**WEST UNIVERSITY OF TIMIŞOARA**
**FACULTY OF MATHEMATICS AND COMPUTER SCIENCE**
**BACHELOR/MASTER STUDY PROGRAM: Artificial Intelligence and Distributed Computing**

# MASTER THESIS

**SUPERVISORS:**
Prof. Dr. Gabriel Istrate
Conf. Dr. Cosmin Bonchiş

**GRADUATE:**
Victor Bogdan

**TIMIŞOARA**
**2021**

**WEST UNIVERSITY OF TIMIŞOARA**
**FACULTY OF MATHEMATICS AND COMPUTER SCIENCE**
**BACHELOR/MASTER STUDY PROGRAM: Artificial Intelligence and Distributed Computing**

# Layer Centrality in Multilayered Networks

**SUPERVISORS:**
Prof. Dr. Gabriel Istrate
Conf. Dr. Cosmin Bonchiş

**GRADUATE:**
Victor Bogdan

**TIMIŞOARA**
**2021**

# Abstract

In recent times, the interest regarding multiplex networks has increased, especially in the case of measuring the influence of certain layers on the nodes and the centrality of the nodes. In this work we are going to present our proposed solution for determining the layer centrality in a multilayered network for any given node, which is based on existing centrality measures for single layered networks and techniques from game theory. We have used together with our approach additive centrality measures in order to determine the influence of each layer over each node in a multi-layered network. The obtained results show that by using our approach we get relevant information regarding the centrality of layers for each node in a multilayered network and an overall understanding of the influence of each layer at network level.

# Contents

# Chapter 1

# Introduction

## 1.1 State of the Art

During the last few years a couple of methods of computing node and layer centralities in multilayered networks were developed. In [RIAB18] the authors propose an algorithm, MultiRank, for ranking nodes and layers in multiplex networks. It is based on a coupled set of equations which determines the centrality of the nodes and the influence of the layers in the multiplex network. The centrality of a node is higher if the respective node is present highly influential layers from the network, but also to other central nodes. Also, a layer has more influence if it contains more central nodes than other layers, at network level. In their experiments, the authors used three multiplex network datasets. Their results show that the MultiRank algorithm is flexible, it can be applied to undirected and directed multiplex networks which have weighted or unweighted links. Also, it extracted relevant information from all three datasets. Thus, the proposed algorithm can be used for many multiplex network dependent applications in fields such as transportation, economics, sociology and others.

Another similar approach is presented in [TVY21]. In this paper, the authors propose an algorithm based on the Adapted PageRank and the two-layer PageRank approach for biplex networks. Thus, the algorithm offers the possibility of controlling the importance assigned to the network topology. Also, it offers a solution to the nodes which are isolated in any of the layers by introducing a residual value for all nodes. The biplex algorithm is then generalised further such that it can be used on a multiplex network. The athuors have demonstrated the the possibilities and characteristics of the algorithm by using a dataset, a biplex network, which contains aggregated origin-destination flows of private cars in Rome. The biplex network contains a private car flow layer and local urban bus connectivity layer. The results obtained by applying the algorithm, show the most central locations in the city. In their experiments, the authors use four different cases, in all of which the most central city locations are highlighted but in each case in a different order according to the chosen parameter values.

In [KNR19] the authors study the effects of inter-layer coupling on the centrality measures. As previously mentioned, the number of network applications rose and in some of them, there are also inter-layer interactions, besides the intra-layer node connections. There are many well known centrality measures, but in the context of multiplex networks, defining centrality measures becomes a more challenging task. As the authors present in this paper, the extension of existing centrality measures to multilayer networks is not straightforward. They focus on the impact of three different types of inter-layer connectivity for existing centrality measures in multilayer networks: adjacent layers diagonal coupling, diagonal coupling and cross coupling. The authors conclude that it is a promising direction to extend this study to more complex graph structures, such as hypergraphs and knowledge graphs.

Another interesting multiplex networks study is presented in [BCMR17]. In this paper, the authors offer a taxonomy and experimental evaluation of existing approaches for computing similarities between the layers of a multiplex network. They also extend the existing approaches and offer a set of practical guidelines for applying them on networks. In order to study the presented approaches, a framework was developed, which also offers the possibility of defining functions which were not tested by the authors. The results which are obtained by applying these approaches suggests that for a given multiplex network, not all methods are always appropriate for determining the similarity between the layers. Also, there are cases in which priority should be given to one of the layers. The authors conclude that splitting the problem of computing the layer similarities in two: deciding what to observe and deciding how to compare these observations, offers the possibility of generating custom layer comparisons which are more appropriate for a given real problem or application.

## 1.2 Motivation

In various components of society, such as social, economical, political, etc..., most of the gathered data can be represented as a graph. Some common examples include graphs which represent the friend connections on different social platforms, multiple transactions between entities from different countries or even a graph representing the connections and different number of publications between researchers. As mentioned in Section State of the Art, in the literature there are some approaches which study the influence of the layers in multi-layered networks.

Because the data collected by different platforms and systems continuously increased in size and complexity, other representations of the data have proven to be useful. Multilayered networks can be considered as a suitable representation, where each layer represents a different type of connections between nodes than the others. In the context of social platforms, such a multilayered network could represent the connection between individuals on different platforms. Some of them might be present on all platforms, whereas others only on some of the platforms. The complexity of such a multilayered network can be increased by adding another layer which represents the connections with friends which are from the workplace, and thus create a multilayered network which does not contain data only from social platforms.

Taking a dataset and displaying it and its relations as a multilayered network can be done in numerous ways, by applying certain rules and conditions. In most cases, the analysis of such a multilayered network is a complex task, our case being **determining the most influential layer for a specific node**. As described in the previously mentioned research papers, in the case of single layered networks, the centrality of each node can be directly computed by using centrality measures such as Degree Centrality, Closeness Centrality [Fre78], Betweenness Centrality [Fre77], Katz Centrality [Kat53], Harmonic Centrality [ML00]. In order to try to solve the problem of finding the most influential layer for a node in a multilayered network, we are proposing and algorithm which uses some of the centrality measures, specifically those which are additive, together with techniques from game theory.

# Chapter 2

# Proposed Approach

## 2.1 Overview

**Our approach is based on two mathematical concepts**, **additive centrality measures** and the **Shapley Value** [Sha52], the latter requiring adapting the context in order to be able to apply techniques from game theory. By representing the data as a multilayered network, we can consider each layer as being an agent, thus adapting our context by creating a game between the layers/agents. In order to compute the contribution of each layer/agent to a specific node, we can consider that all layers form a coalition. The initial contribution of each layer/agent is determined by the centrality measure applied to the layers of the multilayered network. By computing the Shapley value, we will obtain a vector of values which represents the contribution (centrality) of each layer to a node. Therefore, like in the case of the centrality measure functions for single layered network which determine the centrality of nodes, we could use the previously described approach to determine the centrality of layers in a multilayered network.

By using the previously described adaption to a game, the Shapley value has a great significance in providing a fair way of distributing the rewards, in our case represented by the influence of each layer for a node. It is a solution concept of fairly distributing both, the costs and gains, to several agents which are in a coalition. The Shapley value can ultimately be interpreted as the average expected marginal contribution of one agent after all possible combinations have been considered.

The computation of the layer centralities for a given multilayered network using our approach requires some specific steps. First, we compute all combinations of all possible lengths of the layers. For each combination of layers, we flatten the layers in order to avoid duplicate connections. For a given node $n$, we then compute its centrality on each of the resulted flattened layers. Finally, we compute the Shapley value using for each flattened layer the centrality which we obtained in the previous step. Thus, we determine the centrality of each layer for a given node $n$. The entire proposed approach is further explained in detail in Section Algorithm.

## 2.2 Centrality measures

Centrality measures are one of the most popular approaches in determining a nodes importance in a network. There are many centrality measure functions which have different results based on the context they are used in. Some of the well known centralities are: Degree Centrality, Closeness Centrality defined by Alex Bavelas in [Fre78], Betweenness Centrality introduced by Linton Freeman in [Fre77] and Eigenvector Centrality proposed by Bonacich [Bon72]. Many of the previously mentioned and also other centrality measures have different variants such as the Harmonic centrality in the case of Closeness Centrality, which was proposed by Massimo Marchiori and Vito Latora in [ML00] and Katz Centrality in the case of Eigenvector Centrality, which was introduced by Leo Katz in [Kat53].

