of $\hat{H}$.

4. EXPERIMENTAL ANALYSIS OF THE RICH-CLUB ORGANIZATION IN THE SCIENTIFIC CO-AUTHORSHIP HYPERNETWORK BASED ON THE APPLICATION OF THE PROPOSED METHOD

The role of this section consists in illustrating the descriptive potential of the introduced rich-club coefficients for discovering the hidden forms of the topological ordering underlying the cooperative behavior of the actors in the real-world complex hypernetworks. As the basis for constructing the experimental sample of the multihypergraph model reflecting the co-authorship hypernetwork, this work uses the open-access dataset [15] containing the detailed records of 2 867 scientific articles in the field of information visualization extracted from the IEEE Xplore database. Notice that this dataset stores the many-to-many relationship between the papers and their authors in the form of two columns matching the unique identifiers of papers with the corresponding lists of author names (delimited by semicolons). Thereby, the representation of this relationship in terms of the multihypergraph model requires performing the preliminary processing of data. Firstly, the set of all individual researchers engaged in the considered relationship was constructed by collecting the items obtained after spitting the author list of every article by semicolons with the subsequent deletion of all duplicates. This operation has resulted in the extraction of 5 086 individual researches depicted by the vertices in the multihypergraph model. The next stage of the model construction consisted in representing every article covered in the considered dataset by the corresponding hyperedge containing all vertices depicting the researchers included in the author list of this article. For convenience, throughout the remainder of this section, the notation $\hat{H}$ refers to the particular multihypergraph sample constructed in such manner.

All non-zero items comprising the vertex degree sequence $\Xi\left(\hat{H}\right)$ of $\hat{H}$ are concentrated within the range of indexes $2\le i\le 61$ and, as shown in Fig. 1a, exhibit almost monotonically decaying behavior with the increase in i. The additional inspection has resulted in concluding that the behavior of $\xi_i\left(\hat{H}\right)$ over whole this range is accurately approximated by the power-law function

$$\xi_i\left(\hat{H}\right)=3598.5111\cdot\left(i-1\right)^{-2.3328}$$

(at which both coefficient of determination and its adjusted version exceed 0.99995, thereby, pointing to the extreme goodness of the fit). In turn, all non-zero items of the hyperedge cardinality sequence $\Omega\left(\hat{H}\right)$ associated with $\hat{H}$ have the indexes $1\le i\le 17$. Figure 1b illustrates that the value of $\omega_i\left(\hat{H}\right)$ depends on i in the more complicated way with the presence of the sharp peak at $i=2$ (caused by the rarity of the solely authored articles within the considered dataset). Moreover, the smooth fragment of this dependence at $3\le i\le 14$ is appropriately approximated by the power-law function

$$\omega_i\left(\hat{H}\right)=\left(0.0014+4.7825\times10^{-7}\cdot i^{5.0542}\right)^{-1}$$

Journal of Theoretical and Applied Information Technology
15th December 2018. Vol.96. No 23
© 2005 – ongoing JATIT & LLS
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

(with both coefficient of determination and its adjusted version of more than 0.9991). The rest of this section gives the results of analyzing the rich-club organization in $\hat{H}$ with respect to the degree as the richness metric and the accompanying discussion. Remark that only five nodes in $\hat{H}$ have the degree exceeding 38, while the weak submultihypergraph induced by them contains one non-loop hyperedge. Thereby, we limit the consideration area to the indexing set of the degree thresholds $\{1,2,...,39\}$. Remark that due to the absence of vertices with no incident hyperedges in $\hat{H}$ (since every researcher within the considered dataset serves as the author of at least one article), the submultihypergraphs $\hat{H}\left(deg_{\hat{H}},t\right)$ and $\hat{H}^*\left(deg_{\hat{H}},t\right)$ extracted at the degree threshold t of one are identical to the whole $\hat{H}$.

