Rough Net Approach for Community Detection Analysis in Complex Networks

Peer-reviewed author version

FUENTES HERRERA, Ivett; Pina, Arian; NAPOLES RUIZ, Gonzalo; Arco, Leticia & VANHOOF, Koen (2020) Rough Net Approach for Community Detection Analysis in Complex Networks. In: Rough Sets, p. 401 -415.

DOI: 10.1007/978-3-030-52705-1_30

Handle: http://hdl.handle.net/1942/32437

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, Cuba
[2] Faculty of Business Economics, Hasselt University, Belgium
[3] Faculty of Technical Sciences, University of Sancti Spritus, Cuba
[4] AI Lab, Computer Science Department, Vrije Universiteit Brussel, Belgium
ivett@uclv.cu

Abstract. Rough set theory has many interesting applications in circumstances which are characterized by vagueness. In this paper, the applications of rough set theory in community detection analysis is discussed based on the Rough Net definition. We will focus the application of Rough Net concept in community detection validity in both *monoplex* and *multiplex* complex networks. Also, the topological evolution estimation between adjacent layers in dynamic networks is discussed and a new visualization schema combining both complex network representation and Rough Net definition is adopted contributing to the understanding of the community structure. We provide some examples demonstrating 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* complex 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* complex networks the key feature of the analysis is the community structure detection [11,19].

Community detection (CD) analysis like clustering analysis is a part of the unsupervised field and consists in identify dense subgraphs whose nodes are densely connected within itself, but sparsely connected with the rest of the network [9]. CD in *monoplex* complex networks is obviously a very similar task to classical clustering, with one 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 cases: a partition of the set of objects (nodes), which is called community structure [9].

Several methods have been proposed to detect communities in *monoplex* complex 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 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 been proposed [2, 13]. Due to the performance may be judged differently depending on which measure are used, to be more confident in results one should use several measures. Although, the modularity is the most widely used, it suffers the resolution limit problem [9]. Another goal of the community detection analysis is the understanding of the structure evolution in a dynamic network, 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) has often proved to be an excellent tool for the analysis the quality of information, means inconsistency or ambiguity which follows from information granulation [14]. To apply the advantages of RST in some fields of community detection analysis, the goal of our research is to define the new concept Rough Net. This concept is defined starting from a community structure by the application of a CD algorithm to *monoplex* or *multiplex* networks. Rough Net allows us obtaining the upper and the 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 in community detection validity and topological evolution estimation in dynamic networks. Also, this concept supports the visualization of the community detection quality.

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

2 Extended Rough Set Theory

RST, introduced by Z. Pawlak [15], has often proved to be an excellent mathematical tool for the analysis of the quality of information, means inconsistency or ambiguity which follows from information granulation in a knowledge system. 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 extends the classic approach to RST by considering that objects which are not indiscernible but similar can be grouped in the same class [14]. The aim is constructing a similarity relation R' from the relation R by relaxing the original conditions for indiscernibility. 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 the covering induces by R' on U may not be a partition. However, any similarity relation is reflexivity. The rough approximation of a set $X \subseteq U$, using the similarity relation R', has been introduced as a pair of sets called $R' - lower$ and $R' - upper$ approximations of X. A general definition of these approximations which can handle any reflexive R' are defined respectively by equations 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 information 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 Equation 4 expresses the percentage of objects which can be correctly classified into the class X. Moreover, $0 \leq \alpha(X) \leq \gamma(X) \leq 1$, 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 $C_1, \cdots, C_m$ corresponds to the decision classes of the decision system DS. Notice that, if a value is equal to 1, then DS is consistent, otherwise is inconsistent [14]. Equation 6 shows the Accuracy of Classification, which measures the averages the accuracy per classes with different importance levels and a weighted version is formalized by Equation 7 [4].

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

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

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

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

3 Rough Net Definition

$Monoplex$ and $Multiplex$ networks are special cases of complex networks, in which the key feature of the topological analysis is the community detection. $Monoplex$ (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 with the same behavior in most of 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 community structure may be judged differently depending on which measure are used, to be more confident in results one should use several measures [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 application of the Rough Net concept in community validity or topological evolution estimation in dynamic networks. This concept supports the visualization of the community structure quality.

We use a similarity relation R' in our definition of Rough Net, because two nodes of V can be similar but not equal. Let $s : V \times V \rightarrow R'$ a function that measures the similarity between nodes of V, we define the similarity class of the node x, denoted $R'(x)$, as shown in Equation 8. We use the definition of R'-lower and R'-upper approximations for each similarity class taking into account equations 1 and 2 respectively.

