A Flow Approach to Communities Detection in Complex Network and Multilayer Network Systems

Olexandr Polishchuk

Pidstryhach Institute for Applied Problems of Mechanics and Mathematics, National Academy of Sciences of Ukraine, Naukova str, 3"b", Lviv, 79060, Ukraine

Abstract

A flow approach to community detection in complex network and multilayer network systems is proposed. Two methods have been developed to search for communities in a network system (NS). The first of them is based on the calculation of flow influence parameters of NS's subsystems, selected according to the principle of nesting hierarchy. The second method uses the concept of flow core of network system. Two methods are also proposed for community detection in multilayer network system (MLNS). The first of them is based on the concept of MLNS aggregate-network and subsequent allocation of its flow core. The second method uses the concept of flow core of the process of intersystem interactions in general. All developed methods are based on the use of flow criterion that the selected group of nodes really forms a community. The results of application of developed approaches are illustrated by examples for which known methods are ineffective.

Keywords [1]

Complex network, Network system, Intersystem interactions, Flow model, Hierarcy, Flow core, Aggregate-network, Influence, Community

1. Introduction

One of important problems investigated in the theory of complex networks (TCN) is the search of groups of interconnected nodes, the identification of which contributes to a better understanding of principles of organizing the structure and operation processes of complex network systems. In real NS, the most common groups are so-called communities – subnets, the connections between nodes of which are denser and stronger than between them and other network nodes [1, 2]. Communities exist in the physical world, wildlife, economy, transportation, urban infrastructure [3, 4], etc. In human society, communities can be considered public organizations, political parties, religious denominations, national diasporas, groups in social networks [5, 6] and so on. Currently, the main attention is paid to the development of communities detection methods, which are based on the structural characteristics of network systems – the smallest cut, hierarchical clustering, modularity or entropy evaluation, spectral properties of network or random walk [7, 8], etc.

No less important and difficult is the problem of finding communities in MLNS, which describe the processes of intersystem interactions in suprasystem formations of various types [9, 10]. In this case, the methods and approaches listed above are usually also used [11]. The main drawback of known communities detection methods, along with the computational complexity and resource intensity, is the lack of reliable theoretically based criterion that the group of nodes determined by any of these methods really forms a community, because if the term "network density" in TCN is sufficiently clear and easily calculated by well-known formula, the concept of "stronger" or "weaker" connection from a structural point of view is not sufficiently unambiguous [8]. This circumstance sometimes compels the use of visual research methods [12]. An additional drawback of existing structural methods is that they are usually aimed at finding already formed and sufficiently stable

Information Technology and Implementation (IT&I-2023), November 20-21, 2023, Kyiv, Ukraine
EMAIL: od_polishchuk@ukr.net (Olexandr Polishchuk)
ORCID: 0000-0002-0054-7159 (Olexandr Polishchuk)

communities, which consist of a sufficiently large number of nodes, but do not track the appearance of such communities in the network and their rapid development (increase, decrease, disappearance). Even dynamic structural models, i.e. models that take into account changes in the structure of NS and MLNS over time, are generally not able to solve this problem [4]. At the same time, in modern society there are many important and massive events organized by communities of various orientations, the course of which is limited to a few weeks or even days. The large number of existing methods for communities detection in NS and MLNS indicates a great interest in this issue and its importance in system research [13]. The purpose of this article is to develop criteria and methods for finding communities in such entities based on flow models of complex network and multilayer network systems.

2. Dynamics of communities formation and development in human society

One of the most interesting and relevant objects of research in TCN are communities that arise in human society and in one way or another influence its development [6, 14]. Beginning with primitive tribes, such communities have often played a significant role in the historical process, both positive and negative. The creation of world religions and new states, changes in socio-economic formations always began with small, but highly motivated and sufficiently active groups of like-minded people. The emergence and activity of such groups usually had a positive effect on the development of society (communities of collectors created museums and libraries, lovers of culture - philharmonic societies and art galleries, scientists – universities and research laboratories, etc.). At the same time, the emergence and spread of racist, fascist, communist and other misanthropic ideologies had a negative impact on the course of historical processes. The events of recent years show that the threat of recurrence of similar phenomena, and with much more catastrophic consequences, has not disappeared. Along with this, diverse terrorist (Al-Qaeda, ISIS), hacker (Anonymous, LockBit), organized criminal (mafia, drug cartels) groups and religious sects (People's Temple, Aum Shinrikyo) constantly arose and are still emerging, which to one degree or another influenced and often now affect public safety and peace of citizens. Relatively small communities are constantly emerging that create suicidal moods in teenagers (Blue whale), force them to organize simultaneous mass fights in many cities of several countries around the world (PVC Redan), make them pessimistic about their future, tempt them to consume narcotic substances or involve them in extremist organizations of various kinds. The identification of such communities has not only scientific interest, but also the great social importance, since stopping their activities before proceeding to specific actions allows avoiding many victims and broken destinies.

The spread and development of world religions continued for centuries, Nazi and communist ideologies for decades, and various criminal groups for years. In today's world, with the development of information and communication technologies (ICT), the formation of communities can take days and even hours. That is, if previously such processes took years, decades and even centuries and were prompted by serious crisis situations, such as wars, famine, epidemics of dangerous infectious diseases, now with the use of social networks, the birth and activation of communities can be carried out very quickly. Usually, such processes are provoked by incorrect political and economic decisions or actions that disturb social consciousness (violation of human rights, inadequate doings of the police or authorities and so on). Only the beginning of 21^{st} century is full of such events – Maidans in Ukraine in 2004 and 2013, revolutions in Georgia, Kyrgyzstan, Libya, Tunisia, political disturbances in Kazakhstan, Belarus and France, etc. The defining feature of these events was the speed of unification of large groups of people and their mass performances, which was practically impossible in "pre-informatization" times. Simultaneously, such communities arose precisely in civil society, and social networks, as a tool of ICT, served as means that contributed to their formation as soon as possible. However, this tool makes it possible to quantitatively monitor the process of birth and development of such communities. It should also not be forgotten that many communities of different directions and interests exist in social networks themselves, which is also an interesting phenomenon of human life.