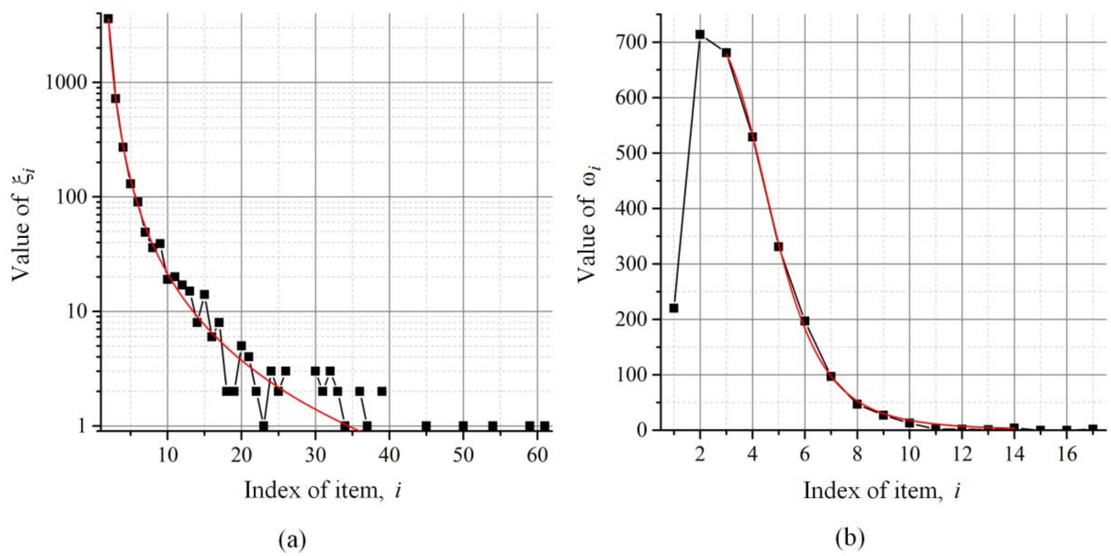

Figure 1: Vertex degree $\Xi\left(\hat{H}\right)$ (a) and hyperedge cardinality $\Omega\left(\hat{H}\right)$ (b) sequences of the multihypergraph model $\hat{H}$ representing the co-authorship hypernetwork given along with the results of their approximation by the power-law functions (refer text for details).

Figure 2 shows that the ordinary rich-club coefficients $\alpha_w\left(\hat{H},deg_{\hat{H}},t\right)$, $\beta_w\left(\hat{H},deg_{\hat{H}},t\right)$, $\gamma_w\left(\hat{H},deg_{\hat{H}},t\right)$, and $\delta_w\left(\hat{H},deg_{\hat{H}},t\right)$ monotonically increase at low t, reach the global maximum, and then demonstrate the falling tendency with the formation of multiple local maxima and minima. Remark that the non-smooth decline of these coefficients at high t is caused by the enhanced influence of fluctuations following from the extreme reduction in the number of vertices contained in the corresponding weak submultihypergraphs $\hat{H}\left(deg_{\hat{H}},t\right)$. At the same time, the coefficient $\delta_w\left(\hat{H},deg_{\hat{H}},t\right)$ achieves the global maximum at the degree threshold $t=4$, while the global maxima of $\beta_w\left(\hat{H},deg_{\hat{H}},t\right)$ and $\gamma_w\left(\hat{H},deg_{\hat{H}},t\right)$ are shifted to the right and registered at $t=8$. In turn, $\alpha_w\left(\hat{H},deg_{\hat{H}},t\right)$ has the most right-shifted global maximum at $t=12$. Another important difference between the considered dependences consists in the rate of their growth before reaching the global maximum. Looking into detail, at $t\le 3$, the numerical derivatives of these dependences (calculated based on applying the two-point estimation method) are arranged in the decreasing order of their values as follows:

$$\frac{d\beta_w\left(\hat{H},deg_{\hat{H}},t\right)}{dt}>\frac{d\alpha_w\left(\hat{H},deg_{\hat{H}},t\right)}{dt}>$$
$$>\frac{d\gamma_w\left(\hat{H},deg_{\hat{H}},t\right)}{dt}>\frac{d\delta_w\left(\hat{H},deg_{\hat{H}},t\right)}{dt}.$$

On the contrary, at $4\le t\le 8$, $\alpha_w\left(\hat{H},deg_{\hat{H}},t\right)$ becomes the most rapidly growing dependence.

