Rough Net Approach for Community Detection Analysis in Complex Networks

Ivett Fuentes[1,2(✉)], Arian Pina[3], Gonzalo Nápoles[2], Leticia Arco[4],
and Koen Vanhoof[2]

[1] Computer Science Department, Central University of Las Villas, Santa Clara, Cuba
ivett@uclv.cu
[2] Faculty of Business Economics, Hasselt University, Hasselt, Belgium
[3] Faculty of Technical Sciences, University of Sancti Spíritus, Sancti Spíritus, Cuba
[4] AI Lab, Computer Science Department, Vrije Universiteit Brussel,
Brussels, Belgium

Abstract. Rough set theory has many interesting applications in circumstances characterized by vagueness. In this paper, the applications of rough set theory in community detection analysis are discussed based on the Rough Net definition. We will focus the application of Rough Net on community detection validity in both *monoplex* and *multiplex* networks. Also, the topological evolution estimation between adjacent layers in *dynamic* networks is discussed and a new community interaction visualization approach combining both complex network representation and Rough Net definition is adopted to interpret the community structure. We provide some examples that illustrate how the Rough Net definition can be used to analyze the properties of the community structure in real-world networks, including *dynamic* networks.

Keywords: Extended rough set theory · Community detection analysis · Monoplex complex networks · Multiplex complex networks

1 Introduction

Complex networks have proved to be a useful tool to model a variety of complex systems in different domains including sociology, biology, ethology and computer science. Most studies until recently have focused on analyzing simple static networks, named *monoplex* networks [7,17,18]. However, most of real-world complex networks are dynamics. For that reason, *multiplex* networks have been recently proposed as a mean to capture this high level complexity in real-world complex systems over time [19]. In both *monoplex* and *multiplex* networks the key feature of the analysis is the community structure detection [11,19].

Community detection (CD) analysis consists of identifying dense subgraphs whose nodes are densely connected within itself, but sparsely connected with the rest of the network [9]. CD in *monoplex* networks is a very similar task to classical clustering, with one main difference though. When considering complex

networks, the objects of interest are nodes, and the information used to perform the partition is the network topology. In other words, instead of considering some individual information (attributes) like for clustering analysis, CD algorithms take advantage of the relational one (links). However, the result is the same in both: a partition of objects (nodes), which is called community structure [9].

Several CD methods have been proposed for *monoplex* networks [7,8,12,16–18]. Also, different approaches have been recently emerged to cope with this problem in the context of *multiplex* networks [10,11] with the purpose of obtaining a unique community structure involving all interactions throughout the layers. We can classify latter existing approaches into two broad classes: (I) by transforming into a problem of CD in simple networks [6,9] or (II) by extending existing algorithms to deal directly with *multiplex* networks [3,10]. However, the high-level complexity in real-world networks in terms of the number of nodes, links and layers, and the unknown reference of classification in real domain convert the evaluation of CD in a very difficult task. To solve this problem, several quality measures (internal and external) have emerged [2,13]. Due to the performance may be judged differently depending on which measure is used, several measures should be used to be more confident in results. Although, the modularity is the most widely used, it suffers the resolution limit problem [9]. Another goal of the CD analysis is the understanding of the structure evolution in *dynamic* networks, which is a special type of *multiplex* that requires not only discovering the structure but also offering interpretability about the structure changes.

Rough Set Theory (RST), introduced by Pawlak [15], has often proved to be an excellent tool for analyzing the quality of information, which means inconsistency or ambiguity that follows from information granulation in a knowledge system [14]. To apply the advantages of RST in some fields of CD analysis, the goal of our research is to define the new Rough Net concept. Rough Net is defined starting from a community structure discovered by CD algorithms applied to *monoplex* or *multiplex* networks. This concept allows us obtaining the upper and lower approximations of each community, as well as, their accuracy and quality. In this paper, we will focus the application of the Rough Net concept on CD validity and topological evolution estimation in *dynamic* networks. Also, this concept supports visualizing the interactions of the detected communities.

This paper is organized as follows. Section 2 presents the general concepts about the extended RST and its measures for evaluating decision systems. We propose the definition of Rough Net in Sect. 3. Section 4 explains the applications of Rough Net in the community detection analysis in complex networks. Besides, a new approach for visualizing the interactions between communities based on Rough Net is provided in Sect. 4. In Sect. 5, we illustrate how the Rough Net definition can be used to analyze the properties of the community structure in real-world networks, including *dynamic* networks. Finally, Sect. 6 concludes the paper and discusses future research.