Communities can exist both in separate layers-systems of MLNS, and in the process of interactions between them, constantly arising, combining, overlapping or leveling each other. Therefore, for a better understanding of intersystem interactions processes, the search for communities must be carried out both in separate layers and in MLNS as a whole. Communities in the modern world, in particular civil and social, are usually quite dynamic structures that can appear, develop, and disappear quickly, and the methods of identifying such formations must take this feature into account. Structural models of complex network systems and intersystem interactions are usually not able to fully solve this problem. Therefore, dynamic models that describe the operation processes of NS and MLNS become extremely important. Let's consider a flow approach to communities detection in such systems and intersystem formations.

3. Structural and flow models of multilayer network system

The structure of intersystem interactions is described by multilayer networks (MLNs) and represented in the form [1]

$$G^M = \left(\bigcup_{m=1}^{M} G_m,\ \bigcup_{m,k=1,\ m\neq k}^{M} E_{mk} \right), \tag{1}$$

where $G_m = (V_m, E_m),\ G_m \in R^n,\ n = 2,3,$ determines the structure of m^{th} network layer of MLN; V_m is the set of nodes of network G_m; E_m is the set of edges of network G_m, E_{mk} is the set of connections between the nodes of sets V_m and V_k, $m \neq k$, $m,k = \overline{1,M}$, M is a number of layers (interconnected systems) of MLN. The set

$$V^M = \bigcup_{m=1}^{M} V_m$$

will be call the total set of nodes, and the set

$$E^M = (\bigcup_{m=1}^{M} E_m) \bigcup (\bigcup_{m,k=1,\ m\neq k}^{M} E_{mk})$$

will be call the total set of edges of multilayer network, N^M, L^M are the numbers of elements of the sets V^M and E^M respectively.

The multilayer network G^M is completely described by the adjacency matrix

$$\mathbf{A}^M = \{\mathbf{A}^{km}\}_{m,k=1}^{M},$$

in which the value a_{ij}^{km}=1 if there is an edge connecting nodes n_i^k and n_j^m, and a_{ij}^{km}=0, $i, j = \overline{1, N^M}$, if there is no such edge. At the same time, blocks $\mathbf{A}^{mm}$ describe the structure of intralayer interactions in m^{th} layer, and blocks $\mathbf{A}^{km}$ describe the structure of interlayer interactions between m^{th} and k^{th} layers of MLN, $m \neq k$, $m,k = \overline{1,M}$. If all blocks of the matrix $\mathbf{A}^M$ are defined for the total set of MLN nodes, then the problem of coordination of node numbers in the case of their independent numbering for each layer is removed.

Most real-world intersystem interactions are multipurpose and multifunctional. This is primarily expressed in the multiflow nature of such formations, i.e. ensuring the movement of various types of flows. In TCN, the structure of such intersystem interactions is represented by so-called multidimensional networks [15]. A multidimensional network is MLN, in which each layer reflects the structure of system, which ensures the movement of flows a type of which is generally different from flows in other layers. As an example, consider a general transport system that provides the movement of two main types of flows – passenger and cargo, that is, its structure can be depicted in the form of two-dimensional network. A feature of this structure, like most multidimensional networks, is the impossibility of flow transition from one layer to another (transformation of passengers into cargo and vice versa). To simplify the analysis of intersystem interactions process in two-dimensional general transport system, it can be divided into two four-layer monoflow MLNSs, the layers of which (railway, road, aviation and water) ensure the movement of only one type of flow – passenger or cargo. A characteristic feature of monoflow transport MLNS is the difference of flow carriers in each layer (trains, motor vehicles, planes, ships). In general, when detailing the structure of real multidimensional networks, it is advisable at first distinguish the layers that ensure the movement of various types of flows, and then depict each of these monoflow layers as MLNS, each layer of which ensures the movement of these flows by a specific carrier or operator system. Communication in social and other information and communication networks is carried out by exchanging information flows. That is, such formations can be considered as monoflow multilayer systems. Separate layers of such systems usually reflect the operation process various systems-operators of information flows, as it happens in the systems of mobile and fixed phone communication, cable and satellite television, e-mail and regular mail, Internet, social networks and so on.

We will represent the flow model of monoflow MLNS [16] in the form of adjacency matrix $\mathbf{V}^M(t)$, the elements of which are determined by the volumes of flows that have passed through the edges of MLNS (1) during period $[t-T,t]$ until the current moment of time $t \geq T$:

$$\mathbf{V}^M(t) = \{V_{ij}^{km}(t)\}_{i,j=1}^{N},\ {}_{k,m=1}^{M}, \quad V_{ij}^{km}(t) = \frac{\tilde{V}_{ij}^{km}(t)}{\max\limits_{s,g=\overline{1,M}}\ \max\limits_{l,p=\overline{1,N^M}} \{\tilde{V}_{lp}^{sg}(t)\}}, \tag{2}$$

where

$$\tilde{V}_{ij}^{km}(t) = \int_{t-T}^{t} v_{ij}^{km}(\tau)d\tau\,; \quad v_{ij}^{km}(t) = \int_{(n_i^k,n_j^m)} \rho_{ij}^{km}(t,\mathbf{x})dl; \quad \boldsymbol{\rho}(t,x) = \{\rho_{ij}^{km}(t,\mathbf{x})\}_{i,j=1}^{N^M},\ {}_{k,m=1}^{M};$$

and $\rho_{ij}^{km}(t,\mathbf{x})$ is the density of flow that pass through the edge (n_i^k, n_j^m) MLNS in the current moment of time $t > 0$, $\mathbf{x} \in (n_i^k, n_j^m) \subset R^n$, $n = 2, 3, \ldots$, $i, j = \overline{1, N^M}$, $k, m = \overline{1, M}$. The elements of flow adjacency matrix $\mathbf{V}^M(t)$ are determined on the basis of empirical data about movement of flows through its edges. Currently, with the help of modern means of information extraction, it is quite easy to obtain such data for many natural and the vast majority of man-made systems, including information NS [17]. The matrix $\mathbf{V}^M(t)$ has a block structure, in which the diagonal blocks $\mathbf{V}^{mm}(t)$ describe the volumes of intralayer flows in the m^{th} layer, and the blocks $\mathbf{V}^{km}(t)$ describe the volumes of flows between the m^{th} and k^{th} layers of MLNS, $m \neq k$, $m,k = \overline{1,M}$.