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

There is a variety of distances and similarities for comparing nodes [1], such as Salton, HDI, HPI, similarities based on the topological structure, and Dice and Cosine coefficients which capture the attributes relations. In this paper, we use the Jaccard (Equation 9) similarity for computing the similarities based on the topological structure because it has the attraction of simplicity and normalization. In this paper, we emphasize the use of the network topology necessary to apply the concepts of RST in complex networks. For that reason, the Equation 9 has been explicitly describe, where $\Gamma(X)$ denotes the neighborhood of node x including it.

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

3.1 Decision system for applying RST on *monoplex* networks

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 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 systems $DS_{monoplex} = (V, A \cup d)$, where A is a finite set of topological 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.

3.2 Decision system for applying RST on *multiplex* networks

Multiplex are powerful enough though to allow modeling different types of networks including:

1. **Multi-relational networks,** in which each layer encodes one relation type. Also, it is possible to be considered as a *monoplex* network by applying a flattening process.
2. **Attributed networks,** in which additional layers can be defined over the node set as a similarity graph induced by a similarity measure applied to the set of node attributes, where each layer represents a different context. Notice that, each subset of attributes or layer topology corresponds to the features for a specific context of the problem.
3. **Dynamic networks,** in which each layer corresponds to the network state at a given time stamp (or each layer represent a snapshot). Like a time-series analysis if attributes are capture in each time, a dynamic complex networks can be represented as an attributed dynamic network [19].

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^k_{ij}

indicates the relationships between nodes i and j in the k-th layer. If we apply some community detection algorithm to the whole *multiplex* network topology by considering *multiplex* community detection approaches, 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 community result as a decision systems $DS_{layer_k} = (V, A_k \cup d)$, 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.

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 other more complex approaches used in the deep learning. Taking into account these variant, we can consider the combination of the topological structure of the transformed network and the CD results as a decision systems $DS_{monoplex} = (V, A \cup d)$, where $A = \bigcup_{k \in L}$ 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 briefly describe three important tasks in the community detection analysis: community detection validity, the evolutionary estimation in dynamic networks and the interpretable visualization schema for community detection.

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 aim of a community detection algorithm is discovering dense subgraphs by considering the structural information in terms of network linkages between nodes. 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. Several methods and measures have been proposed to detect and evaluate communities, respectively in both *monoplex* and *multiplex* networks [2, 3, 13]. 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 Section 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, next we provide a general procedure with implementation details using the Rough Net definition. Notice that, $R'_k(x)$ is computed by considering the attributes or topological features of networks in the k-th layer, as shown in Equation 8. The algorithm 1 allows us to measure the quality of the community structure using Rough Net, by considering the quality and precision of each community.

Algorithm 1 Community Detection Validity

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

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

if G is a *monoplex* network **then**
    $DS[1] \leftarrow DS_{monoplex}$ (See section 3.1)
    $C[1] \leftarrow communities(G, d)$
else if G is a *multiplex* network **then**
    for k in L **do**
        $DS[k] \leftarrow DS_{layer_k}$ (See section 3.2)
        $C[k] \leftarrow communities(layer(G, k), d)$
    end for
end if
for k in $(1 : size(DS))$ **do**
    Obtain the similarity class $R'_k(x)$ based on Equation 8
    for X in $C[k]$ **do**
        Calculate $R'_{k*}(X)$ and $R'^{*}_{k}(X)$ approximations (See Equations 1 and 2)
        Calculate $\alpha(X)$ and $\gamma(X)$ approximation measures (See Equations 3-4)
    end for
    Calculate $\gamma(DS_k)$, $\alpha(DS_k)$ and $\alpha_w(DS_k)$ in DS_k (See Equations 4-7)
    $\gamma_G(DS) += \gamma(DS_k)$, $\alpha_G(DS) += \alpha(DS_k)$ and $\alpha_{w_G}(DS) += \alpha_w(DS_k)$
end for
$\gamma_G(DS) = \gamma_G(DS)/L$, $\alpha_G(DS) = \alpha_G(DS)/L$ and $\alpha_{w_G}(DS) = \alpha_{w_G}(DS)/L$

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 link with their partners than with the other nodes [19]. For that reason, the key feature of the community detection analysis in dynamics 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 a new measure based on the Rough Net definition for estimating in a real number the change level during a specific window time-stamp.