2   Extended Rough Set Theory

The rough sets philosophy is based on the assumption that with every object of the universe U there is associated a certain amount of knowledge expressed through some attributes A used for object description. Objects having the same description are indiscernible with respect to the available information. The indiscernibility relation R induces a partition of the universe into blocks of indiscernible objects resulting in information granulation, that can be used to build knowledge. The extended RST considers that objects which are not indiscernible but similar can be grouped in the same class [14]. The aim is to construct a similarity relation R' from the relation R by relaxing the original indiscernibility conditions. This relaxation can be performed in many ways, thus giving many possible definitions for similarity. Due to that R' is not imposed to be symmetric and transitive, an object may belong to different similarity classes simultaneously. It means that R' induces a covering on U instead of a partition. However, any similarity relation is reflexive. The rough approximation of a set $X \subseteq U$, using the similarity relation R', has been introduced as a pair of sets called R'-lower (R'_*) and R'-upper (R'^*) approximations of X. A general definition of these approximations which can handle any reflexive R' are defined respectively by Eqs. (1) and (2).

$$R'_*(X) = \{x \in X : R'(x) \subseteq X\} \tag{1}$$

$$R'^*(X) = \bigcup_{x \in X} R'(x) \tag{2}$$

$$\alpha(X) = \frac{|R'_*(X)|}{|R'^*(X)|} \tag{3}$$

The extended RST offers some measures to analyze decision systems, such as the accuracy and quality of approximation and quality of classification measures. The accuracy of approximation of a rough set X, where $|X|$ denotes the cardinality of $X \neq \varnothing$, offers a numerical characterization of X. Equation (3) formalizes this measure such that $0 \leq \alpha(X) \leq 1$. If $\alpha(X) = 1$, X is crisp (exact) with respect to the set of attributes, if $\alpha(X) < 1$, X is rough (vague) with respect to the set of attributes. The quality of approximation formalized in Eq. (4) expresses the percentage of objects which can be correctly classified into the class X. Note that $0 \leq \alpha(X) \leq \gamma(X) \leq 1$, and $\gamma(X) \leq 0$ if $\alpha(X) \leq 0$, while $\gamma(X) \leq 1$ if $\alpha(X) \leq 1$ [14]. Quality of classification expresses the proportion of objects which can be correctly classified in the system; Equation (5) formalizes this coefficient where $C_1, \cdots, C_m$ correspond to the decision classes of the decision system DS. Notice that if the quality of classification value is equal to 1, then DS is consistent, otherwise is inconsistent [14]. Equation (6) shows the accuracy of classification, which measures the average the accuracy per classes with different importance levels. Its weighted version is formalized in Eq (7) [4].

$$\gamma(X) = \frac{|R'_*(X)|}{|X|} \tag{4}$$

$$\gamma(DS) = \frac{\sum_{i=1}^{m} |R'_*(C_i)|}{|U|} \tag{5}$$

$$\alpha(DS) = \frac{\sum_{i=1}^{m} \alpha(C_i)}{m} \tag{6}$$

$$\alpha_w(DS) = \frac{\sum_{i=1}^{m} (\alpha(C_i) \cdot |C_i|)}{\sum_{i=1}^{m} |C_i|} \tag{7}$$

3   Rough Net Definition

Monoplex (simple) networks can be represented as graphs $G = (V, E)$ where V represents the vertices (nodes) and E represents the edges (interactions) between these nodes in the network. *Multiplex* networks have multiple layers, where each one is a *monoplex* network. Formally, a *multiplex* network can be defined as a triplet $< V, E, L >$ where $E = \bigcup E_i$ such that E_i corresponds to the interactions on layer i-th and L is the number of layers. This extension of graph model is powerful enough though to allow modeling different types of networks including *dynamic* and *attributed* networks [9]. CD algorithms exploit the topological structure for discovering a collection of dense subgraphs (communities). Several *multiplex* CD approaches emphasize on how to obtain a unique community structure throughout all layers, by considering as similar nodes that ones with the same behavior in most of the layers [3,10]. In the context of *dynamic* networks, the goal is to detect the conformation by layers for characterizing the evolutionary or stationary properties of the CD structures. Due to the quality of the community structure may be judged differently depending on which measure is used, to be more confident in results several measures should be used [9]. In this section, we recall some basic notions related to the definition of the extension of RST in complex networks. Also, we will focus on the introduction of the Rough Net concept by extrapolating these notions to the analysis of the consistency of the detected communities in complex networks. This concept supports to validate, visualize, interpret and understand the communities and also their evolution. Besides, it has a potential application in labeling and refining the detected communities. As was mentioned, it is necessary to start from the definition of the decision system, the similarity relation, and the basic concepts of lower and upper approximations.

Definition 1. *(Rough Net): Given a complex network G, where V represents the nodes. Let $s : V \times V \to R'$ be a function that measures the similarity between nodes of V. The Rough Net comprises the combination of the topological structure and the CD results as a decision system $DS = (V, A \cup d)$, where A is a finite set of topological (i.e., the adjacency tensor of G) or non-topological features which may additionally be available if the network is attributed and $d \notin A$ is the decision attribute resulting from the detected communities.*