Some of the aforementioned centrality measures have been proposed by authors which have been active researches in the field or closely related fields of sociology. The Betweenness Centrality has been developed by Linton Freeman, an American structuralist sociologist, in order to quantify in a social network, the control of a human on the communication between other humans. Thus, one could determine the influence of one node over the communication of other nodes in a given network. Another example is the Closeness Centrality, which was introduced by Alex Bavelas, an American psychosociologist.

Naturally, besides the mathematical significance and because of the numerous use cases in sociology, centrality measures have been used in studies which contained networks representing communities of different types such as political, economical and social, which fundamentally represent different components of the society. In our approach we use a social multilayered network presented in Section AUCS dataset, whose layers represent different types of social connections between individuals.

As we have already mentioned, there are many centrality measures, with different properties. Because our approach focuses on determining the layer centrality in a multilayered network, we are going to use only additive centrality measures. This is intuitively related to the fact that by increasing the number of layers on which a respective node is present, the centrality of the node at multilayered network level increases, in the flattened version of the network the node has an increased number of direct and indirect connections, while the centralities of the layers for the respective nodes are altered based on the connections in each layers. It is also important to mention that we are going to use the non-normalised results of the centrality measure functions, in order to obtain additivity.

## 2.3 Shapley value

In order to use game theory techniques, we have adapted the context by creating a game between the layers of a given multilayered network, where for each node, the layers form a coalition. The marginal contribution of each layer will be computed using the Shapley value, Equation 2.1, considering in our case that the total contribution of the layers is equal to 1.

$$\phi_i(N) = \frac{1}{|N|!} \sum_{S \subseteq N \setminus \{i\}} |S|!(|N| - |S| - 1)! \left(v(S \cup \{i\}) - v(S)\right) \tag{2.1}$$

In our context of a coalitional game between the layers, we define the set $N$ as the set of all layers in the multilayered network. Also, in our approach, the function $v$, which gives the value for any subset of layers, is represented by an additive centrality measure. In order to avoid duplicate connections between the nodes when applying the function $v$, the respective combination of layers is flattened. It is important to mention that by using this approach, we compute the marginal contribution of layer $i$ for a specific node $n$. The algorithm which we developed in order to adapt our approach to Equation 2.1 is described in the following section.

## 2.4 Algorithm

As we have previously mentioned, the Shapley Value and the additive centrality measure functions are at the core part our proposed algorithm. We have used many centrality measures functions with our algorithm, most of them having different properties and results, which are highlighted in the upcoming chapters. Some of the centrality measures, by definition, take into consideration multiple aspects such as total number of connections of a node, total number of nodes in the layer, centrality of neighbouring nodes and many more.

### Pseudocode

**Algorithm 1** Layer Centrality

**Require:** $node$
**Require:** $layerPermList$
  $shapMap \leftarrow initShapMap()$
  $N \leftarrow layerPermList.size$
  **for** $i = 1, \ldots, N$ **do**
    $layerPerm \leftarrow layerPermList[i]$
    **for** $j = 1, \ldots, layerPerm.size$ **do**
      **if** j == 1 **then**
        $layer \leftarrow layerPerm[j]$
        **if** node is in $layer$ **then**
          $shapMap[layer] \leftarrow shapMap[layer] + v_{node}(layer)/N$
        **end if**
      **else**
        $layersBefore \leftarrow layerPerm[1 : j-1]$
        $layersWith \leftarrow layerPerm[1 : j]$
        **if** node is in $layersBefore$ **then**
          $contribLayer \leftarrow v_{node}(flatten(layersBefore)) - v_{node}(flatten(layersWith))$
          $shapMap[j] \leftarrow shapMap[j] + contribLayer/N$
        **end if**
      **end if**
    **end for**
  **end for**
  **return** $shapMap$

Because our approach uses the Shapley Value, we have developed Algorithm 1, in which for a given $node$ we compute the contribution of each layer in the multilayered network, by using a list of all possible layer permutations, denoted with *layerPermList*. The algorithm computes the marginal contribution of each layer and stores them into *shapMap*, a map which is initialised with the value 0 for each layer present in the multilayered network. After computing the marginal contributions of all layers, we set them to be equal to 1 and thus we obtain percentage values. Thus, we can offer a more tangible way of interpreting the results, which helps in further analyzing the influence of the layers on the nodes in the multilayered network.

The Shapley value computes the marginal contributions by taking into account the order of arrival for each layer permutation. Thus, we can identify two cases, one in which a layer in on the first position, $j == 1$, and its value is equal to the centrality value of *node* on the respective layer, denoted with $v_{node}(layer)$. The second case is represented by a permutation where the layer is not on the first position, and thus we have two subgroups of layers which we flatten, $layersBefore$ and $layersWith$, and then we apply the centrality measure function for *node*, $v_{node}(flatten(\text{layersBefore}))$ and $v_{node}(flatten(layersWith))$. $layersBefore$ represents a subgroup of layers which are as order of arrival before the layer. $layersWith$ is the union between the respective layer and the subgroup $layersBefore$.

## 2.5 Layer Marginal Contributions

In Figure 2.1 we can observe a simple example of a multilayered network and its layers, which contains 3 layers and a total of 6 nodes. We are going to use this example for a demonstration of our approach and algorithm, by offering details for each part of Algorithm 1. Some of the centrality measures require or can be tuned using different hyper-parameters. In the case of the Katz Centrality, we can choose multiple values for the attenuation factor $\alpha$. A more common hyper-parameter is the number of maximum iterations which can be set in multiple centrality measures computations, such as the Eigenvector Centrality, Katz Centrality and many more. Therefore, we are going to consider that all the hyper-parameters have been already chosen when applying the centrality measure $v$ to the considered *node*.

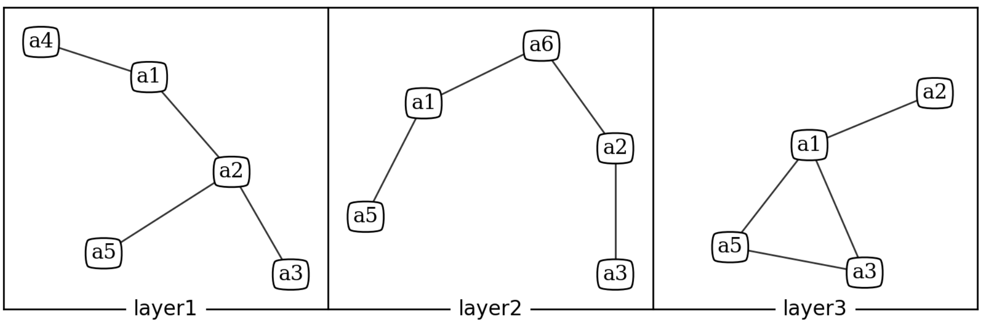


Figure 2.1: Multilayered Network Example

According to the Shapley value, in order to compute the contribution of the layers we first need to compute all possible layer permutations, which can be seen in the first column in Table 2.1. In the following columns, we can notice the computations of the marginal contributions for each layer, where $v$ represents the centrality measure function, which in our case will compute only the centrality of *node*. Thus, the first column in Table 2.1 represents the $layerPermList$ in Algorithm 1 and the rest of the columns represent the the nested for loops which compute the marginal contributions of each layer for the considered *node*. After finishing all computations, for each layer we sum up the results and divide them by the number of permutations $N$.

| Order of Arrival | Layer 1 Contribution | Layer 2 Contribution | Layer 3 Contribution |
|---|---|---|---|
| L1L2L3 | $v(L1)$ | $v(L1, L2) - v(L1)$ | $v(L1, L2, L3) - v(L1, L2)$ |
| L1L3L2 | $v(L1)$ | $v(L1, L2, L3) - v(L1, L3)$ | $v(L1, L3) - v(L1)$ |
| L2L1L3 | $v(L1, L2) - v(L2)$ | $v(L2)$ | $v(L1, L2, L3) - v(L1, L2)$ |
| L2L3L1 | $v(L1, L2, L3) - v(L2, L3)$ | $v(L2)$ | $v(L2, L3) - v(L2)$ |
| L3L1L2 | $v(L1, L3) - v(L3)$ | $v(L1, L2, L3) - v(L1, L3)$ | $v(L3)$ |
| L3L2L1 | $v(L1, L2, L3) - v(L2, L3)$ | $v(L2, L3) - v(L3)$ | $v(L3)$ |

Table 2.1: Marginal contribution equations example for all layers from Figure 2.1

We can notice that there are computations which contain subgroups of layers, such as $v(L1, L2) - v(1)$ for $layer2$ and $v(L1, L2, L3) - v(L1, L2)$ for $layer3$. In those cases, we apply the flattening method to those subgroups before computing the centrality of the considered node. The subgroups $v(L1, L2)$ and $v(L1, L2, L3)$ are represented by $layersWith$ in Algorithm 1 whereas $v(1)$ and $v(L1, L2)$ by $layerBefore$. Therefore, by using this approach we compute, using multiple centrality measure functions, the layer contribution for each node in a multilayered network.