We calculate the values of matrix $\mathbf{V}^M(t)$ elements on the time interval $[t-T,t]$, $t \geq T$, in order to level out random disturbances that may occur during the movement of flows at certain moments of time. These values are dynamic, as they are determined up to the current moment of time, and therefore change continuously. The duration of interval T depends on the dynamics of system's behavior. For example, for the mass social disturbances mentioned in the previous section, which took place in many countries of the world, this interval should not usually exceed one day. This is evidenced by the fact that even on November 29, 2013, almost no one could predict that the beating of students in Kyiv on the night of November 30 would provoke the appearance of a new Maidan in Ukraine almost the next day. For models of Covid-19 spreading, the interval T (to smooth out the difference in the number of newly diagnosed infections on weekdays and weekends) was usually equal to a week [18]. Communities can arise both in separate layers-systems and in MLNS as a whole. At the same time, they can both strengthen, overlap or intersect, and level each other. Thus, the writing society is quite clearly divided into communities of authors of detective, fantasy, historical and other genres. However, the author of crime novels can introduce into them elements of fiction or historical events, fiction – a detective or historical component, etc. Simultaneously, communities of the sports society, which consist of athletes of various kinds of sport, intersect much less. Indeed, it is rare that a football player does weightlifting, and a weightlifter does chess or marathon running at the same professional level. This means that it is appropriate to start the communities detection from separate layers of multilayer network system.

4. Communities in layers of multilayer network systems

To simplify the presentation, in this section we will denote the flow adjacency matrix of arbitrary layer-system of MLNS as $\mathbf{V}(t) = \{V_{ij}(t)\}_{i,j=1}^{N}$, the elements of which $V_{ij}(t)$ are equal to the relative volumes of flows that passed through the edge (n_i, n_j), $i, j = \overline{1, N}$, of this layer during the time period $[t-T, t]$, $t \geq T$, N is the number of layer nodes. Let us consider two approaches to defining communities in such network system.

4.1. Communities detection based on the nestedness hierarchy

In real large scale systems, the first "candidates" for communities are subsystems of different levels of hierarchy built according to the nesting principle, when the smaller is part of the larger [19] (Fig. 1). Indeed, students of class communicate more among themselves than with students of other classes or courses, people of certain professions – more often than with representatives of other specialties, representatives of different social groups or age categories also prefer to communicate with people of the same groups or categories. That is, to singled out by a certain feature of homogeneity of elements, NS's subsystems have a higher probability for the formation of community.

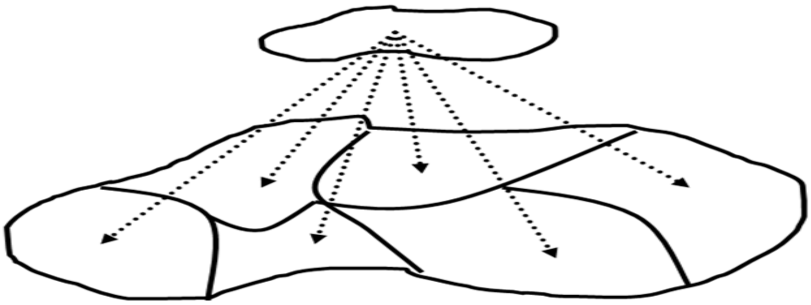


Figure 1: An example of hierarchy built on the nesting principle

Let us the source network system S is divided according to the nesting principle into L subsystems of the lowest level of hierarchy

$$S_l \subset S = \bigcup_{l=1}^{L} S_l ,$$

sets of nodes $H_{S_l} = \{n_i^l\}_{i=1}^{N_l}$ of which do not intersect, N_l is the number of subsystem S_l nodes, $l = \overline{1, L}$. Denote by $G_{S_l}^{out}$ the set of all nodes-generators of flows distributed by the network system S, which are included in the set H_{S_l}. Define using parameter

$$\xi_{S_l}^{out}(t) = \sum_{i \in G_{S_l}^{out}} \xi_i^{out}(t) / s(\mathbf{V}(t))$$

the power of influence of subsystem S_l on the NS in general. Here, $\xi_i^{out}(t)$ is the volume of output flows generated at node n_i from the set $G_{S_l}^{out}$ and spread out through the network system S, and $s(\mathbf{V}(t))$ is the total volume of flows that passed through the system S during the time interval $[t–T, t]$, i.e.

$$s(\mathbf{V}(t)) = \sum_{i,j=1}^{N} V_{ij}(t) , \ t \geq T .$$

Let us

$$R_{S_l}^{out} = \bigcup_{i \in G_{S_l}^{out}} R_{l,i}^{out}$$

is the set of numbers of nodes – final receivers (not transitors) of flows generated in the nodes of the set $G_{S_l}^{out}$. Let's divide the set $R_{S_l}^{out}$ into two subsets, namely

$$R_{S_l}^{out} = R_{S_l, int}^{out} \cup R_{S_l, ext}^{out} ,$$

where $R^{out}_{S_l,int}$ is the subset of $R^{out}_{S_l}$ which belong to H_{S_l}, and $R^{out}_{S_l,ext}$ is the subset of nodes $R^{out}_{S_l}$ which belong to complement of H_{S_l} in the sourse network system *S*. The set $R^{out}_{S_l,ext}$ will be call the domain of output influence of subsystem S_l on NS. External and internal output power of influence of the nodes – generators of flows belonging to the set $G^{out}_{S_l}$ on subnets $R^{out}_{S_l,ext}$ and $R^{out}_{S_l,int}$ determines by means of parameters

$$\xi^{out}_{S_l,ext}(t) = \sum_{i \in R^{out}_{S_l,ext}} \xi^{out}_i(t)/s(\mathbf{V}(t)) \quad \text{and} \quad \xi^{out}_{S_l,int}(t) = \sum_{i \in R^{out}_{S_l,int}} \xi^{out}_i(t)/s(\mathbf{V}(t))$$

respectively. Then the value

$$\varpi^{out}_{S_l}(t) = \frac{\xi^{out}_{S_l,ext}(t)}{\xi^{out}_{S_l,int}(t)}$$

determines the relative power of influence of subsystem S_l on the network system *S* in general. Namely, the smaller the value of parameter $\varpi^{out}_{S_l}(t)$, $t \geq T$, the smaller is the influence of subsystem S_l on NS. In other words, in this case, the flow connections between the generator nodes and nodes – final receivers of flows are much stronger within the subsystem S_l, $l=\overline{1,L}$, than between them and other nodes – receivers of flows in the system *S* at a whole.