We use a similarity relation R' in our definition of Rough Net, because two nodes of V can be similar but not equal. The similarity class of the node x is denoted

by $R'(x)$, as shown in Eq. (8). The R'-lower and R'-upper approximations for each similarity class are computed by Eqs. (1) and (2) respectively. There is a variety of distances and similarities for comparing nodes [1], such as Salton, Hub Depressed Index (HDI), Hub Promoted Index (HPI), similarities based on the topological structure, and Dice and Cosine coefficients which capture the attribute relations. In this paper, we use the Jaccard similarity for computing the similarities based on the topological structure because it has the attraction of simplicity and normalization. The Jaccard similarity, which also allows us to emphasize the network topology necessary to apply RST in complex networks, is defined in Eq. (9), where $\Gamma(X)$ denotes the neighborhood of the node x including it.

$$R'(x) = \{y \in V : yR'x, \text{ iff } s(x,y) \geq \xi\} \tag{8}$$

$$s(x,y) = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|} \tag{9}$$

3.1 Decision System for Applying RST on *monoplex* Networks

An adjacency tensor for a *monoplex* (i.e., single layer) network can be reduced to an adjacency matrix. The topological relation between nodes comprises an $|V| \times |V|$ adjacency matrix M, in which each entry $M_{i,j}$ indicates the relationships between nodes i and j weighted or not. The weight can be obtained as a result of the application of both a flattening process in a multi-relational network or a network construction schema when we want to apply network-based learning methods to vector-based datasets. If we apply some CD algorithm to this adjacency matrix, then we can consider the combination of the topological structure and the CD results as a decision system $DS_{monoplex} = (V, A \cup d)$, where A is a finite set of topological or non-topological features and $d \notin A$ is the decision attribute resulting from the detected communities over the network.

3.2 Decision System for Applying RST on *multiplex* Networks

Multiplex are powerful enough though to allow modeling different types of networks including *multi-relational*, *attributed* and *dynamic* networks [11]. Note that *multiplex* networks explicitly incorporate multiple channels of connectivity in which entities can have a different set of neighbors in each layer. In a *dynamic* network each layer corresponds to the network state at a given time-stamp (or each layer represents a snapshot). Like a time-series analysis, if attributes are captured in each time, a complex network can be represented as a *dynamic* network [19].

An adjacency tensor for a *dynamic* network with dimension L, which corresponds to the number of layers, represents a collection of adjacency matrices. The topological interaction between nodes within each layer k-th of a *multiplex* network comprises an $|V| \times |V|$ adjacency matrix M_k, in which each entry M_{ij}^k indicates the relationships between nodes i and j in the k-th layer. If we apply a CD algorithm to the whole *multiplex* network topology by considering

multiplex CD approaches [10,19] in order to compute the unique final community structure, then we can consider the application of RST concepts over the *multiplex* network as the aggregation of the application of the RST concepts over each layer k-th. Consequently, the decision system for the k-th layer is the combination of the topological structure M_k and the CD results, formalized as $DS_{layer_k} = (V, A_k \cup d)$, where A_k is a finite set of topological or non-topological features in the k-th layer and $d \notin A$ is the decision attribute resulting from the detected communities in the *multiplex* topology (i.e., each node and their counterpart in each layer represent a unique node that belongs to a specific community). Besides, it is possible to transform a *multiplex* into a *monoplex* network by a flattening process. The main flatten approaches are the binary flatten, the weighted flatten and another based on deep learning [10]. Taking into account these variants, we can consider the combination of the topological structure of the transformed network and the CD results as a decision system $DS_{monoplex} = (V, A \cup d)$, where $A = \bigcup_{k \in L} A_k$ is a finite set of topological or non-topological features that characterize the networks and $d \notin A$ is the decision attribute resulting from the detected communities. The multiple instance or ensemble similarity measures are powerful for computing the similarity between nodes taking into account the similarity per layers (contexts).

4   The Application of Rough Net in the Community Detection Analysis

In this section, we describe the application of Rough Net in important tasks of the CD analysis: the validation and visualization of detected communities and their interactions, and the evolutionary estimation in *dynamic* networks.

4.1   Community Detection Validity

A community can be defined as a subgraph whose nodes are densely connected within itself, but sparsely connected with the rest of the network, though other patterns are possible. The existence of communities implies that nodes interact more strongly with the other members of their community than they do with nodes of the other communities. Consequently, there is a preferential linking pattern between nodes of the same community (being modularity [13] one of the most used internal measures [9]). This is the reason why link densities end up being higher within communities than between them. Although the modularity is the most widely quality measure used in complex networks, it suffers the resolution limit problem [9] and, therefore, it is unable to judge in a correct way community structure of the networks with small communities or where communities may be very heterogeneous in size, especially if the network is large. Several methods and measures have been proposed to detect and evaluate communities in both *monoplex* and *multiplex* networks [2,3,13]. As well as modularity, Normalized Mutual Information (NMI), Adjusted Rand (AR), Rand, Variation of Information (VI) measures [2] are widely used, but the latter ones need an

---

Algorithm 1. Community Detection Validity

Input: A *Monoplex* or *multiplex* network G (attributed or not), detected communities, a threshold ξ and a similarity s (topological or non-topological features)

Output: Values of quality, accuracy and weighted accuracy of classification measures