In this section we are going to analyse the results of applying the Degree Centrality and the Katz Centrality on the multilayered example from Figure 2.1. These centrality measures can offer 2 different perspectives when interpreting the results. The Degree Centrality is based only on the number of nodes with which a node $n$ is connected with. On the other hand, the Katz Centrality is a generalization of the Degree Centrality by incorporating the importance of those neighbours, in the case of an undirected graph.

As previously mentioned, the multilayered network contains a total of six nodes, each present on all or only some of the layers. By analysing the layers we can notice that in the case of node $a5$, $layer1$ intuitively has more influence than $layer2$, because it offers a new connection, with node $a2$. On the other hand, in $layer2$ the node $a5$ is connected only to node $a1$, a connection which is also present in $layer3$. Therefore, although both layers, $layer1$ and $layer2$, offer only 1 connection, $layer1$ offers a unique one. By using the Degree Centrality and then the Shapley Value, we obtain results presented in Table 2.2, which show that indeed, $layer1$ is more has more influence. The values represent the percentages of the centrality of each layer for each node. Thus, we can say that for node $a5$, $layer3$ is the most central layer, whereas $layer1$ is more central than $layer2$ but less than $layer3$.

| Node | Layer 1 | Layer 2 | Layer 3 | Shannon Entropy |
|---|---|---|---|---|
| a1 | 30.0% | 30.0% | 40.0% | 1.57 |
| a2 | 50.0% | 37.50% | 12.50% | 1.40 |
| a3 | 16.67% | 16.67% | 66.67% | 1.25 |
| a4 | 100.0% | 0.0% | 0.0% | 0.0 |
| a5 | 33.33% | 16.67% | 50.0% | 1.46 |
| a6 | 0.0 % | 100.0% | 0.0% | 0.0 |

Table 2.2: Layer Centrality values in percentages after applying Degree Centrality

By considering the sum of marginal contributions to be equal to 1, we can say that the marginal contribution of each layer represents a probability. Therefore, we can make use of the Shannon Entropy to find out the amount of information we obtain for each node, which can have a great use when analysing the results of our proposed algorithm. In Table 2.2, we can notice that the Shannon Entropy obtains high values for the cases in which the differences between the values of the layer contributions for a given node are small. Thus, because multilayered networks can have a lot of nodes, we can find nodes which are influenced by multiple layers by searching high values obtained using the Shannon Entropy.

| Node | Layer 1 | Layer 2 | Layer 3 | Shannon Entropy |
|---|---|---|---|---|
| a1 | 32.55% | 32.41% | 35.04% | 1.58 |
| a2 | 36.5% | 33.92% | 29.57% | 1.58 |
| a3 | 30.69% | 30.37% | 38.94% | 1.57 |
| a4 | 98.46% | 0.75% | 0.79% | 0.13 |
| a5 | 33.49% | 30.46% | 36.05% | 1.58 |
| a6 | 1.44% | 97.52% | 1.04% | 0.19 |

Table 2.3: Layer Centrality values in percentages after applying Degree Centrality

The results of using the Katz Centrality on the same multilayered network presented in Figure 2.1 can be observed in Table 2.3. For nodes $a1$, $a2$, $a3$ and $a5$, we obtain almost the same amount of information according to the Shannon Entropy, around 1.58. This is related to the fact, that for those nodes, the differences between the layer centrality values is small. We can also notice that in the case of nodes $a4$ and $a6$, the layers on which they are not present still have a small centrality value. In this case, $layer3$ has a higher centrality value for node $a5$ than $layer1$, whereas when we used the Degree Centrality, $layer1$ had the highest centrality value. This is due to the fact that each centrality measure computes differently the centrality of the nodes on the layers, thus we can obtain different results in terms of the centrality of each layer to a respective node using our proposed approach.

## 2.6 AUCS dataset

In order to further analyze our proposed algorithm, we have used the AUCS dataset [RM15]. It contains 5 layers, one of which is a digital social connections layer, Facebook, Figure 2.3, and the rest are real life social connections layers: Coauthor, Figure 2.2, Leisure, Figure 2.4, Lunch, Figure 2.5, and Work, Figure 2.6. Each layer is undirected and all layers together build a multilayered network. In total, there are 61 unique nodes, which are actually anonymous employees from the research department of an unmentioned university. Each employee is present on one or more of these layers and can have from one to multiple connections.

| Layer | Nodes | Edges | Max. No. of Edges | Leaf Nodes |
|---|---|---|---|---|
| Coauthor | 25 | 21 | 5 | **15** |
| Facebook | 32 | 124 | 15 | 0 |
| Leisure | 47 | 88 | 14 | 10 |
| Lunch | **60** | 193 | 15 | 2 |
| Work | **60** | **194** | **27** | 1 |

Table 2.4: AUCS Dataset Layer Statistics

The AUCS dataset has a wide variety of properties when it comes down to the layers of the multilayered network. In Table 2.4, we can notice some of those, such as the total number of nodes, total number of edges, maximum number of edges for a single node and the total number of leaf nodes, which represent nodes with a single connection. Overall, these properties, but not only, affect the most the centrality measures which we used. In the case of the Degree Centrality, the number of edges represents the actual centrality value, therefore the maximum number of edges of a node on the layer represents the node with the highest centrality value. Also, in the case of the Katz Centrality, the centrality of the neighbours of a node affects its own centrality and therefore the overall number of edges in the layer matters.

We can observe that the Lunch layer and the Work layer have the most nodes and also the most connections. Regarding the number of leaf nodes, the aforementioned layers have the least whereas the Coauthor layer has the most leaf nodes. Interestingly, the Coauthor layer has the least number of nodes and edges, the maximum number of edges for a node being equal to 5. Regarding the latter property, the Work layer has at least one node with 27 edges and only one leaf node. The Leisure layer falls in between the maximum and minimum values for all properties. On the Facebook layer, there are no leaf nodes, thus each node has at least 2 connections.

From a social point of view we can determine some interesting aspects just by analyzing the statistics in Table 2.4. The Work Layer contains at least one node with 27 edges, whereas the Coauthor Layer the most connections of a node is equal to 5. We can also notice, that although this data belongs to a university department, the least number of edges and nodes is on the Coauthor layer, which might suggest that it is harder to connect with other persons through scientific work. On the other hand, the Lunch and Work layers suggest that in this case of a university department, most social interactions happen physically, as they contain the most number of nodes and edges.

Based on the centrality measures which we will use, the results will offer a different perspective about the layers and also about the nodes. In the case of the Subgraph Centrality, a large number of edges on a respective layer might benefit the nodes because of the total number of walks of all lengths starting and ending at a given node $n$. In the upcoming chapters, we are going to analyze the results of multiple centrality measures on the AUCS dataset.