Denote by $G^{in}_{S_l}$ the set of all nodes– final receivers of flows which are included in the set H_{S_l}. Define using a parameter

$$\xi^{in}_{S_l}(t) = \sum_{i \in G^{in}_{S_l}} \xi^{in}_i(t)/s(\mathbf{V}(t))$$

the power of influence of network system *S* on subsystem S_l. Here, $\xi^{in}_i(t)$ is the volume of input flows generated by the nodes of system *S* and finally received in the node n_i from the set $G^{in}_{S_l}$. Let us

$$R^{in}_{S_l} = \bigcup_{i \in G^{in}_{S_l}} R^{in}_{l,i}$$

is the set of numbers of nodes-generators of flows which finally received in the nodes of the set $G^{in}_{S_l}$. Let's divide the set $R^{in}_{S_l}$ into two subsets, namely

$$R^{in}_{S_l} = R^{in}_{S_l,int} \cup R^{in}_{S_l,ext},$$

where $R^{in}_{S_l,int}$ is the subset of $R^{in}_{S_l}$ which belong to H_{S_l}, and $R^{in}_{S_l,ext}$ is the subset of nodes $R^{in}_{S_l}$ which belong to complement of H_{S_l} in the sourse network system *S*. The set $R^{in}_{S_l,ext}$ will be call the domain of input influence of network system *S* on subsystem S_l. External and internal input power of influence of the nodes – finally receivers of flows belonging to the set $G^{in}_{S_l}$ on subnets $R^{in}_{S_l,ext}$ and $R^{in}_{S_l,int}$ determines by means of parameters

$$\xi^{in}_{S_l,ext}(t) = \sum_{i \in R^{in}_{S_l,ext}} \xi^{in}_i(t)/s(\mathbf{V}(t)) \quad \text{and} \quad \xi^{in}_{S_l,int}(t) = \sum_{i \in R^{in}_{S_l,int}} \xi^{in}_i(t)/s(\mathbf{V}(t))$$

respectively. Then the value

$$\varpi^{in}_{S_l}(t) = \frac{\xi^{in}_{S_l,ext}(t)}{\xi^{\text{int}}_{S_l,int}(t)}$$

determines the relative power of influence of the system *S* on subsystem S_l. Namely, the smaller the value of parameter $\varpi^{in}_{S_l}(t)$, $t \geq T$, the smaller is the influence of the system *S* on subsystem S_l. In other words, in this case, the flow connections between the nodes – final receivers and nodes-generators of flows are much stronger within the subsystem S_l, $l=\overline{1,L}$, than between them and other nodes-generators of flows in the system S at a whole.

It is natural to assume that the greater the volumes of flow movement between two NS nodes, the stronger the relationship between them. This statement defines a sufficiently justified criterion for the existence of community within a certain group (subsystem) of NS nodes. Therefore, a pair of parameters $(\varpi_{S_l}^{out}(t), \varpi_{S_l}^{in}(t))$, namely the compatible fulfillment of conditions

$$\varpi_{S_l}^{out}(t) \le \varpi_C << 1, \quad \varpi_{S_l}^{in}(t) \le \varpi_C << 1, \tag{3}$$

where ϖ_C is a predetermined value, makes it possible to determine a sufficiently objective criterion that the subsystem S_l forms a community in the network system S. Indeed, the smaller the value of these parameters, the smaller the external interaction of subsystem S_l, $l=\overline{1,L}$, with the system S as a whole and the greater the interactions within this subsystem, which is, in fact, the definition of community.

Summarizing, since by definition the community is considered as a certain group of nodes (subsystem S^*) of system S, the connections between which are denser (structural indicator) and stronger (functional indicator) than between them and other nodes of network system S, then objective criteria for existence of community within the subsystem S^* can be considered:

1) indicator of relative density of connections (edges) of subsystem S^*, which is calculated according to the formula

$$\mu(S^*, S) = \left(\frac{L^*}{N^*}\right) \Big/ \left(\frac{L}{N}\right)$$

where N^* is the number of nodes and L^* is the number of connections of subsystem S^*, N is the number of nodes and L is the number of connections of system S;

2) indicator of relative flow power of interconnections of subsystem S^*, which is calculated according to the formula

$$\theta(S^*, S) = \left(\frac{s(\mathbf{V}_{S^*}(t))}{N^*}\right) \Big/ \left(\frac{s(\mathbf{V}(t))}{N}\right),$$

and determines the ratio of specific power of interconnections between nodes in subsystem S^* and system S, respectively. If the indicators of specific density and flow power of interconnections for subsystem S^* significantly exceed the corresponding indicators for system S, then it is quite reasonable to assume that the subsystem S^* forms a community in the source network system. Obviously, the concept of "significantly exceed" is quite vague. Therefore, it is advisable to determine the concrete values of $\mu(S^*, S)$ and $\theta(S^*, S)$, which establish the level of such excess, namely, the fulfillment of conditions

$$\mu(S^*, S) \ge \mu^* >> 1 \tag{4}$$

and / or

$$\theta(S^*, S) \ge \theta^* >> 1, \tag{5}$$

where μ^* and θ^* are the predetermined values. Note that condition (5) is weaker than condition (3) and generally does not impose such strict restrictions on the interaction of subsystem S^* with the rest of system S. In addition, the value

$$\nu(S^*, S) = \frac{s(\mathbf{V}_{S^*}(t))}{s(\mathbf{V}(t))}$$

makes it possible to track the role of subsystem S^* in functioning of network system S and the dynamics of changes in this role over time. If no communities are found at this level of nesting hierarchy, then we move to the next, higher level of this hierarchy. The main drawback of method discussed above is that it is focused on identifying communities formed over a sufficiently long period of time, which enables for the construction of appropriate nesting hierarchies. Consider the method that allows us to track the emergence and dynamic development of communities in network system, focusing primarily on the predominant power of interconnections between its elements.