---

1: **if** G is a *monoplex* network **then**
2:     $DS[1] \leftarrow DS_{monoplex}$ (see Section 3.1)
3:     $C[1] \leftarrow communities(G, d)$
4: **else if** G is a *multiplex* network **then**
5:     **for** k in L **do**
6:         $DS[k] \leftarrow DS_{layer_k}$ (see Section 3.2)
7:         $C[k] \leftarrow communities(layer(G, k), d)$
8:     **end for**
9: **end if**
10: **for** k in $(1 : size(DS))$ **do**
11:     Obtain the similarity class $R'_k(x)$ based on Equation (8)
12:     **for** X in $C[k]$ **do**
13:         Calculate $R'_{k*}(X)$ and $R'^{*}_k(X)$ approximations (see equations (1)–(2))
14:         Calculate $\alpha(X)$ and $\gamma(X)$ approximation measures (see equations (3)–(4))
15:     **end for**
16:     Calculate $\gamma(DS_k)$, $\alpha(DS_k)$ and $\alpha_w(DS_k)$ in DS_k (see equations (4)–(7))
17:     $\gamma_G(DS)+ = \gamma(DS_k)$, $\alpha_G(DS)+ = \alpha(DS_k)$ and $\alpha_{w_G}(DS)+ = \alpha_w(DS_k)$
18: **end for**
19: $\gamma_G(DS) = \gamma_G(DS)/L$, $\alpha_G(DS) = \alpha_G(DS)/L$ and $\alpha_{w_G}(DS) = \alpha_{w_G}(DS)/L$

---

external reference classification to produce a result. However, it is very difficult to evaluate a community result because the major of complex networks occur in real world situations since reference classifications are usually not available. We propose to use quality, accuracy and weighted accuracy of classification measures described in Sect. 2 to validate community results, taking into account the application of accuracy and quality of approximation measures to validate each community structure. Aiming at providing more insights about the validation, we provide a general procedure based on Rough Net. Notice that $R'_k(x)$ is computed by considering the attributes or topological features of networks in the k-th layer, by using Eq. (8). Algorithm 1 allows us to measure the quality of the community structure using Rough Net, by considering the quality and precision of each community. Rough Net allows judging the quality of the CD by measuring the vagueness of each community. For that reason, if boundary regions are smaller, then we will obtain better results of quality, accuracy and weighted accuracy of classification measures.

4.2   The Evolutionary Estimation in Dynamic Networks

A huge of real-world complex networks are dynamic in nature and change over time. The change can be usually observed in the birth or death of interactions within the network over time. In a *dynamic* network is expected that nodes of the same community have a higher probability to form links with their partners than

---

Algorithm 2. Evolution Analysis

Input: Two-consecutive layers CL of G, a threshold ξ and a similarity s
Output: The evolutionary estimations
1: $d_{k-1} \leftarrow CD(layer(G, k-1))$
2: $d_k \leftarrow CD(layer(G, k)), d_{k-0} \leftarrow d_k$
3: $DS_{k-1} \leftarrow (V, A_{k-1} \cup d_k)$ (see Section 3.2)
4: $DS_k \leftarrow (V, A_k \cup d_{k-1}), DS_{k-0} \leftarrow DS_k$ (see Section 3.2)
5: **for** i in $(0 : 1)$ **do**
6:     **for** X in d_{k-i} **do**
7:         Obtain the similarity class $R'_{k-i}(x)$ based on Equation (8)
8:         Calculate $R'_{k-i} * (X)$ and $R'^*_{k-i}(X)$ approximations (see equations (1)–(2))
9:         Calculate $\alpha_{k-i}(X)$ and $\gamma_{k-i}(X)$ measures (see equations (3)–(4))
10:     **end for**
11:     $\gamma_i = \gamma(DS_{k-i}), \alpha_i = \alpha(DS_{k-i})$ and $\alpha_{wi} = \alpha_w(DS_{k-i})$ in DS_{k-i}
12: **end for**
13: $\gamma_{CL} = (\gamma_0 + \gamma_1)/2, \alpha_{TC} = (\alpha_0 + \alpha_1)/2$ and $\alpha_{w_{CL}} = (\alpha_{w0} + \alpha_{w1})/2$

---

with other nodes [19]. For that reason, the key feature of the community detection analysis in *dynamic* networks is the evolution of communities over time. Several methods have been proposed to detect these communities over time for specific time-stamp windows [3, 10]. Often more than one community structure is required to judge if the network topology has suffered transformation over time for specific window size. To the best of our knowledge, there is no measure able which captures this aspect. For that reason, in this paper, we propose measures based on the average of quality, accuracy and weighted accuracy of classification for estimating in a real number the change level during a specific window time-stamp. We need to consider two-consecutive layers for computing the quality, accuracy and weighted accuracy of classification measures in the evolutionary estimation (see Algorithm 2). For that reason, we need to apply twice the Rough Net concept for each pair of layers. The former Rough Net application is based on the decision system $DS = (V, A_k \cup d_{k-1})$, where A_k is a set of topological attributes in the layer k and $d_{k-1} \notin A_k$ is the result of the community detection algorithm in the layer $k-1$ (decision attribute). The latter Rough Net application is based on the decision system $DS = (V, A_{k-1} \cup d_k)$, where A_{k-1} is a set of topological attributes in the layer $k-1$ and $d_k \notin A_{k-1}$ is the result of the community detection algorithm in the layer k (decision attribute). The measures can be applied over a window size K by considering the aggregation of the quality classification between all pairs of consecutive (adjacency) layers. Values nearer to 0 express the topology is evolving over time.