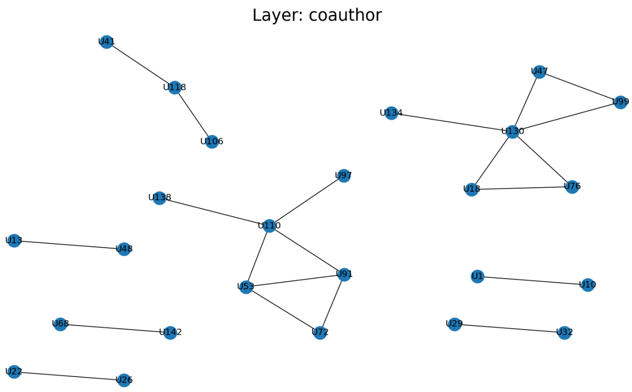


Figure 2.2: AUCS Coauthor Layer

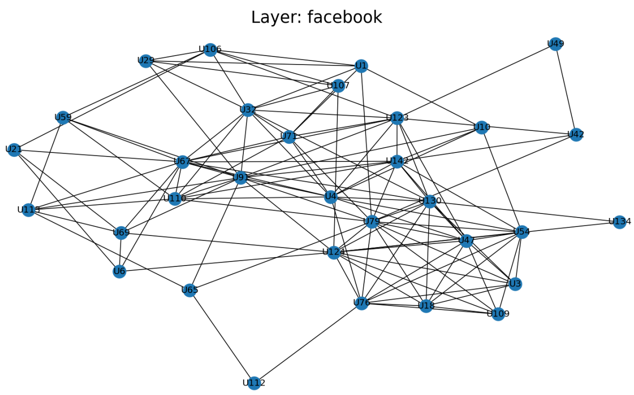


Figure 2.3: AUCS Facebook Layer

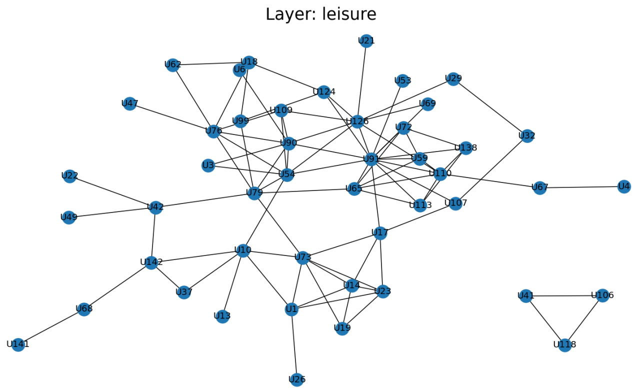


Figure 2.4: AUCS Leisure Layer

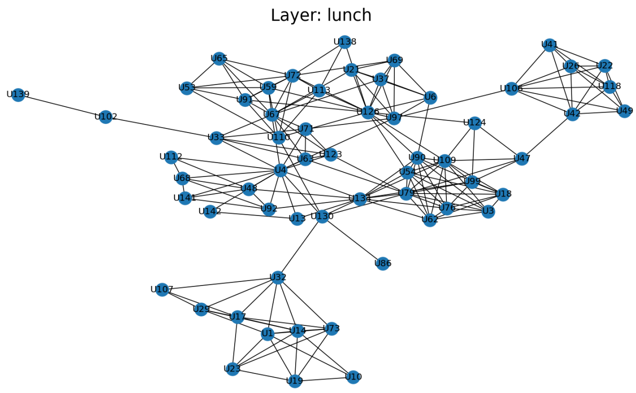


Figure 2.5: AUCS Lunch Layer

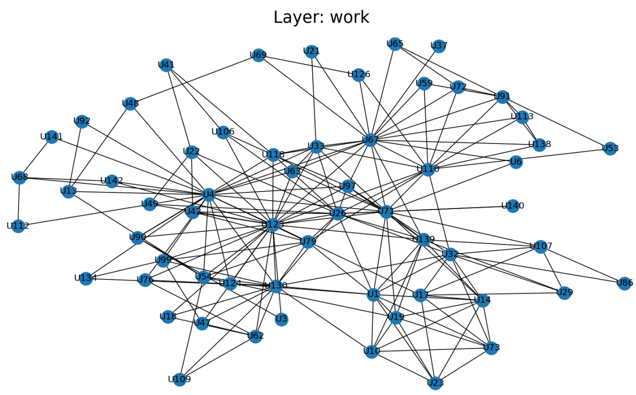


Figure 2.6: AUCS Work Layer

# Chapter 3

# Degree Centrality

We have mentioned in Chapter Proposed Approach, in order to determine the layer centrality in the AUCS dataset, one of the centrality measures which we have used in our proposed algorithm is the Degree Centrality, Equation 3.1. It computes the fraction of nodes to which a node $n$ is connected to. It is a widely known centrality measure which has led to the development of other more complex centrality measure such as the Katz Centrality. It is also important to mention that in order to obtain additivity, we have computed the Degree Centrality using Equation 3.1, without normalising the results on each layer.

$$C_{Degree}(j) = \sum_{n}^{j=1} A_{ij} \tag{3.1}$$

The results which we obtained highlight the fact that most of the nodes are highly influenced by the Lunch and Work layers. In Table 3.1, we can notice that the average influence of the Lunch layer is 37.45% and of the Work layer is 32.35%. The nodes are less influenced by the leisure, Facebook and Coauthor layers. The Coauthor layer has the least average influence, around 2%, which caused by the low number of nodes and connections present on the layer, see Table 2.4. On the last columns we can notice the number of nodes which where mostly influenced by each layer. In this case, most of the nodes were mainly influenced by the Lunch and Work layers, whereas the Coauthor layer did not influence any node the most. For node $U97$, both layers Lunch and Work, had the highest centrality value equal to 46.6. Also, in the case of the Degree Centrality, all layers have nodes for which their centrality is 0%.

| Layer name | Max. Influence | Min. Influence | Avg. Influence | Highest Influence |
|---|---|---|---|---|
| Coauthor | 20.97% | 0% | 2.08% | 0 |
| Facebook | 63.1% | 0% | 17.24% | 14 |
| Leisure | 41.87% | 0% | 10.88% | 2 |
| Lunch | **100%** | 0% | **37.45%** | **25** |
| Work | **100%** | 0% | 32.35% | 21 |

Table 3.1: Degree Centrality Results Statistics

In Figures 3.3, 3.4, 3.5, 3.6, 3.7, we can notice the layer centrality results for each node on each layer of the AUCS dataset. On the Coauthor layer, none of the nodes have a higher centrality value than 30% which suggests that this layer has the smallest overall contribution to the nodes. The Work and Lunch layers are the only layers which have a centrality value higher than 75%. We can also notice that the Facebook layer contains a couple of nodes for which the centrality is higher than 50%, in the cases of nodes $U142$, $U124$ and $U47$.

A large number of the nodes are present on multiple layers and by computing the Shannon Entropy on the results, most of the nodes have an entropy equal or higher than 1, up to 2.16. Thus, there are many nodes which are strongly influenced by more than one layer. Such an example would be node U18, which is present on all layers, and which we are going to further analyze. Table 3.2 and Figure 3.1 contains information about node U18 regarding the centrality of each layer on which it is present, the total number of connections and the number of connections of each order. The order, Definition 1, represents the number of layers on which U18 and another node are connected, more specifically in the case of U18, it is connected with node U47 only on the facebook layer.

| Node U18 | | | | | | |
|---|---|---|---|---|---|---|
| Layer name | Centrality | Total | Order 1 | Order 2 | Order 3 | Order 4 |
| Coauthor | 4.86% | 2 | 0 | 0 | 1 | 1 |
| Facebook | 34.03% | 8 | 1 | 5 | 1 | 1 |
| Leisure | 13.19% | 4 | 0 | 2 | 1 | 1 |
| Lunch | 34.03% | 8 | 1 | 5 | 1 | 1 |
| Work | 13.89% | 3 | 1 | 0 | 2 | 0 |

Table 3.2: Node U18 layer centrality analysis

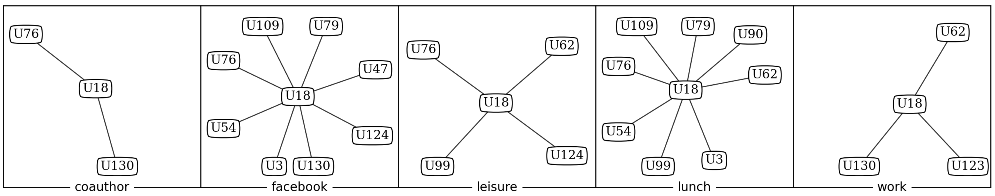


Figure 3.1: Node U18 subnetwork

By comparing the centrality values of layers facebook and lunch, we can notice that they are equal. Also, both layers have 8 connections, and for all orders, they have the same number of connections. In the case of the leisure layer, by comparing its centrality with those of layers work and coauthor we can observe that there is a relation between the number of connections which each layer offers but also the order of those connections. The *coauthor* layer offers U18 2 connections, of orders 3 and 5 whereas the leisure layer offer 4 connections of order 2, 3 and 4, thus, the leisure layer has more connections which are also, overall, of higher order. On the other hand, the work layer has a higher centrality and offers only 3 connections, out of which 1 is of order 1 and 2 of order 3.