4.2. Communities and flow cores of network systems

Let us introduce the concept of flow λ-core of network system [20], as the largest subsystem of the source NS, for which the elements of flow adjacency matrix $\mathbf{V}(t)$ satisfy the inequalities

$$V_{ij}(t) \geq \lambda,\ i, j = \overline{1, N},\ t \geq T, \lambda \in [0,1].$$

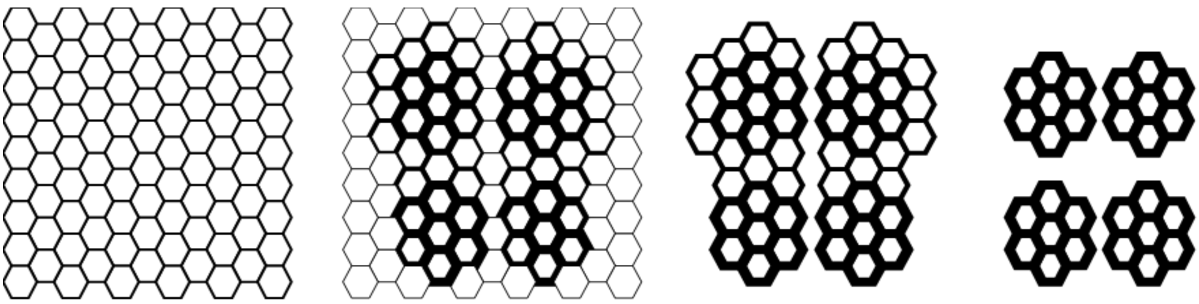
The concept of NS's flow core allows us to build, on the basis of criterion (3) or (5), the following communities detection algorithm in the network system (Fig. 2a – the structure of source NS, 2b – the source NS with reflected volumes of flow movement during the period $[t-T, t]$, $t \geq T$, the thickness of lines is proportional to the volume of flows):

1) take the values $i = 0$, $S = S_{\lambda_0}$, where λ_0 is the minimum nonzero value of parameter $\lambda \in [0,1]$;

2) gradually increase the value λ until the condition (3) or (5) is fulfilled for a certain $\lambda = \lambda_{i+1}$, or at least one of the components of λ_i-core detected earlier is divided into unconnected components (Fig. 2c, 2d);

3) if the value $\lambda_{i+1} < 1$, then accept $i = i+1$ and proceed to point 2, otherwise, finish the execution of algorithm.

By adjusting the value T in the direction of decrease or increase, we make the procedure for communities detection using the NS's flow model and λ-core method more or less sensitive to rapid changes in the structure of detected communities. Note that for the shown in Fig. 2a regular network, criterion (4) is not fulfilled for any of its subnets.



a) NS's structure b) flow distribution by NS's edges c) λ_1-core d) λ_2-core

Figure 2: Use of flow λ-cores for communities detection in network system

However, in the case of irregular networks, this criterion can be used to check whether the connections in NS's subsystem detected by criterion (5) are indeed denser than the average in network. That is, the functional criteria (3) or (5) make it possible to identify communities in the network system for which known structural methods do not work. It is obvious that the structure and composition of nodes and links of communities detected using the algorithm described above is easily determined from the matrix $\mathbf{V}(t)$, $t \geq T$. It is obvious that the NS's flow model enables continuous monitoring of changes in the volume of flows moving between nodes of network system. This makes it possible to monitor the processes of emergence and development of communities in the network almost in real time, which is much more difficult to do with the help of structural methods.

Thus, until 2014, Donbas was one of the most industrially developed region of Ukraine with very close connections between mines, deposits, mining and processing and metallurgical enterprises located on its territory. This was accompanied by the need for a transport and energy infrastructure that was much denser than the national average. In general, such formation can be considered as industrial community. However, after 2014, as a result of closure of a large part of mines and deposits and the cessation of work of many metallurgical plants, this community practically ceased to exist, although the dense transport and energy networks did not disappear anywhere. Many similar examples can be cited: the autoindustry center in Detroit, coal industry in Great Britain and Germany, wine industry in France and many other industrial regions, the demand for products manufactured in them gradually decreased and disappeared, or the mineral deposits mined in these regions were exhausted. That is, from a structural point of view, according to criterion (4), a subsystem may form a community, but functionally, according to criteria (3) or (5), it not form it, and vice versa.

Note, that here and below we intentionally use such simple examples of network system structures, since known numerical and visual structural methods of communities detection practically do not work on them. Similar examples can be given for much more complex real network structures, for example, the system of interconnections between the regions of Ukraine (Fig. 3), which are also generally regular, despite the visual complexity.