4.3   Discovering Interactions Between Communities

In many applications more than a unique real value that expresses the quality of the community conformation is required for the understanding of the interactions throughout the networks. Besides, real-world complex networks usually are

---

Algorithm 3. Visualization for Community Structure Analysis

Input: A complex network G, detected communities, a threshold ξ and a similarity s
Output: Community network representation

1: Create an empty network $G'(V', E')$
2: **for** x in V **do**
3:     Obtain the similarity class $R'(x)$ based on Equation (8)
4: **end for**
5: **for** X in $communities(G, d)$ **do**
6:     Calculate $R'_*(X)$ and $R'^*(X)$ approximations (see equations (1)–(2))
7:     Calculate $\alpha(X)$ and $\gamma(X)$ approximation measures (see equations (3)–(4))
8:     Add a new node X where the size corresponds to quality or accuracy
9: **end for**
10: **for** X, Y in $communities(G), X \neq Y$ **do**
11:     Calculate the similarity s_{BN} between communities X-th and Y-th
12:     Add a new edge (I, Y, w_{XY}) where the weighted $w_{ij} = s_{BN}(X, Y)$
13: **end for**

---

composed by many nodes, edges, and communities, making difficult to interpret the obtained results. Thus, we propose a new approach for visualizing the interactions between communities taking into account the quality of the community structure by using the combination of the Rough Net definition and the complex network representation. Our proposal, formalized in Algorithm 3, allows us to represent the quality of the community structure in an interpretable way.

The similarity measure used for weighted the interactions between communities in the network representation is formalized in Eq. (10). The $s_{BN}(X, Y)$ captures the proportion of nodes members of the community X, which cannot be unambiguously classified into this community but belong to the community Y and vice-versa. The above idea is computed based on the boundary region BN of both communities X and Y. The Rough Net approach allows us to evaluate the interaction between the communities and its visualization facilitates interpretability. In turn, it helps experts redistribute communities and change granularity based on the application domain requirements.

$$s_{BN}(X, Y) = \frac{\frac{|BN(X) \cap Y|}{|BN(X)|} + \frac{|BN(Y) \cap X|}{|BN(Y)|}}{2} \tag{10}$$

5   Illustrative Examples

For illustrating the performance of the Rough Net definition in the community detection analysis, we apply it to three networks, two known to have *monoplex* topology and the third *multiplex* one. To be more confident in results, we should use several measures for judging the performance of a CD algorithm [2,5].Thus, we compare our approach to validate detected communities (i.e., accuracy and quality of classification) with the most popular internal and external measures used for community detection validity: modularity, AR, NMI, Rand, VI [2]. Modularity [13] quantifies when the division is a good one, in the sense of having many

within-community edges. It takes its largest value (1) in the trivial case where all nodes belong to a single community. A value near to 1 indicates strong community structure in the network. All other mentioned measures need external references for operating. All measures except VI, express the best result though values near to 1. For that reason, we use the notation VIC for denoting the complement of VI measure (i.e., $VIC = 1 - VI$).

5.1   Zachary Network

Zachary is the much-discussed network[1] of friendships between 34 members of a karate-club at a US university. Figure 1 shows the community structures reported by the application of the standard CD algorithms Label Propagation (LP), Multilevel Louvain (LV), Fast Greedy Optimization (FGO), Leading Eigenvector (EV), Infomap (IM) and Walktrap (WT) to the Zachary network. Each community has been identified with a different colour. These algorithms detect communities, which mostly not correspond perfectly to the reference communities, except the LP algorithm which identically matches. For that reason, we can affirm that the LP algorithm reported the best division. However, in Fig. 2 we can observe that the modularity values not distinguish the LP as the best conformation of nodes into communities, while the proposed accuracy and quality
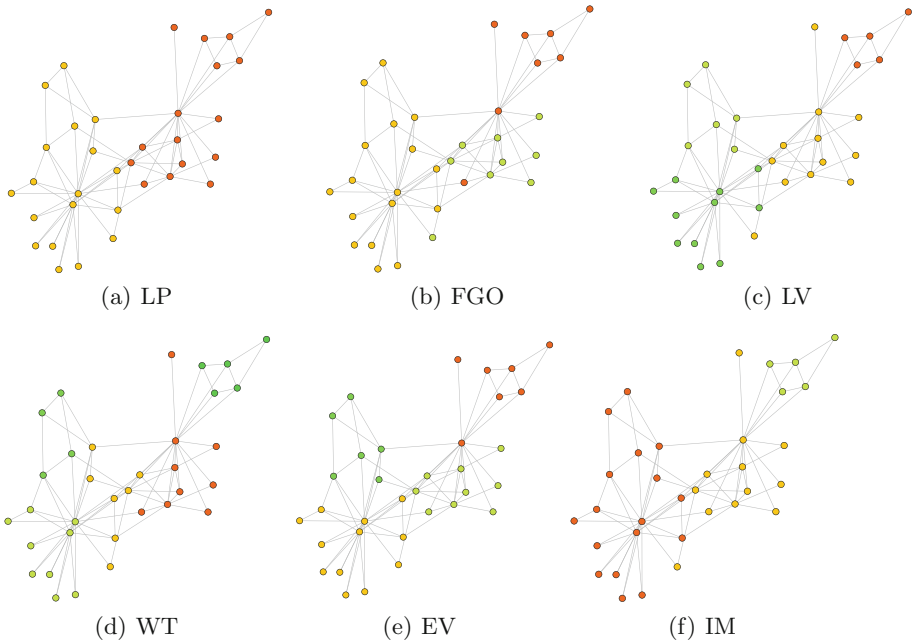