Another similar example would be in the case of node U142, which we can observe in Table 3.3 and Figure 3.2. In the case of the work layer, which has 3 connections, it has a more connections than both, the coauthor and the lunch layer, but a smaller centrality than the lunch layer. Thus, by analyzing nodes U18 in Table 3.2 and U147 in Table 3.3, the results indicate that not only the number of connections but also the order of the connections have an impact on the influence of a layer for a node.

| Node U142 | | | | | |
|---|---|---|---|---|---|
| Layer name | Centrality | Total | Order 1 | Order 2 | Order 3 |
| Coauthor | 2.38% | 1 | 0 | 0 | 1 |
| Facebook | 63.10% | 11 | 7 | 3 | 1 |
| Leisure | 16.67% | 4 | 1 | 2 | 1 |
| Lunch | 9.52% | 2 | 1 | 0 | 1 |
| Work | 8.33% | 3 | 0 | 1 | 2 |

Table 3.3: Node U148 layer centrality analysis

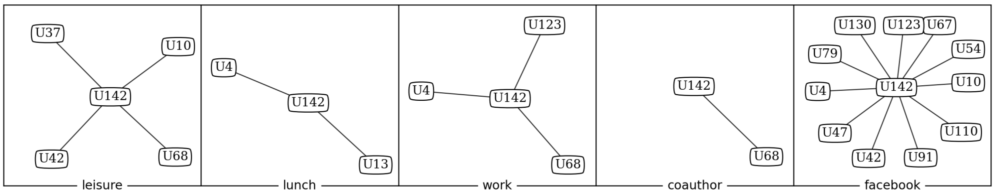


Figure 3.2: Node U142 subnetwork

**Definition 1.** *We define the connection order of a node $n$, which is connected to the node $x$, as the total number of connections between $n$ and $x$ on all layers in the multilayered network.*

**Definition 2.** *We define $\alpha$ as the order of a node present on layer $l$ connected to the node $x$.*

$$w_{l_x} = \sum_{1}^{i} \frac{1}{\alpha_i} (3.2) \tag{3.2}$$

Based on the connection order, Definition 1, and the total number of connections which a layer provides to a node, we can define Equation 3.2, where $c_{l_x}$ represents the centrality of layer $l$ for node $x$ and $\alpha$ the order of a node on layer $l$ connected to the node $x$ as per Definition 2. Thus, by applying Equation 3.2, we obtain the same centrality values for the layers in the examples presented in Table 3.2 and Table 3.3. Therefore in the case of the Degree Centrality we can also use Equation 3.2 in order to compute the layer centrality for any given node.

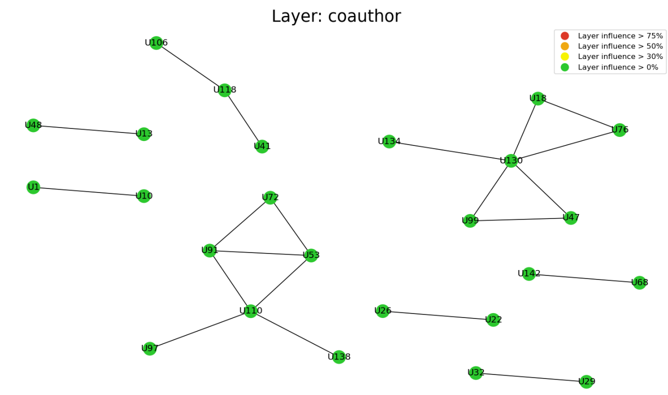


Figure 3.3: AUCS Dataset Coauthor Layer Degree Centrality Results

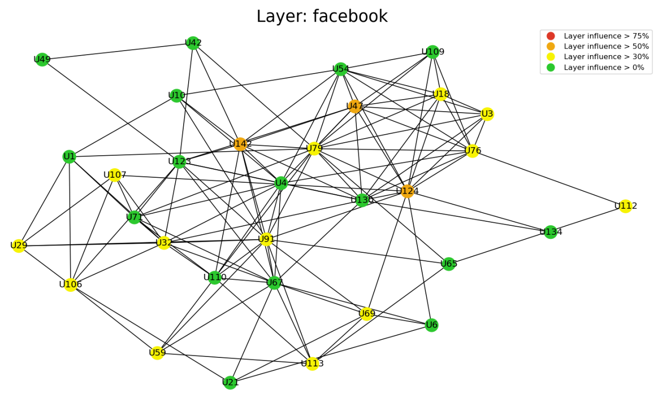


Figure 3.4: AUCS Dataset Facebook Layer Degree Centrality Results

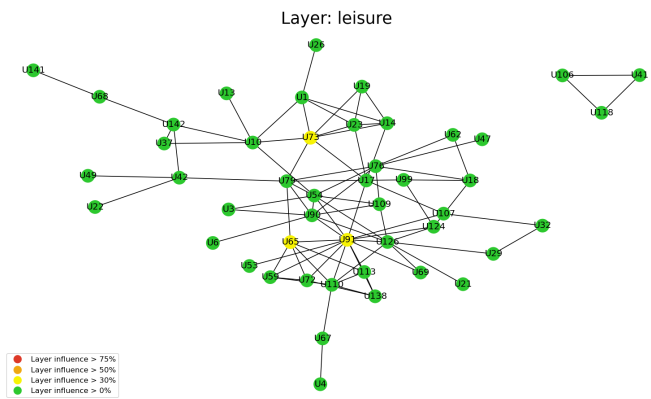


Figure 3.5: AUCS Dataset Leisure Layer Degree Centrality Results

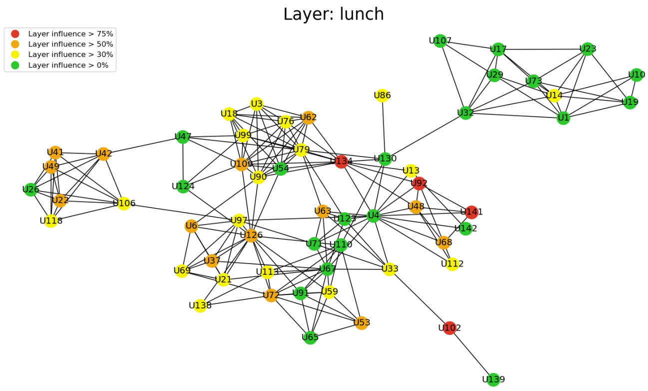


Figure 3.6: AUCS Dataset Lunch Layer Degree Centrality Results

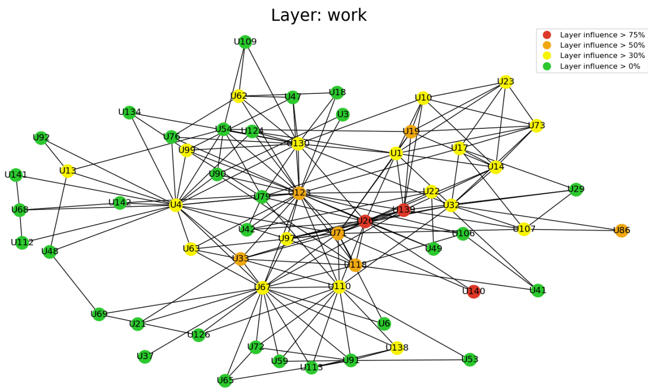


Figure 3.7: AUCS Dataset Work Layer Degree Centrality Results

# Chapter 4

# Katz Centrality

Another centrality measure which we used is the Katz Centrality [Kat53], introduced by Leo Katz in 1953. It is a centrality measure which takes into account the total number of walks between a pair of nodes. The Katz Centrality, Equation 4.1, is similar to PageRank [PBMW99] and Eigenvector Centrality [Bon72]. By taking attenuation factor $\alpha = \frac{1}{\lambda_{max}}$, where $\lambda_{max}$ is the largest eigenvalue of the adjacency matrix $A$, and $\beta = 1$, we obtain the Eigenvector Centrality.