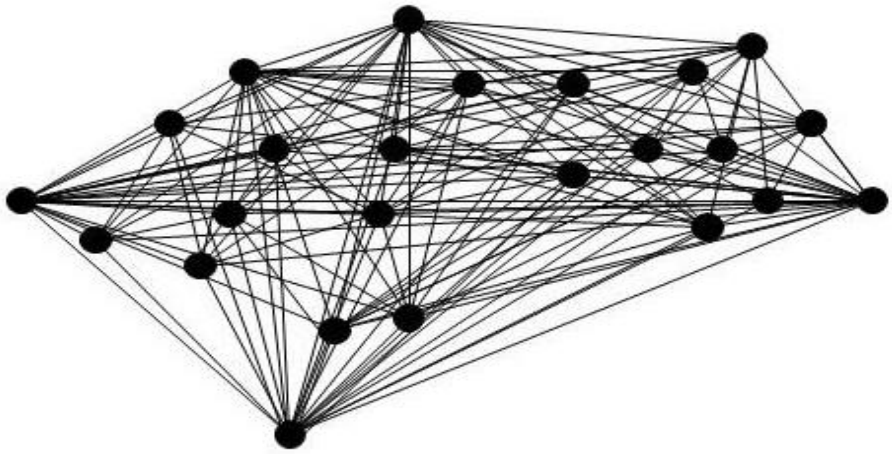


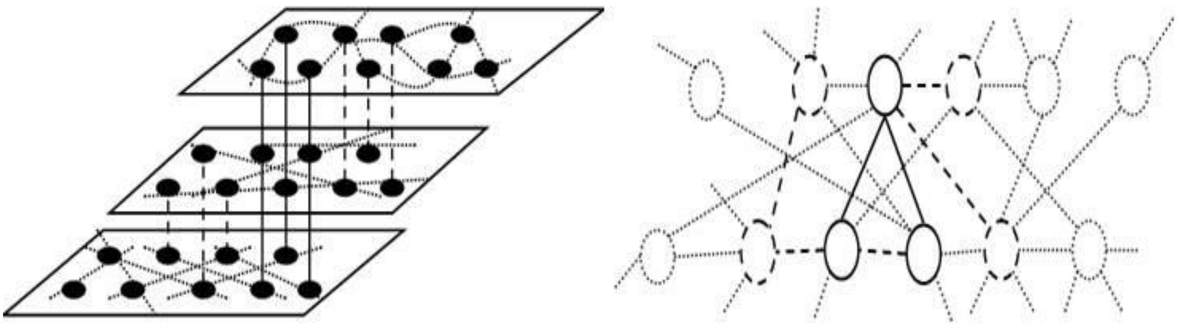
Figure 3: Structure of economical interconnections between regions of Ukrain

5. Aggregate networks and cores of multilayer network systems

The local characteristic ε_{ij} of edge (n_i, n_j) of the general set of multilayer network edges E^M, where n_i and n_j are nodes from the general set of nodes V^M, which we will call its structural aggregate-weight, is the number of layers in which such edge is present. The structural aggregate-weight ε_{ii} of node n_i of MLN is the number of layers of which it is a part, $i, j = \overline{1, N^M}$. For an arbitrary multilayer network, the adjacency matrix $\mathbf{E} = \{\varepsilon_{ij}\}_{i,j=1}^{N^M}$ completely determines the weighted network, which we will call the structural aggregate-network of MLN. The elements of matrix **E** determine the integral structural characteristics of multilayer network's nodes and edges (Fig. 4). For monoflow MLNs, the weight of each edge reflects the number of possible carriers or operator-systems that can ensure the movement of corresponding type of flow and the weight of each node is the number of systems it is a part of. The transition to aggregate-networks can be used to develop structural methods for communities detection in MLN [7]. We will call the structural p_{ag} -core of MLNS aggregate-network the network whose the adjacency matrix elements are determined by the ratio

$$\varepsilon_{ij}^{p_{ag}} = \begin{cases} \varepsilon_{ij}(t), \text{if } \varepsilon_{ij} \geq p_{ag}, \\ 0, \quad \text{if } \varepsilon_{ij} < p_{ag}, \; i, j = \overline{1, N^M}. \end{cases}$$

As the value p_{ag} increases, p_{ag}-cores can be considered to be among the most likely "candidates" in community, since the duplication of connections in the NS usually occurs for two reasons, namely, when these connections are sufficiently important for the system and if through them, the movement of large volumes of flows is distributed.



a) b)

Figure 4: Fragment of three-layer MLN (a) and its aggregate-network (b – ____ – for ε_{ij} =3, _ _ _ – for ε_{ij} =2, ….. – for $\varepsilon_{ij} = 1$, $i, j = \overline{1, N^M}$)

We shall call the flow aggregate-network of MLNS the network system whose elements of the adjacency matrix $\mathbf{F}(t) = \{f_{ij}(t)\}_{i,j=1}^{N^M}$ are determined by the relations

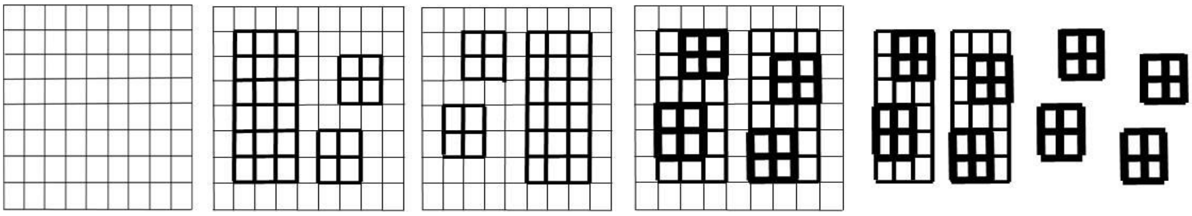
$$f_{ij}(t) = \sum_{k,m=1}^{M} V_{ij}^{km}(t) / M^2 , \; i, j = \overline{1, N^M} , \; t \geq T .$$

The elements of matrix $\mathbf{F}(t)$ determine the integral flow characteristics of nodes and edges of multilayer network system. The flow λ_{ag}-core of aggregate-network of monoflow MLNS is determined using the adjacency matrix, the elements of which are calculated according to the ratio

$$f_{ij}^{\lambda_{ag}}(t) = \begin{cases} f_{ij}(t), & \text{if } f_{ij}(t) \geq \lambda_{ag}, \\ 0, & \text{if } f_{ij}(t) < \lambda_{ag}, \end{cases} \; i, j = \overline{1, N^M}, t \geq T, \; \lambda_{ag} \in [0,1].$$

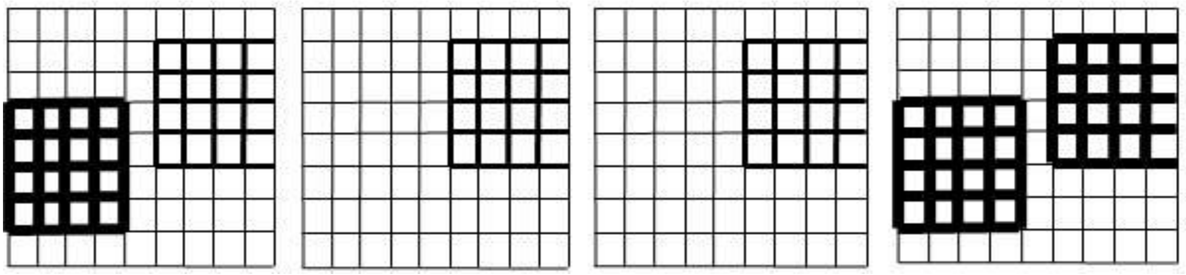
An algorithm described in section 4.2 can be used for communities detection in MLNS's aggregate-network. In fig. 5a – 5c are shown the communities contained in layers 1-3 of three-layer MLNS, respectively. In Fig. 5d shows the flow aggregate-network of MLNS, selected at the moment of time $t \geq T$. Fig. 5e and 5f contain images of communities obtained using the λ-core method for $\lambda_{ag} = \lambda_1$ and $\lambda_{ag} = \lambda_2$. The use of flow λ-cores method for MLNS's aggregate-network makes it possible to determine the presence of communities in multilayer system, primarily those that are simultaneously formed in different layers and as a result of process of intersystem interactions. However, this method cannot determine the specific contribution of each layer in creation of such communities and the intensity of interactions between layers as components of different systems, as well as the structural methods of communities detection, which are based on the use of its p_{ag}-core concept. An example of such situation for three-layer network system is shown in Fig. 6. In particular, one of the communities is fully formed in the first layer and practically does not exist in other layers (Fig. 6a, lower left corner). The second of communities is formed in all layers of MLNS (Figs. 6a – 6c, upper right corner), but it stands out in them relatively weakly. However, this community is commensurate with the power of interconnections with the first in aggregate-network of multilayer system in general (Fig. 6d). Such communities can tentatively include the above-mentioned examples of writers and scientists communities. Another disadvantage of flow λ-cores method for MLNS's aggregate-network is the possibility of leveling communities that exist in separate layers (Fig. 7), which does not contribute to a better understanding of processes that take place in multilayer system.