Journal of Theoretical and Applied Information Technology
15th December 2018. Vol.96. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

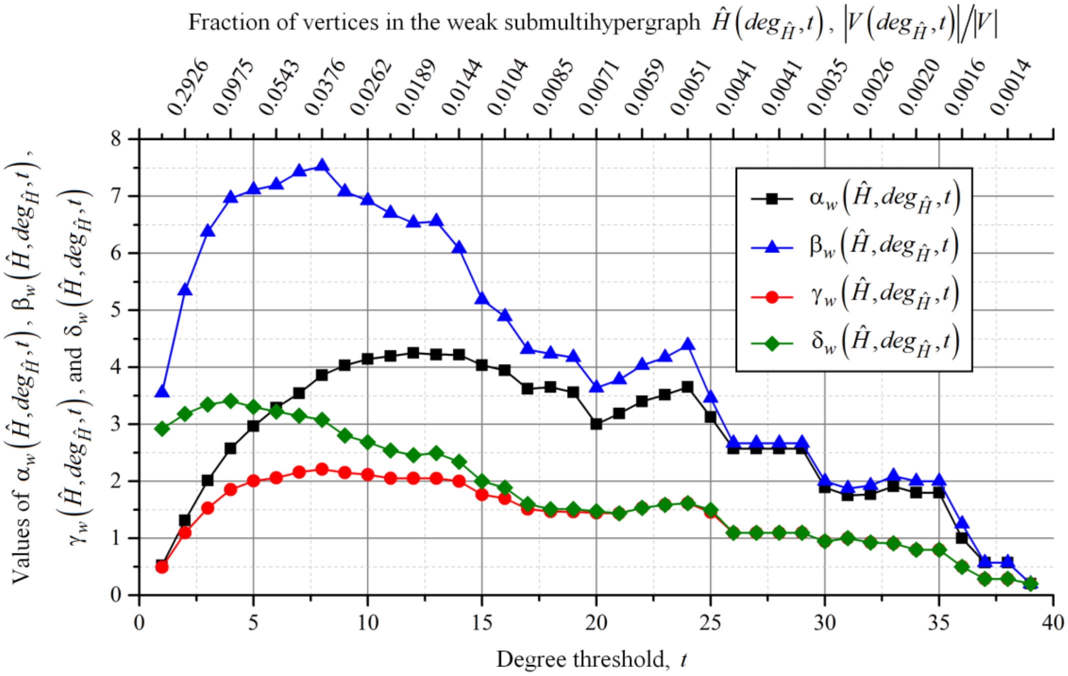

Figure 2: Dependences of the ordinary α -, β -, γ -, and δ -coefficients characterizing the weak form of the rich-club organization in the multihypergraph model $\hat{H}$ of the co-authorship hypernetwork with respect to the metric of degree $deg_{\hat{H}}$ on the threshold value t .