We need to consider 2-consecutive layers for computing the quality, accuracy and weighted accuracy of classification measures in the evolutionary estimation. For that reason, we need to apply the RST for community validity two times. The former RST 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 RST 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 measure can be applied over a window size K by considering the aggregation of the quality classification between all pairs of consecutive (adjacency) layers.

4.3 Visualization schema for community detection analysis

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. Below we present a schema for visualizing the interaction 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. The algorithm 2 allows us to represent the community structure quality in an interpretable way. Notice that, real-world complex networks usually are composed by many nodes, edges and communities.

The community similarity used for weighted the interactions between communities in the network representation is formalized in Equation 11. The $BN'(X, Y)$, computed by $s_{BN}(X, Y)$, captures the proportion of nodes member of the community X, which cannot be unambiguously classified into this community but belong to the community Y. The above idea is computing based on the boundary region BN of both communities X and Y. Due to BN' is non-symmetric, the measure s_{BN} is computed by considering the application of BN' (Equation 10) in both senses (i.e., $X \to Y$ and $Y \to X$).

$$BN'(X, Y) = \frac{BN(X) \in \cap Y}{BN(X)} \tag{10}$$

$$s_{BN}(X, Y) = \frac{BN'(X, Y) + BN'(Y, X)}{2} \tag{11}$$

Algorithm 2 Visualization for Community Structure Analysis

Input: A complex network G (attributed or not), detected communities, a similarity threshold ξ and a similarity function between nodes (topological or non-topological features)

Output: Community network representation

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

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. The performance of a CD algorithm may be judged differently depending on which measure is used. To be more confident in results we should use several measures [2, 5]. Thus, we compare the reported result of our new community detection validity measures (Accuracy and Quality of Classification) with the more used internal and external measures used for community detection validity: modularity, adjusted rand (AR), normalized mutual information (NMI), rand, variation of information (VI) measures [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 an external references for operating. These measures except VI, like a modularity, 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 $VIC = 1 - VI$).

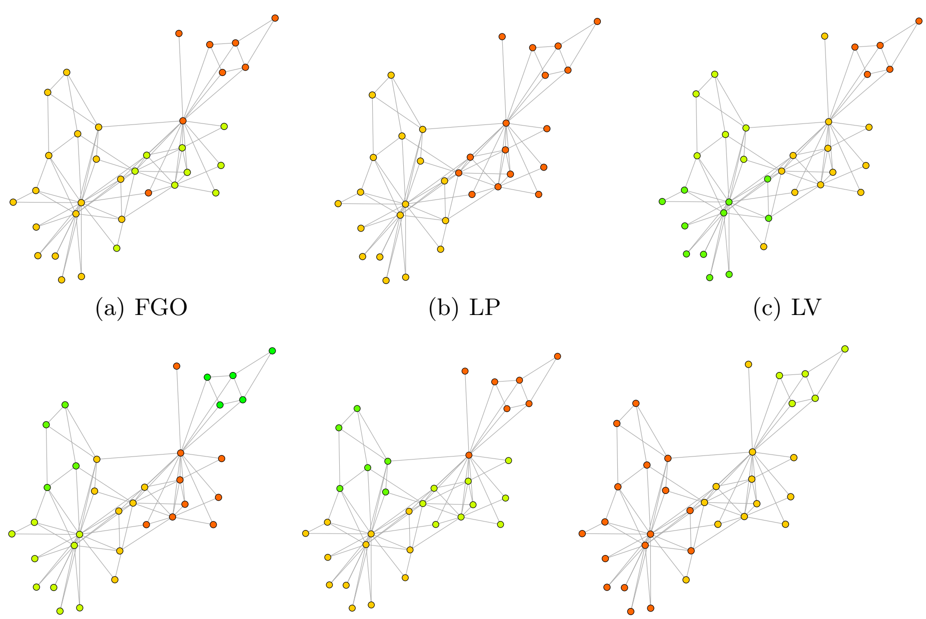
5.1 Zachary Network

Zachary is the much-discussed network [5] of friendships between 34 members of a karate-club at a US university. The two real communities (ground-truth) coincide identically with the division of the network founded by the LP algorithm, as

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

shown in Figure 1(b). Here, we explore the performance of the Rough Net definition in the quality evaluation of the community structure reached by popular CD algorithms.