$$C_{Katz}i = \alpha \sum_{i} A_{ij} C_{Katz}j + \beta \tag{4.1}$$

By modifying the $\beta$ parameter, we can provide extra weight to immediate neighbours. The connections with distant neighbours are penalized by the attenuation factor $\alpha$, where $\alpha < \frac{1}{\lambda_{max}}$ and $\alpha > 0$. The standard value for $\beta$ is 1. One could use other values for $\beta$, but the computations might not always converge as the required number of iterations might increase a lot. This would also be the case for different values for the attenuation factor, as it can have any value between the previously mentioned boundaries.

In our approach, in order to compute the attenuation factor, we had to take into consideration all the layers in the multilayered network. In order to compute $\alpha$, we computed first the maximum eigenvalues of the adjacency matrix for each possible layer combination of any length. In the case of multiple layers combinations, the adjacency matrix corresponds to the result of the flattening of the respective layers. In the next step, we have chosen the minimum out of all the maximum eigenvalues we have previously computed and used it as $\lambda_{max}$ in order to compute the attenuation factor. We have used this approach in order to maintain the additivity, by choosing the smallest of the maximum eigenvalues and therefore the centrality is the same or increases with the addition of new layers.

By using the Katz centrality, the results of the layer centralities suggest that overall, the Work layer is for the most of the nodes the most influential one. In Table 4.1 we can notice that the average influence of the Work layer is around 32.76%. In the case of the Lunch layer, although it does not have the highest average influence, it is the most influential layer for 27 nodes, more than the Work layer. These statistics do in fact reflect the AUCS dataset statistics from Table 2.4. The Coauthor layer does not influence any node the most, and has an average influence of 5.6%, the smallest by a wide margin.

| Layer name | Max. Influence | Min. Influence | Avg. Influence | Highest Influence |
|---|---|---|---|---|
| Coauthor | 21.31% | 0.02% | 5.6% | 0 |
| Facebook | 40.9% | 1.21% | 17.23% | 14 |
| Leisure | 33.31% | 0.2% | 14.61% | 1 |
| Lunch | 89.31% | 2.14% | 29.79% | **27** |
| Work | **93.81%** | **9.4%** | **32.76%** | 20 |

Table 4.1: Katz Centrality Results Statistics

By analysing the minimum Influence, we can notice that all layers have a minimum influence greater than 0%. This suggests that even in the case of nodes which are not present on a certain layer, the layer still offers influence to these nodes. For example, in the case of node $U102$, which is present only on the Lunch layer, the other layers also have a small influence: Coauthor 0.06%, Facebook 1.21%, Leisure 0.2% and Work 9.4%. Another similar example is node $U140$, which is present only on the Work layer, but is influenced by the other layers as well: Coauthor 0.13%, Facebook 3.52%, Leisure 0.4% and Lunch 2.14%. Therefore, some centrality measures like the Katz Centrality used together with our approach, can deduce that **layers which do not include a certain node**, **can in fact still offer it a certain influence at network level**, which is also due to the fact that in our approach we consider all combinations of layers when determining the centrality of each layer for the respective node.

The results of using the Katz centrality can also be observed in Figures 4.1, 4.2, 4.3, 4.4, 4.5. Layers Work and Lunch are the only ones which have a influence greater than 75% for a node, 89.13% for node $U102$ in the case of the Lunch layer and 93.81% for node $U140$ in the case of the Work layer. Also, the Coauthor layer does not influence any node more than 30%. Compared with other centrality measures used with our approach, the differences between the layer centrality values obtained using the Katz Centrality are smaller.