An example of such situation is the sports society already mentioned in section 3, in which separate communities by kinds of sports practically do not overlap. In this case, the independent communities that exist in 1-3 layers of three-layer MLNS (Figs. 7a – 7c) practically "merge" into its aggregate-network, forming a single community (Fig. 7d), which in fact does not correspond to reality.



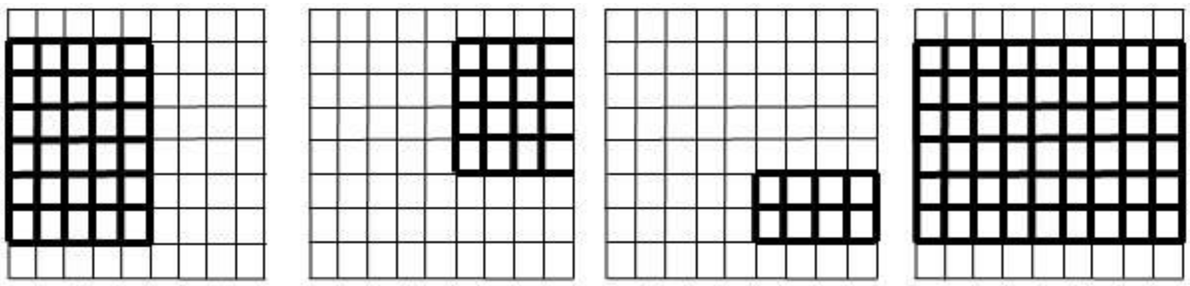
a) 1[st] layer b) 2[nd] layer c) 3[rd] layer d) aggregate-network of MLNS e) communities detection for $\lambda = \lambda_1$ f) communities detection for $\lambda = \lambda_2$

Figure 5: Use of flow λ-core method for communities detection in MLNS's aggregate-network



a) 1[st] layer b) 2[nd] layer c) 3[rd] layer d) overlapping communities in aggregate-network

Figure 6: Communities in separate layers of MLNS and its flow aggregate-network



a) 1[st] layer b) 2[nd] layer c) 3[rd] layer d) leveling communities in aggregate-network

Figure 7: The leveling of communities that exist in separate layers into aggregate-network of multilayer network system

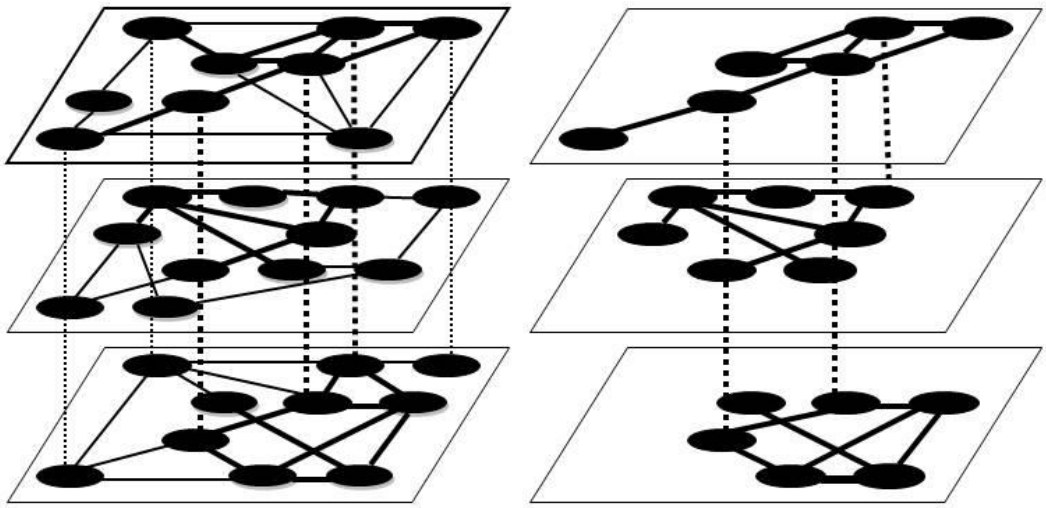
Note that this method cannot determine the specific contribution of each layer in the "fusion" of such communities, as well as structural methods based on the use of its p_{ag}-core concept. In addition, the flow λ_{ag}-cores of MLNS's aggregate-network do not make it possible to establish intersystem interactions between communities that exist in different layers of multilayer system. Therefore, the development of communities detection methods in MLNS, in particular, the selection of such formations in separate layers and the establishment of interactions between them is no less important. Obviously, such communities usually also have the appearance of multilayer system.