Figure 1 shows the community structures reported by the application of the standard community detection algorithms 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 division observed in real life (ground-truth). The community structure reached by the LP algorithm coincides identically with the ground-truth. For that reason, we can affirm that the LP algorithm reported the best division. However, in Figure 2 we can observe that the modularity values not distinguish the LP as a best conformation of nodes into communities, while the proposed accuracy and quality measures based on the Rough Net definition, assign the higher value to the LP conformation regardless of the threshold value used. On the other hand, our measures grant the lowest quality results for the CD 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 the most neighbors of this node are in another community. Indeed, the FGO and WT are the following lowest results reported by our measures.



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

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

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

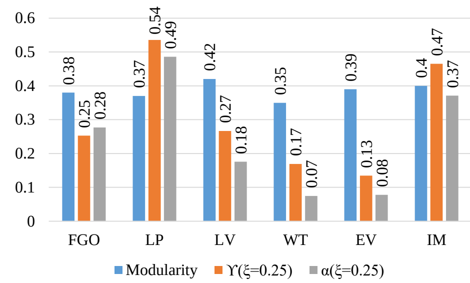


Fig. 2: Performance of the internal measures on the evaluation of the Zachary community structure.

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.

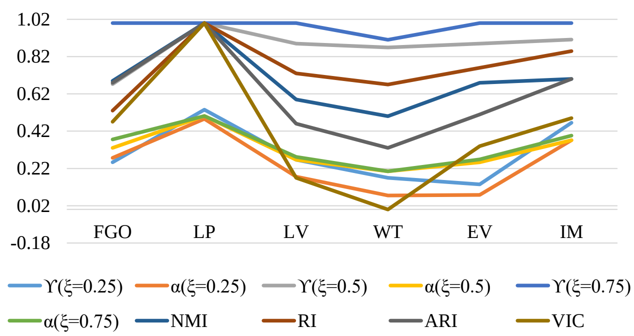
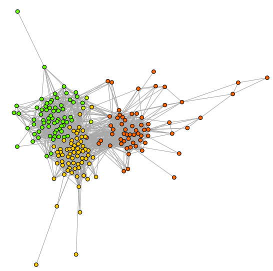


Fig. 3: Performance of the proposed measures on the Zachary CD evaluation.

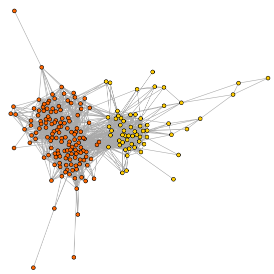
5.2 Jazz Network

The Jazz network [6] represents the collaboration between jazz musicians, where each node represent a jazz musician and interactions denoting that two musicians playing together in a band. Six CD algorithms were applied to this network with the objective of subsequently exploring the behavior of validation 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

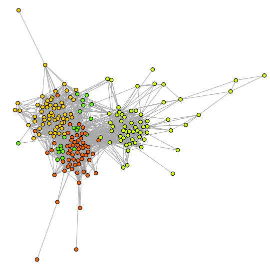
[6] http://konect.cc/networks/arenas-jazz/



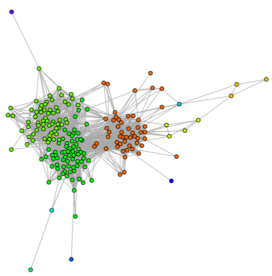
(a) FGO



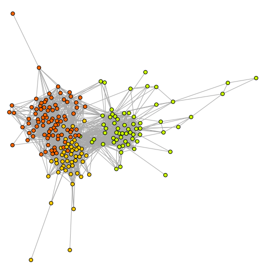
(b) LP



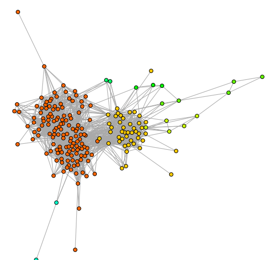
(c) LV



(d) WT



(e) EV



(f) IM

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

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 Figure 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 in this example, than the modularity measure. On the other hand, 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.