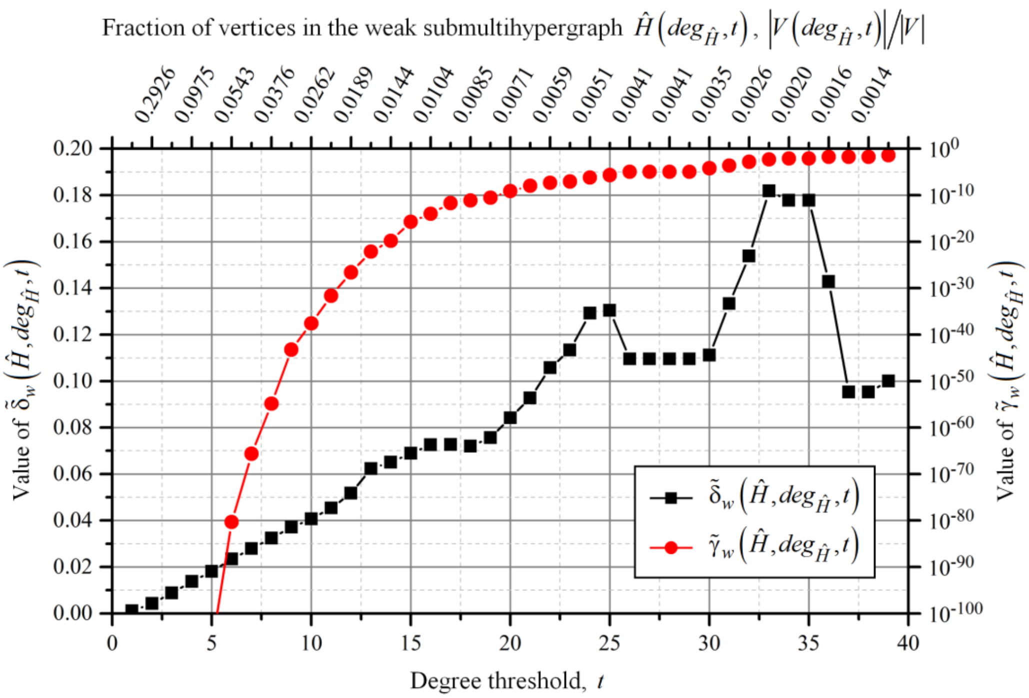

Figure 3: Dependences of the ordinary $\tilde{\gamma}$ - and $\tilde{\delta}$ -coefficients characterizing the weak form of the rich-club organization in the multihypergraph model $\hat{H}$ of the co-authorship hypernetwork with respect to the metric of degree $deg_{\hat{H}}$ on the threshold value t .

In turn, Fig. 3 illustrates that the $\tilde{\gamma}_w\left(\hat{H},deg_{\hat{H}},t\right)$ undergoes the catastrophic growth stemming from the loopless density $\tilde{\lambda}\left(H\left[\hat{H}\right]\right)=2.29310\times10^{-1528}$ of $H\left[\hat{H}\right]$, while $\tilde{\delta}_w\left(\hat{H},deg_{\hat{H}},t\right)$ almost monotonically increases over $t\leq 25$ with the subsequent series of jumps associated with the influence of fluctuations. In sum, all these results allow viewing the constructed model $\hat{H}$ as exhibiting the complete weak ordinary rich-club phenomenon with respect to the metric of degree over $t\leq 4$ (the submultihypergraph $\hat{H}\left(deg_{\hat{H}},4\right)$ contains less than 10% of vertices existing in $\hat{H}$). Moreover, $\hat{H}$ could be interpreted as preserving the partial form of such phenomenon at all higher thresholds t selected for the consideration (with four attesting coefficients over $5\leq t\leq 8$, two over $9\leq t\leq 12$, and only one otherwise).

On the other hand, as illustrated in Figs. 4 and 5, the dependences of the coefficients $\alpha_s\left(\hat{H},deg_{\hat{H}},t\right)$, $\beta_s\left(\hat{H},deg_{\hat{H}},t\right)$, $\gamma_s\left(\hat{H},deg_{\hat{H}},t\right)$, $\delta_s\left(\hat{H},deg_{\hat{H}},t\right)$, and $\tilde{\delta}_s\left(\hat{H},deg_{\hat{H}},t\right)$ on the threshold t do not have any intervals of the well-expressed growth. The only exception is the standalone peak of $\tilde{\delta}_s\left(\hat{H},deg_{\hat{H}},t\right)$ at high t, which could be dismissed as originating from the fluctuations. In turn, the coefficient $\tilde{\gamma}_s\left(\hat{H},deg_{\hat{H}},t\right)$, similarly to $\tilde{\gamma}_w\left(\hat{H},deg_{\hat{H}},t\right)$, demonstrates the meaningless catastrophic growth and almost reproduces the exponential function included in its denominator. These observations allow concluding that neither complete nor partial strong ordinary rich-club phenomenon with respect to the metric of degree is present in $\hat{H}$ to the significant extent.