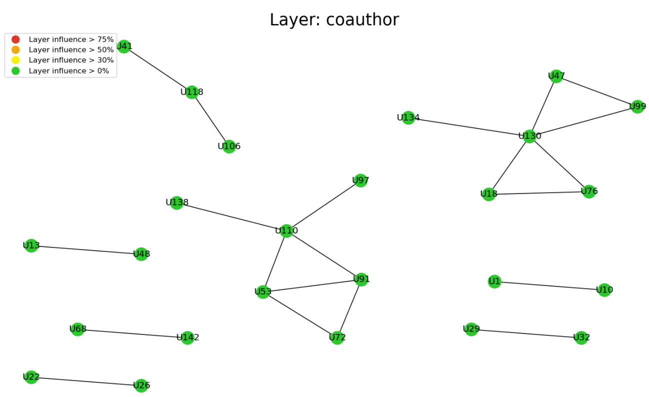


Figure 4.1: AUCS Dataset Coauthor Layer Katz Centrality Results

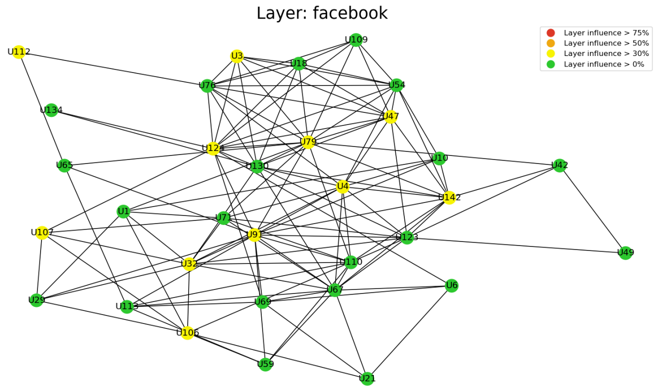


Figure 4.2: AUCS Dataset Facebook Layer Katz Centrality Results

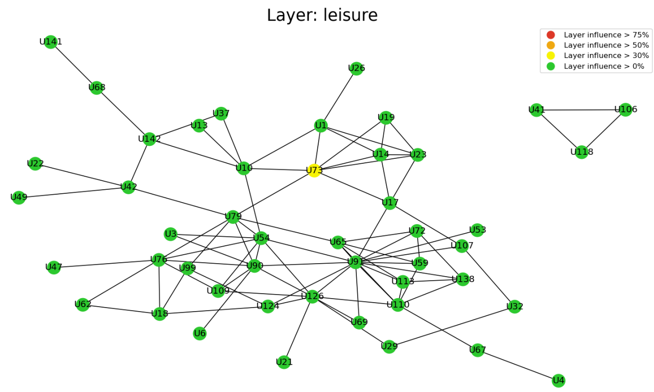


Figure 4.3: AUCS Dataset Leisure Layer Katz Centrality Results

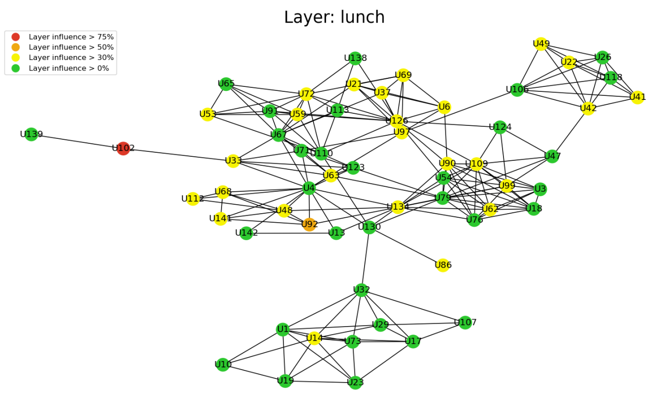


Figure 4.4: AUCS Dataset Lunch Layer Katz Centrality Results

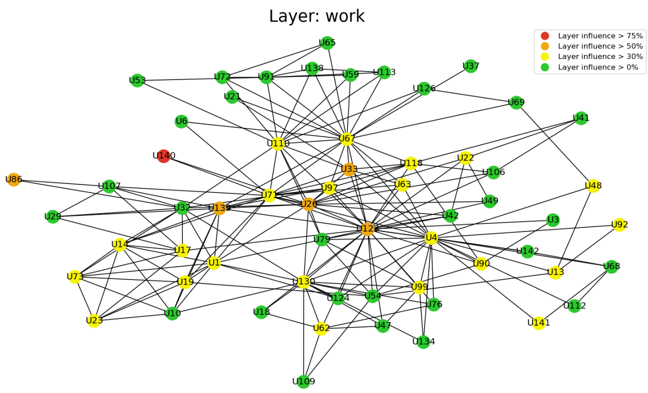


Figure 4.5: AUCS Dataset Work Layer Katz Centrality Results

# Chapter 5

# Subgraph Centrality

The Subgraph Centrality, presented in [ERV05], is another additive centrality measure which we have used together with our algorithm. It is equal to the sum of weighted closed walks of all lengths, starting from a given node $n$ and ending at the same node. In a graph $G$, a closed walk is an alternating sequence of nodes and edges, which starts and ends at the same node.

$$C_{subgraph}(u) = \sum_{j=1}^{N} (v_j^u)^2 e^{\lambda_j} \tag{5.1}$$

$$C_{subgraph}(u) = (e^A)_{uu} \tag{5.2}$$

This centrality measure characterizes the participation of all nodes in all subgraphs of a network, where the smaller subgraphs have more weight than the larger subgraphs, this being achieved by decreasing the weights as the path length increases. In Equation 5.1 we can observe the eigenvalues and eigenvector version of the subgraph centrality, where for a given node $u$, $v_j$ is an eigenvector of the adjacency matrix of a given network corresponding to the eigenvalue $\lambda_j$. The alternative exponential version is presented in Equation 5.2. The latter version of the subgraph centrality algorithm exponentiates the adjacency matrix.

| Layer name | Max. Influence | Min. Influence | Avg. Influence | Highest Influence |
|---|---|---|---|---|
| Coauthor | 6.07% | 0.72% | 1.48% | 0 |
| Facebook | 51.99% | **25.57%** | 31.69% | 17 |
| Leisure | 33.86% | 0.19% | 8.94% | 0 |
| Lunch | 21.09% | 4.54% | 24.7% | 3 |
| Work | **53.51%** | 2.07% | **33.2%** | **42** |

Table 5.1: Subgraph Centrality Results Statistics

In Table 5.1 we can observe the results of using the Subgraph Centrality measure with our proposed algorithm on the AUCS dataset. The Facebook and Work layers have the highest average influence on the nodes in the multilayered network, 33.2% and 31.69% respectively. Also, both of the previously mentioned layers have had the highest influence on the most of the nodes in the network, the Work layer having influenced 42 nodes and the Facebook layer having influenced 17 nodes. On the other hand, the Coauthor and Leisure layers haven't influenced any of the nodes the most.

Similarly to the results of the Katz Centrality presented in Table 4.1, the minimum influence of any of the layers is higher than 0%. Thus, by using the Subgraph Centrality measure together our proposed approach, nodes which are not present on certain layers, can still be influenced by these layers. In this case, on of the examples would be node $U49$, which is present on all layers except the Coauthor layer. Even though it is not present on the Coauthor layer, the respective layer has a centrality value to 1.25% for the node.

Interestingly enough, the Facebook layer has a minimum influence of 25.57%. This is strongly related to the fact that the Facebook layer has the highest ratio of edges per node. We can observe in the dataset statistics presented in Table 2.4, that there are 124 edges and 32 nodes on the Facebook layer, which gives a ratio close to 4. Therefore, because of the high number of edges it benefits the Subgraph Centrality as there are many smaller and also larger subgraphs which result in a high centrality value for each node and ultimately for the layer itself at network level.

In Figures 5.1, 5.2, 5.3, 5.4, 5.5, we can observe the layer influences for the nodes in the network. By using the Subgraph Centrality, there are no layers which have an influence higher than 75%. The Work and Facebook layers contain one node each, for which they have an influence higher than 50%, node $U102$ in the case of the work layer and node $U140$ in the case of the Facebook layer. Overall, for most nodes, the layers have an influence between 0.19% and 50%

Therefore, by using the Subgraph Centrality together with our algorithm we have obtained interesting results, which similarly to the results of the Katz Centrality, suggest that layers which do not contain certain nodes can still offer an amount of influence to these nodes. In this case, the nodes which are not present on a certain layer, can still have a higher number of closed walks when the respective layer is combined with another layer.