(a) LP   (b) FGO   (c) LV

(d) WT   (e) EV   (f) IM

Fig. 1. Communities detected by different algorithms in the Zachary network.

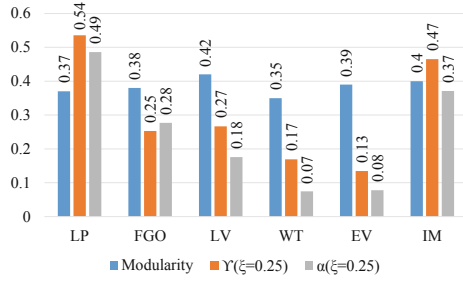[1] http://networkrepository.com/ucidata-zachary.php.

Fig. 2. Performance of the internal measures on the Zachary CD evaluation.
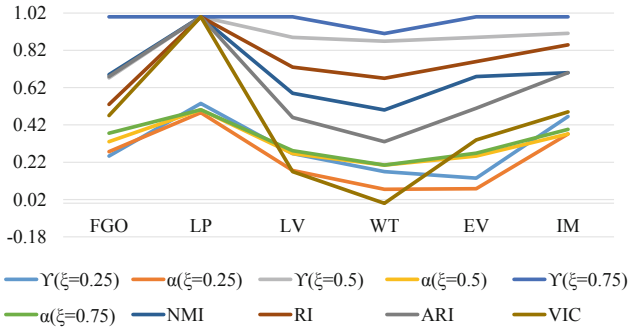

Fig. 3. Performance of the proposed measures on the Zachary CD evaluation.

of classification measures based on the Rough Net definition, assign the higher value to the LP conformation regardless of the used threshold. On the other hand, our measures grant the lowest quality results for the community structure obtained by the EV algorithm as expected. Notice that FGO and EV assign the orange node with high centrality in the orange community structure in a wrong manner. We can notice that most neighbors of this node are in another community. Indeed, the FGO and WT are the following lowest results reported by our measures. Figure 3 shows the performance reported by the application of the standard community detection algorithms before mentioned by using the proposed quality measures and the external ones. All measures exhibit the same monotony behaviors with independence of the selected similarity threshold ξ. Our measures have the advantage that are internal and behave similarly to external measures.

5.2   Jazz Network

The Jazz network[2] represents the collaboration between jazz musicians, where each node represents a jazz musician and interactions denote that two musicians

---

[2] http://konect.cc/networks/arenas-jazz/.

are playing together in a band. Six CD algorithms were applied to this network with the objective of subsequently exploring the behavior of validity measures. Figure 4 displays that LP obtains a partition in which the number of interactions shared between nodes of different communities is smaller than the number of interactions shared between the communities obtained by the FGO algorithm. However, this behavior is not reflected in the estimation of the modularity values, while it manages to be captured by the proposed quality measures, as shown in Fig. 5. Besides, the number of interactions shared between the communities detected by the algorithms LV, FGO, and EV is much greater than the number of interactions shared between the communities detected by the algorithms LP, WT, and IM. Therefore, this behavior was expected to be captured through the Rough Net definition. Figure 5 shows that the results reported by our measures coincide with the expected results. On the one hand, we can observe that our quality measures exhibit a better performance than the modularity measure in this example. Our measures also capture the presence of outliers, this is the reason why the community structure reported by the WT algorithm is higher than the obtained by the LP algorithm.
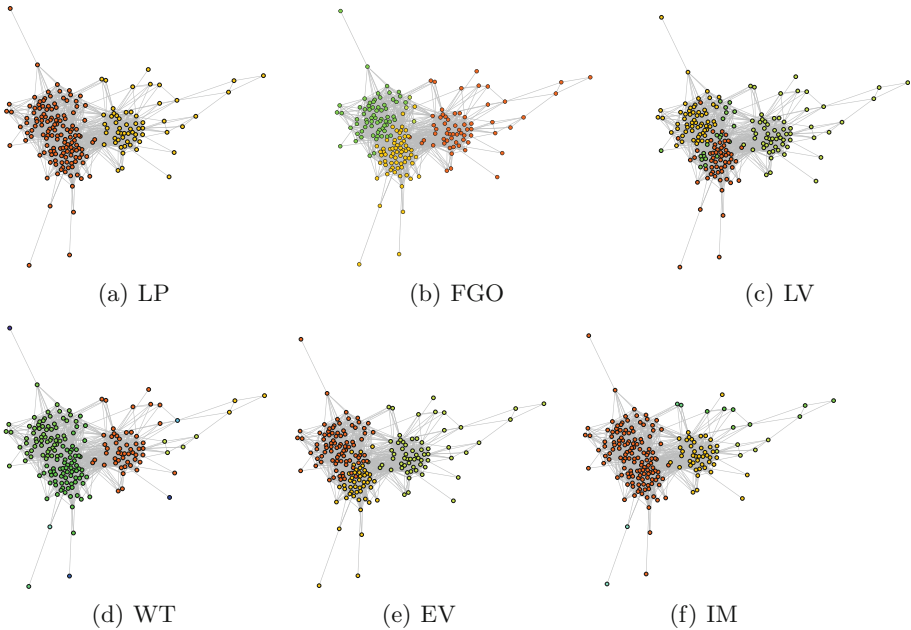