6. Communities detection in multilayer network systems

Determine the concept of flow λ-cores of multilayer network system. Let's form an adjacency matrix $V_{\lambda}^{M}(t)=\{V_{\lambda,ij}^{mk}(t)\}_{i,j=1}^{N^N}{}_{m,k=1}^{M}$ in which

$$V_{\lambda,ij}^{mk}(t)=\begin{cases}V_{ij}^{mk}(t), \text{ if } V_{ij}^{mk}(t)\ge\lambda,\\ 0, \qquad \text{ if } V_{ij}^{mk}(t)<\lambda,\ i,j=\overline{1,N^M},\ m,k=\overline{1,M},\ t\ge T,\ \lambda\in[0,1].\end{cases}$$

Similarly to section 4.2, an algorithm for communities detection in MLNS is built, in which communities in separate layers of multilayer system and the connections between them are identified sequentially as the value λ increases. In Fig. 8a (the thickness of lines is proportional to the volume of flows which pass through the MLNS's edges) shows a fragment of three-layer network system, and in Fig. 8b – its flow λ_1-core selected with the help of this algorithm, the flow connections between nodes of which according to criterion (5) are at least three times stronger than in MLNS average. It is obvious that the aggregate-network of flow λ-core of multilayer system is part of its $\lambda_{ag}=\lambda$-core, that is, the communities that are distinguished by method of λ-cores for MLNS are usually subcommunities of communities obtained by the method of λ-cores for the aggregate-network of MLNS.



a) fragment of MLNS b) flow λ_1-core of MLNS

Figure 8: Detection of communities in MLNS in general by the λ-core method

By projecting the λ-core of MLNS, obtained using the method described above, onto its aggregate-network, we can determine the merging, overlapping, or leveling of communities that exist in its separate layers.

7. Conclusions

The study of phenomena of the communities occurrence and development contributes to a better understanding of operation processes of real complex network systems and intersystem interactions that exist in the physical world, living nature, and human society. That is why a lot of scientific research has been devoted to this issue in recent decades. The structural approach to communities detection in complex network and multilayer network systems, which is currently being developed within the framework of theory of complex networks, has a number of shortcomings, among which the first should be called the lack of well-founded criterion that the connections within detected formation, which is considered as community, are not only denser, but also stronger than the network average. In contrast to structural methods, the flow approach makes it possible to effectively solve this problem, because the statement that the greater the volume of flows connecting two network nodes, the stronger the connection between them, seems to be well-founded. The dynamism of formation and development of communities in modern human society, at least some of which pose an obvious or hidden threat to public peace and security, makes the problem of timely detection of such formations even more urgent. The communities detection methods proposed in the article, which are based on the use of network flow core and MLNS's aggregate-network and the flow cores of multilayer network system concepts, make this problem solvable even in real time. An additional advantage of proposed methods is the possibility of their application in those cases when the structure of network or multilayer network system makes other known approaches practically unworkable.

8. References

[1] G. Bianconi, Multilayer Networks: Structure and Function, Oxford University Press, 2018. doi: 10.1093/oso/9780198753919.001.0001.

[2] V. A. Traag, L. Waltman, N. J. van Eck, From Louvain to Leiden: guaranteeing well-connected communities, Scientific Reports 9 (2019) 5233. doi: 10.1038/s41598-019-41695-z.

[3] C. Wagg et al, Fungal-bacterial diversity and microbiome complexity predict ecosystem functioning, Nature Communications 10 (2019) 4841. doi: 10.1038/s41467-019-12798-y.

[4] G. Rossetti, R. Cazabet, Community Discovery in Dynamic Networks: A Survey, ACM Computing Surveys, 5:2 (2018) 1–37. doi: 10.1145/3172867.

[5] B. S. Khan, M. A. Niazi, Network community detection: A Review and Visual Survey, arXiv: 1708.00977 [cs.SI], 2017. doi: 10.48550/arXiv.1708.00977.

[6] S. Tabassum et al, Social network analysis: An overview, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8:5 (2018), e1256. doi: 10.1002/widm.1256.

[7] M. A. Javed et al. Community detection in networks: A multidisciplinary review, Journal of Network and Computer Applications 108 (2018) 87-111. doi: 10.1016/j.jnca.2018.02.011.

[8] O. D. Polishchuk, Flow approaches to community detection in complex network systems, in: Proc. of the International Scientific and Practical Conference on Information Technologies and Computer Modelling, 2021, Ivano-Frankivsk, Ukraine, 81-84.

[9] W. Liu et al, Finding overlapping communities in multilayer networks, PLOS ONE 13:4 (2018) e0188747. doi: 10.1371/journal.pone.0188747.

[10] C. De Bacco et al, Community detection, link prediction, and layer interdependence in multilayer networks, Physical Review E 95 (2017) 042317. doi: 10.1103/PhysRevE.95.042317.

[11] X. Huang et al, A survey of community detection methods in multilayer networks, Data Mining and Knowledge Discovery 35 (2021) 1–45. doi: 10.1007/s10618-020-00716-6.

[12] C. D. G. Linhares et al, Visual analysis for evaluation of community detection algorithms, Multimedia Tools Applications 79 (2020) 17645–17667. doi: 10.1007/s11042-020-08700-4.

[13] D. Jin et al, A Survey of Community Detection Approaches: From Statistical Modeling to Deep Learning, IEEE Transactions on Knowledge and Data Engineering 35:2 (2023) 1149-1170. doi: 10.1109/TKDE.2021.3104155.

[14] J. Li et al, Community-diversified influence maximization in social networks, Information Systems 92 (2020) 101522. doi: 10.1016/j.is.2020.101522.

[15] M. Berlingerio et al, Multidimensional networks: foundations of structural analysis, World Wide Web 16 (2013) 567–593. doi: 10.1007/s11280-012-0190-4.

[16] O. D. Polishchuk, M. S. Yadzhak, Network structures and systems: I. Flow characteristics of complex networks, System research and informational technologies 2 (2018) 42-54. doi: 10.20535/SRIT.2308-8893.2018.2.05.

[17] A.-L. Barabasi, The architecture of complexity, IEEE Control Systems Magazine 27:4 (2007) 33-42.

[18] K. R. Tuttle, Impact of the COVID-19 pandemic on clinical research, Nature Review Nephrology 16 (2020) 562–564. doi: 10.1038/s41581-020-00336-9.

[19] O. D. Polishchuk, M. S. Yadzhak, Network structures and systems: III. Hierarchies and networks. System research and informational technologies 4 (2018) 82-95. doi: 10.20535/SRIT.2308-8893.2018.4.07.

[20] O. D. Polishchuk, M. S. Yadzhak, Network structures and systems: II. Cores of networks and multiplexes, System research and informational technologies 3 (2018) 38-51. doi: 10.20535/SRIT.2308-8893.2018.3.04.