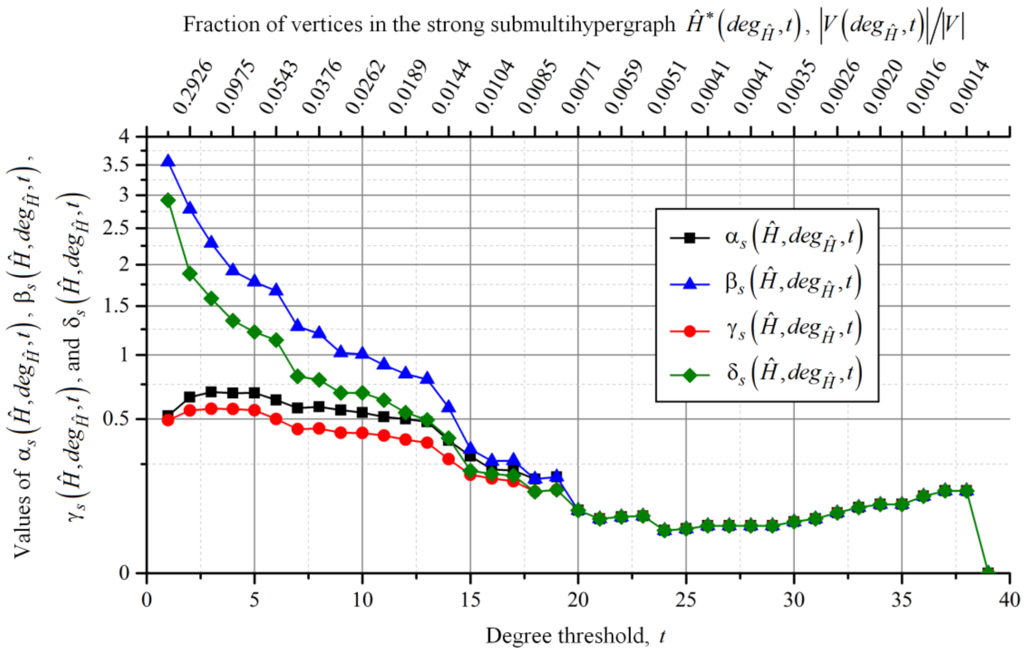

Figure 4: Dependences of the ordinary α-, β-, γ-, and δ-coefficients characterizing the strong form of the rich-club organization in the multihypergraph model $\hat{H}$ of the co-authorship hypernetwork with respect to the metric of degree $deg_{\hat{H}}$ on the threshold value t.

At the next stage of experiment, the calculation of the normalized rich-club coefficients was performed based on the reference ensemble $\varepsilon\left(\hat{H}\right)$ filled with one hundred random multihypergraphs produced by the generating algorithm proposed in the previous section. In particular, Fig. 6 shows that among the coefficients $\alpha_w^n\left(\hat{H},deg_{\hat{H}},t\right)$, $\beta_w^n\left(\hat{H},deg_{\hat{H}},t\right)$, $\gamma_w^n\left(\hat{H},deg_{\hat{H}},t\right)$, and $\delta_w^n\left(\hat{H},deg_{\hat{H}},t\right)$, the value of one is exceeded only by $\beta_w^n\left(\hat{H},deg_{\hat{H}},t\right)$ and only at $2\leq t\leq 14$.

Journal of Theoretical and Applied Information Technology
15th December 2018. Vol.96. No 23

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Thereby, over this relatively wide range of thresholds, the model $\hat{H}$ satisfies the formal criterion for being viewed as exhibiting the partial weak normalized rich-club phenomenon with respect to the metric of degree.

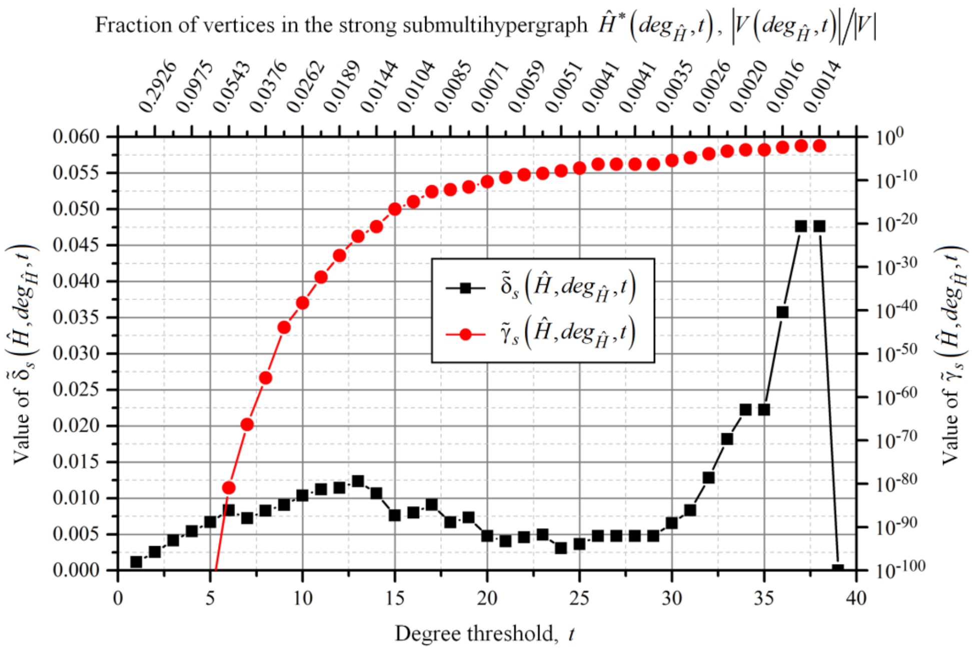

Figure 5: Dependences of the ordinary $\tilde{\gamma}$ - and $\tilde{\delta}$ -coefficients characterizing the strong form of the rich-club organization in the multihypergraph model $\hat{H}$ of the co-authorship hypernetwork with respect to the metric of degree $deg_{\hat{H}}$ on the threshold value t .