(a) LP   (b) FGO   (c) LV

(d) WT   (e) EV   (f) IM

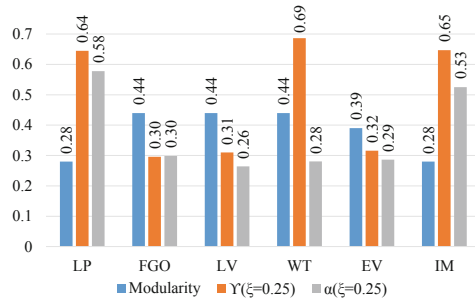**Fig. 4.** Communities detected by different algorithms in the Jazz network.

Fig. 5. Performance of the internal measures on the Jazz CD evaluation.

5.3 CElegans Network

Caenorhabditis elegans connectome (CElegans) is a *multiplex* network[3] that consists of layers corresponding to different synaptic junctions: electric (ElectrJ), chemical monadic (MonoSyn), and polyadic (PolySyn). Figure 6 shows the mapping of the community structure in each network layer, which has been obtained by the application of the MuxLod CD algorithm [10]. Notice that a strong community structure result must correspond to a structure of densely connected subgraphs in each network layer. This reflexion property is not evident for these communities in the CElegans network. For that reason, both the modularity and the proposed quality community detection measures obtain low results ($Modularity = 0.07$, $\alpha(\xi = 0.25) = 0.24$ and $\gamma(\xi = 0.25) = 0.14$). Figure 7 shows the interactions between the communities in each layer by considering the MuxLod community structure and the algorithm described in Sect. 4.3. The community networks show high interconnections and as expected, the results of the quality measures are low. Figure 7 shows that the topologies of the PolySyn and ElectrJ layers do not match exactly. In this sense, let us suppose without loss of generalization, that we want to estimate if there has been a change in
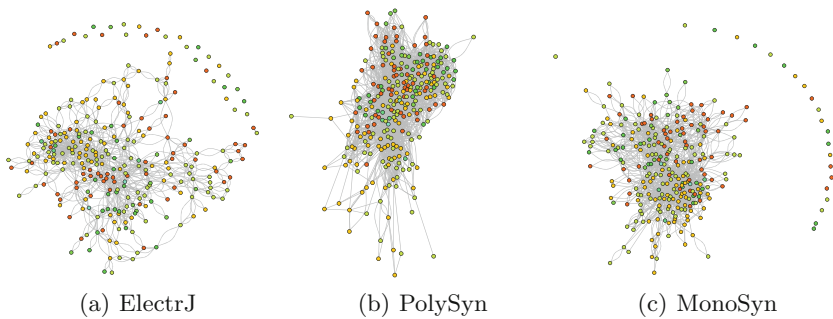

(a) ElectrJ   (b) PolySyn   (c) MonoSyn

Fig. 6. Application of the MuxLod CD to the CElegans network.

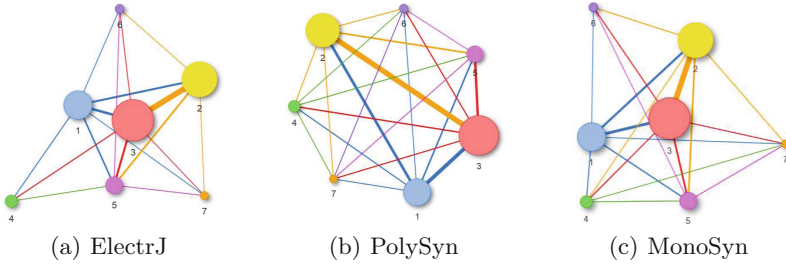(a) ElectrJ   (b) PolySyn   (c) MonoSyn

Fig. 7. Visualization of community quality based on the MuxLod CD to the CElegans network.
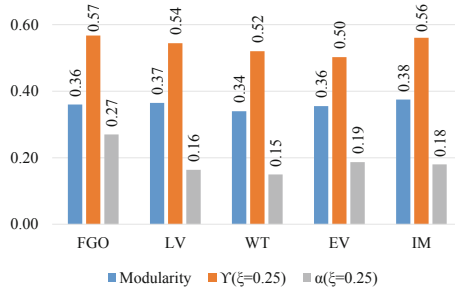