5.3 CElegans network

Caenorhabditis elegans connectome (CElegans) is a *multiplex* network [7], which 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

[7] http://deim.urv.cat/~alexandre.arenas/data/welcome.htm

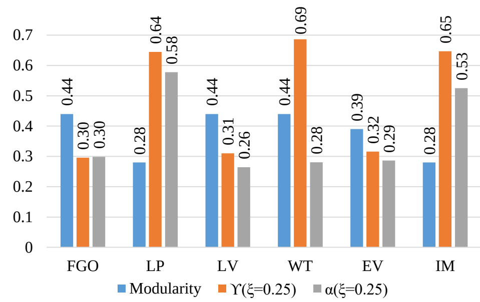
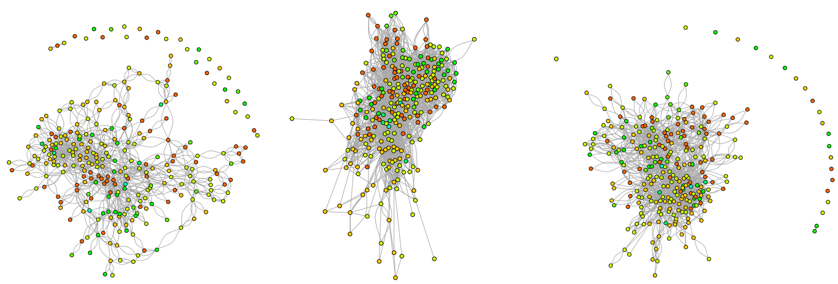


Fig. 5: Performance of the internal measures on the Jazz CD evaluation.

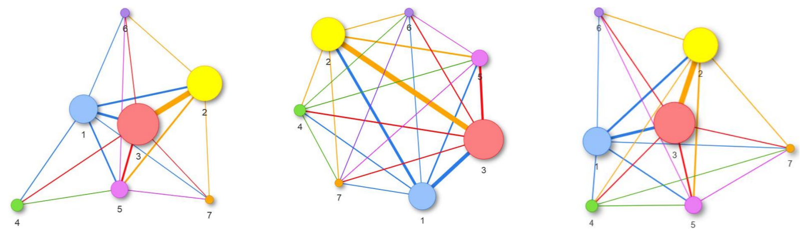
obtained by the application of the MuxLod CD algorithm [10]. Notice that, a strong community structure result must correspond to a structure of densely connected subgraphs in each network layer. This reflexion property is not evident for these communities in the CElegans network. For that reason, both 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



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

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

7 shows the interactions between the communities in each layer by considering the MuxLod community structure and the schema described in section 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 the topology considering these layers as consecutive. To estimate these results, we apply the methodology described in section 4.2. Figure 8 shows the modularity, accuracy and quality obtained values, which reflect that the community structure between the layers does not completely match, so it can be concluded that the topology has evolved (changed).



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

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

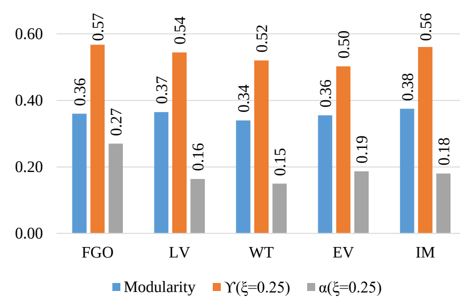


Fig. 8: Performance of the internal measures on the CElegans CD evaluation.

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 visualization schema can be allowed to the experts a deep understanding of complex real systems. For the future work, we propose to extend the applications of Rough Net definition 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. Social Networks 54, 41–49 (2018)
2. Aldecoa, R., Marín, I.: Surprise maximization reveals the community structure of complex networks. Scientific reports 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: Granular Computing: at the Junction of Rough Sets and Fuzzy Sets, pp. 229–248. Springer (2008)
5. Bai, L., Liang, J., Du, H., Guo, Y.: A novel community detection algorithm based on simplification of complex networks. Knowledge-Based Systems 143, 58–64 (2018)
6. Berlingerio, M., Pinelli, F., Calabrese, F.: Abacus: frequent pattern mining-based community discovery in multidimensional networks. Data Mining and Knowledge Discovery 27(3), 294–320 (2013)
7. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment 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 (Dec 2004)
9. De Domenico, M., Solé-Ribalta, A., Cozzo, E., Kivelä, M., Moreno, Y., Porter, M.A., Gómez, S., Arenas, A.: Mathematical formulation of multilayer networks. Physical Review 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 Intelligent Informatics Bulletin 16, 24–28 (2015)
12. Newman, M.E.: Finding community structure in networks using the eigenvectors of matrices. Physical review E 74(3), 036104 (2006)
13. Newman, M.E., Leicht, E.A.: Mixture models and exploratory analysis in networks. Proceedings of the National Academy of Sciences 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. Cybernetics & Systems 29(7), 661–688 (1998)
16. Pons, P., Latapy, M.: Computing communities in large networks using random walks. In: International symposium on computer and information sciences. pp. 284–293. Springer (2005)
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 (Sep 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. Physics Reports 787, 1–97 (2019)