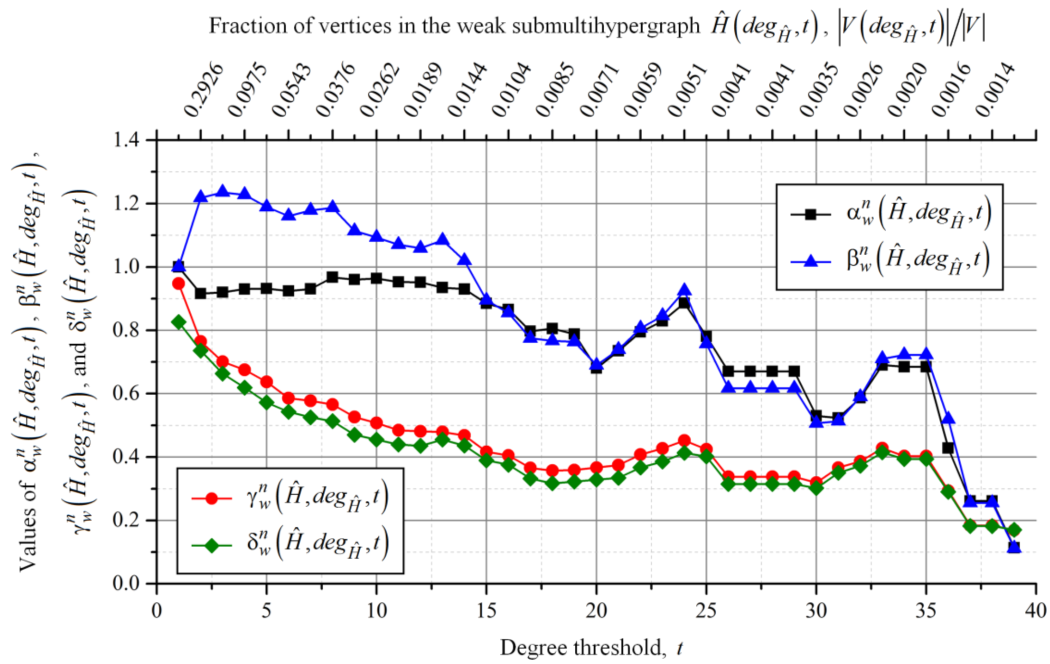

Figure 6: Dependences of the normalized α -, β -, γ -, and δ -coefficients characterizing the weak form of the rich-club organization in the multihypergraph model $\hat{H}$ of the co-authorship hypernetwork with respect to the metric of degree $deg_{\hat{H}}$ on the threshold value t .

At the same time, the dependences of the coefficients $\alpha_s^n\left(\hat{H},deg_{\hat{H}},t\right)$, $\beta_s^n\left(\hat{H},deg_{\hat{H}},t\right)$, $\gamma_s^n\left(\hat{H},deg_{\hat{H}},t\right)$, and $\delta_s^n\left(\hat{H},deg_{\hat{H}},t\right)$ on t illustrated in Fig. 7 are, in general, characterized by the more sharp decline compared to the dependences given in Fig. 6. Only $\beta_s^n\left(\hat{H},deg_{\hat{H}},t\right)$ demonstrates the small peak and exceeds the value of one at $t=2$. However, both extreme narrowness and low absolute value of this peak allow dismissing it as the anomaly. Along with this remark, all obtained results point to the absence of the strong normalized rich-club phenomenon with respect to $deg_{\hat{H}}$ manifested to the considerable extent in the examined multihypergraph $\hat{H}$.