Fig. 8. Performance of the internal measures on the CElegans CD evaluation.

the topology considering these layers as consecutive. To estimate these results, we apply the algorithm described in Sect. 4.2. Figure 8 shows the modularity, accuracy and quality of classification obtained values, which reflect that the community structure between layers does not completely match, so it can be concluded that the topology has evolved (changed).

6 Conclusions and Future Work

In this paper, we have described new quality measures for exploratory analysis of community structure in both *monoplex* and *multiplex* networks based on the Rough Net definition. The applications of Rough Net in community detection analysis demonstrate the potential of the proposed measures for judging the community detection quality. Rough Net allows us to asses the detected communities without requiring the referenced structure. Besides, the proposed evolutionary estimation and the new approach for discovering the interactions between communities allows to the experts a deep understanding of complex real systems mainly based on the visualization of interactions. For the future work, we propose to extend the applications of Rough Net to the estimation of the community structure in the next time-stamp based on the refinement between adjacent layers in *dynamic* networks.

References

1. Ahajjam, S., El Haddad, M., Badir, H.: A new scalable leader-community detection approach for community detection in social networks. Soc. Netw. **54**, 41–49 (2018)
2. Aldecoa, R., Marín, I.: Surprise maximization reveals the community structure of complex networks. Sci. Rep. **3**(1), 1–9 (2013)
3. Amelio, A., Pizzuti, C.: Community detection in multidimensional networks. In: 26th International Conference on Tools with Artificial Intelligence, pp. 352–359. IEEE (2014)
4. Arco, L., Bello, R., Caballero, Y., Falcón, R.: Rough text assiting text mining: focus on document clustering validity. In: Bello, R., Falcón, R., Pedrycz, W., Kacprzyk, J. (eds.) Granular Computing: at the Junction of Rough Sets and Fuzzy Sets. Studies in Fuzziness and Soft Computing, vol. 224, pp. 229–248. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-76973-6_15
5. Bai, L., Liang, J., Du, H., Guo, Y.: A novel community detection algorithm based on simplification of complex networks. Knowl.-Based Syst. **143**, 58–64 (2018)
6. Berlingerio, M., Pinelli, F., Calabrese, F.: ABACUS: frequent pattern mining-based community discovery in multidimensional networks. Data Min. Knowl. Disc. **27**(3), 294–320 (2013)
7. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. J. Stat. Mech: Theory Exp. **2008**(10), P10008 (2008)
8. Clauset, A., Newman, M.E.J., Moore, C.: Finding community structure in very large networks. Phys. Rev. E **70**, 066111 (2004)
9. De Domenico, M., et al.: Mathematical formulation of multilayer networks. Phys. Rev. X **3**(4), 041022 (2013)
10. Falih, I., Kanawati, R.: MUMA: a multiplex network analysis library. In: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 757–760. IEEE (2015)
11. Kanawati, R.: Multiplex network mining: a brief survey. IEEE Intell. Inf. Bull. **16**, 24–28 (2015)
12. Newman, M.E.: Finding community structure in networks using the eigenvectors of matrices. Phys. Rev. E **74**(3), 036104 (2006)
13. Newman, M.E., Leicht, E.A.: Mixture models and exploratory analysis in networks. Proc. Natl. Acad. Sci. **104**(23), 9564–9569 (2007)
14. Orlowska, E.: Incomplete Information: Rough Set Analysis, vol. 13. Physica (2013)
15. Pawlak, Z.: Rough set theory and its applications to data analysis. Cybern. Syst. **29**(7), 661–688 (1998)
16. Pons, P., Latapy, M.: Computing communities in large networks using random walks. In: Yolum, P., Güngör, T., Gürgen, F., Özturan, C. (eds.) ISCIS 2005. LNCS, vol. 3733, pp. 284–293. Springer, Heidelberg (2005). https://doi.org/10.1007/11569596_31
17. Raghavan, U.N., Albert, R., Kumara, S.: Near linear time algorithm to detect community structures in large-scale networks. Phys. Rev. E **76**, 036106 (2007)
18. Rosvall, M., Bergstrom, C.T.: Maps of information flow reveal community structure in complex networks. arXiv preprint physics.soc-ph/0707.0609 (2007)
19. Zou, Y., Donner, R.V., Marwan, N., Donges, J.F., Kurths, J.: Complex network approaches to nonlinear time series analysis. Phys. Rep. **787**, 1–97 (2019)