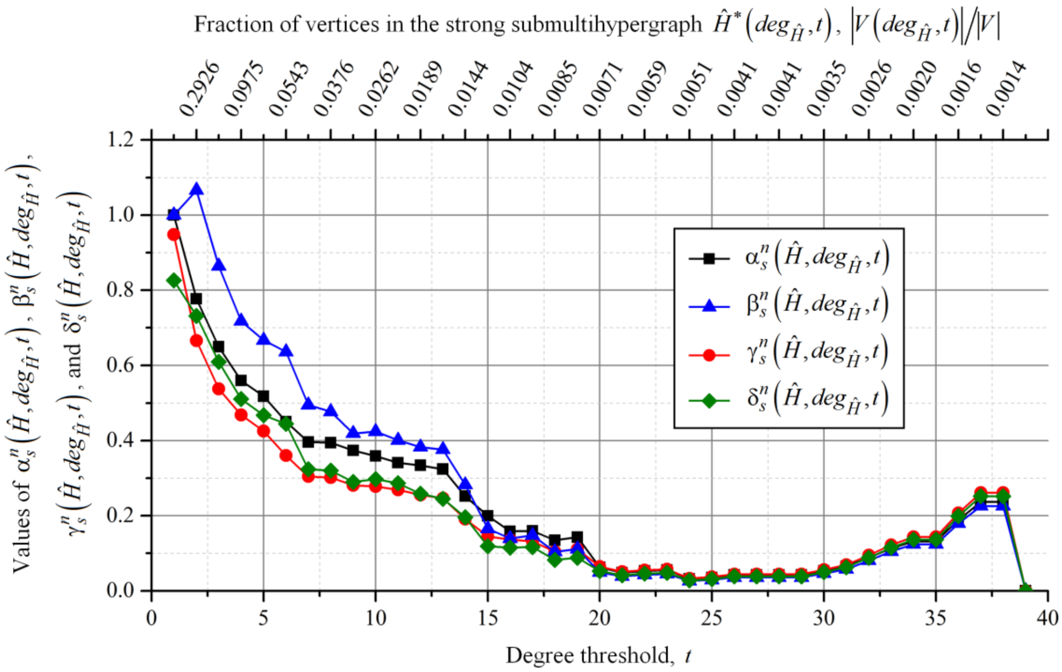

Figure 7: Dependences of the normalized α-, β-, γ-, and δ-coefficients characterizing the strong form of the rich-club organization in the multihypergraph model $\hat{H}$ of the co-authorship hypernetwork with respect to the metric of degree $deg_{\hat{H}}$ on the threshold value t.

5. CONCLUSIONS

The method for detecting and assessing the rich-club organization in an arbitrary multihypergraph $\hat{H}\in\Gamma_{\hat{H}}$ designed in this work is based on introducing the family of twelve ordinary and eight normalized rich-club coefficients. In particular, the ordinary coefficients embody all possible combinations of six metrics $\tilde{\sigma}$, θ, $\tilde{\tau}$, η, $\tilde{\mu}$, and ζ characterizing the tightness of interconnections among the multihypergraph's vertices and two non-equivalent types of submultihypergraphs (weak and strong). For its part, the normalized coefficients represent the results of correcting the corresponding ordinary coefficients with respect to the reference ensemble of random multihypergraphs possessing the same sequences of vertex degrees and hyperedge cardinalities as the examined multihypergraph.

The experimental analysis of the multihypergraph-based model constructed for real-world co-authorship hypernetwork has attested the descriptiveness of all introduced coefficients, excepting $\gamma_w\left(\hat{H},r_{\hat{H}},t\right)$ and $\gamma_s\left(\hat{H},r_{\hat{H}},t\right)$. In total, the proposed method allows discovering eight fundamental forms of the rich-club phenomenon in multihypergraphs (i.e. all combinations of complete/partial, weak/strong, and ordinary/normalized). At the same time, even the most basic substructures of simple graphs (such as cycles) are known to have multiple non-equivalent conceptual generalizations to the cases of simple hypergraphs and multihypergraphs. Thereby, a variety of rich-club coefficients and forms of the rich-club phenomenon in multihypergraphs introduced in this work serves as the natural implication of their complex underlying combinatorial nature.

For comparison, in view of their applicability only to the class of simple graphs, the concepts of the non-normalized and normalized rich-club coefficients reported in the existing literature [11, 12] allow organizing only the indirect assessment of the rich-club ordering in the multihypergraph $\hat{H}$ through analyzing its approximation $G\left[\hat{H}\right]$. The fundamental limitation of such approach consists in the impossibility of detecting all natural forms of the rich-club phenomenon in $\hat{H}$ classified in this article, which clearly points to the conceptual novelty of the proposed method and its enhanced descriptive potential.

REFERENCES

[1] Johnson JH. Embracing n-ary relations in network science. In: Wierzbicki A, Brandes U, Schweitzer F, Pedreschi D, editors. Advances in Network Science. NetSci-X 2016. Lecture Notes in Computer Science 9564. Cham: Springer; 2016. p. 147–60. doi: 10.1007/978-3-319-28361-6_12

[2] Estrada E, Rodriguez-Velazquez JA. Subgraph centrality and clustering in complex hyper-networks. Physica A: Statistical Mechanics and its Applications. 2006 May 15; 364:581–94. doi: 10.1016/j.physa.2005.12.002

[3] Potebnia A. Formation of the multifractal hypergraph structure reflecting the self-similarity properties of the computational complexity classes. In: 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON). Piscataway: IEEE; 2017. p. 953–8. doi: 10.1109/ukrcon.2017.8100392

[4] Potebnia A. Representation of the greedy algorithms applicability for solving the combinatorial optimization problems based on the hypergraph mathematical structure. In: 2017 14th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM). Piscataway: IEEE; 2017. p. 328–32. doi: 10.1109/CADSM.2017.7916145

[5] Dewar M, Healy J, Perez-Gimenez X, Prałat P, Proos J, Reiniger B, et al. Subgraphs in non-uniform random hypergraphs. In: Bonato A, Graham F, Pralat P, editors. Algorithms and Models for the Web Graph. WAW 2016. Lecture Notes in Computer Science 10088. Cham: Springer; 2016. p. 140–51. doi: 10.1007/978-3-319-49787-7_12

[6] Han Y, Zhou B, Pei J, Jia Y. Understanding importance of collaborations in co-authorship networks: A supportiveness analysis approach. In: Proceedings of the 2009 SIAM International Conference on Data Mining. Philadelphia: SIAM; 2009. p. 1112–23. doi: 10.1137/1.9781611972795.95

[7] Potebnia A. Innovative concept of the strict line hypergraph as the basis for specifying the duality relation between the vertex separators and cuts. In: Shakhovska N, Stepashko V, editors. Advances in Intelligent Systems and Computing II. CSIT 2017. Advances in Intelligent Systems and Computing 689. Cham: Springer; 2018. p. 386–403. doi: 10.1007/978-3-319-70581-1_28

[8] Klamt S, Haus U-U, Theis F. Hypergraphs and cellular networks. PLoS Computational Biology. 2009 May 29; 5(5):e1000385. doi: 10.1371/journal.pcbi.1000385

[9] Cinelli M, Ferraro G, Iovanella A. Resilience of core-periphery networks in the case of rich-club. Complexity. 2017 Dec 28; 2017:6548362. doi:10.1155/2017/6548362

[10] Csigi M, Korosi A, Biro J, Heszberger Z, Malkov Y, Gulyas A. Geometric explanation of the rich-club phenomenon in complex networks. Scientific Reports. 2017 May 11; 7:1730. doi: 10.1038/s41598-017-01824-y

[11] Colizza V, Flammini A, Serrano MA, Vespignani A. Detecting rich-club ordering in complex networks. Nature Physics. 2006 Jan 15; 2(2):110–5. doi: 10.1038/nphys209

[12] van den Heuvel MP, Sporns O. Rich-club organization of the human connectome. Journal of Neuroscience. 2011 Nov 2; 31(44):15775–86. doi: 10.1523/jneurosci.3539-11.2011

[13] Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D. Complex networks: Structure and dynamics. Physics Reports. 2006 Jan 10; 424(4–5):175–308. doi: 10.1016/j.physrep.2005.10.009

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[14] Bhuiyan H, Khan M, Marathe M. A parallel algorithm for generating a random graph with a prescribed degree sequence. In: IEEE International Conference on Big Data (Big Data). Piscataway: IEEE; 2017. p. 3312–21. doi: 10.1109/bigdata.2017.8258316

[15] Isenberg P, Heimerl F, Koch S, Isenberg T, Xu P, Stolper C, et al. Visualization Publication Dataset [Internet]. 2015 Jun [cited 2018 Apr 21]; Available from: http://www.vispubdata.org/. Archived copy of the dataset on which the experiment was performed is available from: https://goo.gl/CD6u5D