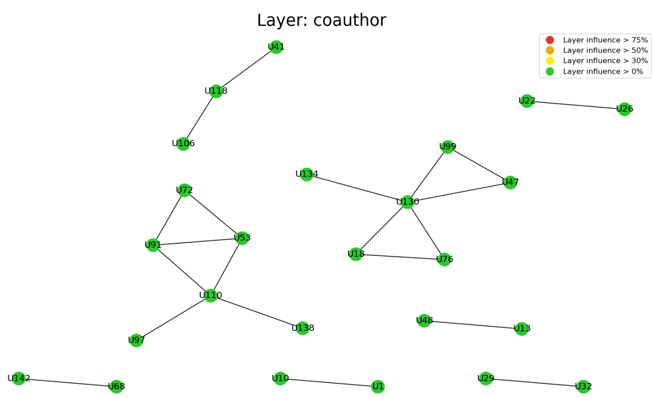


Figure 5.1: AUCS Dataset Coauthor Layer Subgraph Centrality Results

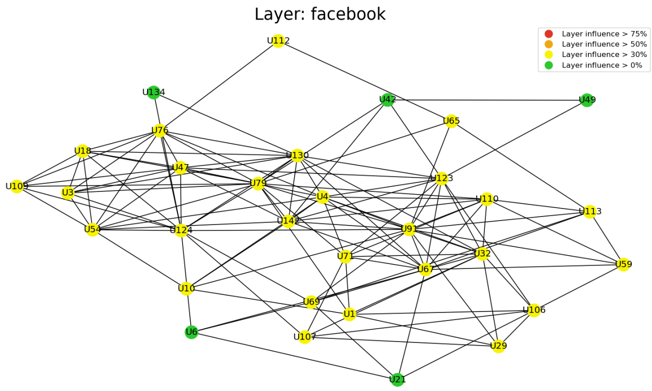


Figure 5.2: AUCS Dataset Facebok Layer Subgraph Centrality Results

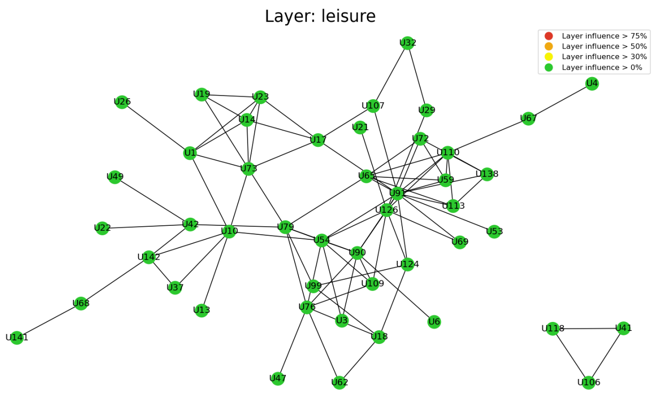


Figure 5.3: AUCS Dataset Leisure Layer Subgraph Centrality Results

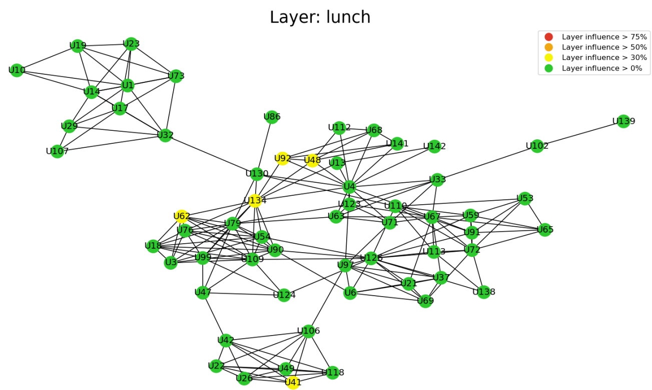


Figure 5.4: AUCS Dataset Lunch Layer Subgraph Centrality Results

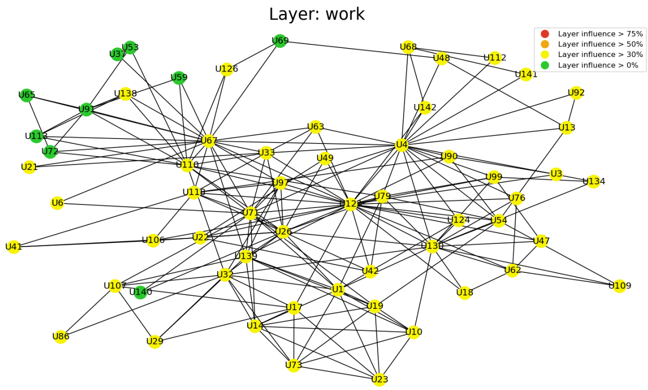


Figure 5.5: AUCS Dataset Work Layer Subgraph Centrality Results

# Chapter 6

# Harmonic Centrality

Another additive centrality measure which we have used together with our proposed algorithm is the Harmonic Centrality Measure. Proposed in [ML00], this centrality measure is a variant of the Closeness Centrality [Fre78], originating from the field of social network analysis. It was invented in order to solve the problem the Closeness Centrality had in the case of the unconnected graphs. An unconnected graph is a graph in which at least two nodes are not connected through any path. In the AUCS dataset there are two such layers which are disconnected graphs, the Coauthor and Leisure layers.

$$C_{harmonic}(u) = \sum_{v \neq u} \frac{1}{d(v,u)} \tag{6.1}$$

The Harmonic Centrality of a node $u$, Equation 6.1, is equal to the reciprocal of the shortest path distances from all other notes to the node $u$. The shortest path distance between the node $u$ and a node $v$ is represented by $d(v,u)$. In our approach we have considered that each edge will have a distance equal to 1. Thus, the shortest path is represented by the minimum number of edges from node $u$ to a node $v$.

| Layer name | Max. Influence | Min. Influence | Avg. Influence | Highest Influence |
|---|---|---|---|---|
| Coauthor | 5.3% | 0.0% | 1.35% | 0 |
| Facebook | 31.26% | 0.0% | 14.54% | 2 |
| Leisure | 32.08% | 0.0% | 13.7% | 0 |
| Lunch | 58.79% | 4.85% | 30.54% | 14 |
| Work | **95.15%** | **24.21%** | **39.87%** | **46** |

Table 6.1: Harmonic Centrality Results Statistics

In Table 6.1 we can notice the results of the Harmonic Centrality measure used with out proposed algorithm. The highest average influence values are obtained by the Work and Lunch layers. These two layers have also had the highest centrality value for the most of the nodes, 46 and 16 respectively. The Facebook and Leisure layers have similar values regarding the average influence, but the Leisure layer did not have the highest influence for any of the nodes whereas the Facebook layer had 2 nodes which it influenced the most. As in the case of the previously used centrality measures, the Coauthor layer has the smallest average average influence.

Regarding the minimum influence, the Coauthor, Facebook and Leisure layers have it equal to 0%. Unlike the Degree Centrality and similar to the Katz and Subgraph centrality measures, layers which do not contain certain nodes, can still offer them influence at network level. For example in the case of node $U62$, it is present only on the Leisure, Lunch and Work layers, but the Coauthor and Facebook layers still offer the respective node some influence which is reflected in the case of the Coauthor layer by the 0.33% centrality value and in the case of the Facebook layer by the 3.31% centrality value.

By using the Harmonic centrality measure, the Coauthor layer obtained the smallest average influence of just 1.35%, like for all the other centrality measures which we have used. Although we can notice that there are different results when using different centrality measures regarding the layer centrality values and also which nodes are influenced the most by which layer, in the case of the Coauthor layer it seems that it does not have influence of the nodes in the network. This comes somewhat as a surprise and in contrast with the context, because the dataset is representing the department of a university, but it still suggests that it is much harder to connect with other people in research, one of the reasons possibly being the variety of research topics.

In Figures 6.1, 6.2, 6.3, 6.4, 6.5, we can notice the influence of each layer when using the Harmonic Centrality measure. The Work layer is the only layer which has nodes on which its centrality value is higher than 75%, for nodes $U139$ and $U140$. Regarding the Coauthor, Facebook and Leisure layers, they do not have an influence higher than 50% for any of the nodes.

Overall the Work layer has the most influence for the nodes in the network and compared with other centrality measure which we have used, none of the had such a dominant layer from this point of view. The closest comparison would come from the Subgraph centrality, on which the Work layer was also the one with the highest average influence but did not have the highest influence for most of the nodes, see Table 5.1. Therefore, by using the Harmonic centrality measure, we have obtained interesting results which, similarly to the Katz and Subgraph centralities, suggest that the nodes can be influenced by layers on which they are not present.

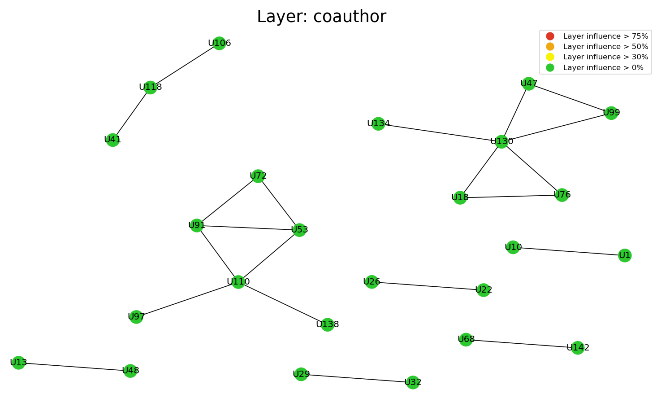


Figure 6.1: AUCS Dataset Coauthor Layer Harmonic Centrality Results

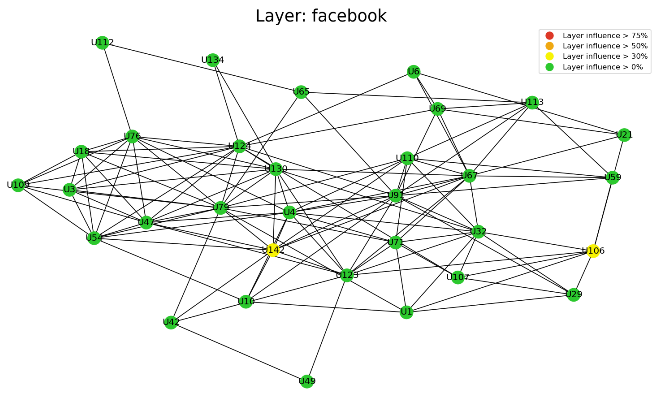


Figure 6.2: AUCS Dataset Facebook Layer Harmonic Centrality Results

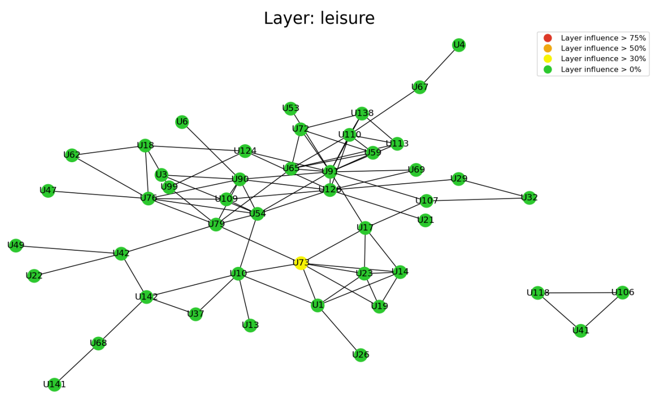


Figure 6.3: AUCS Dataset Leisure Layer Harmonic Centrality Results

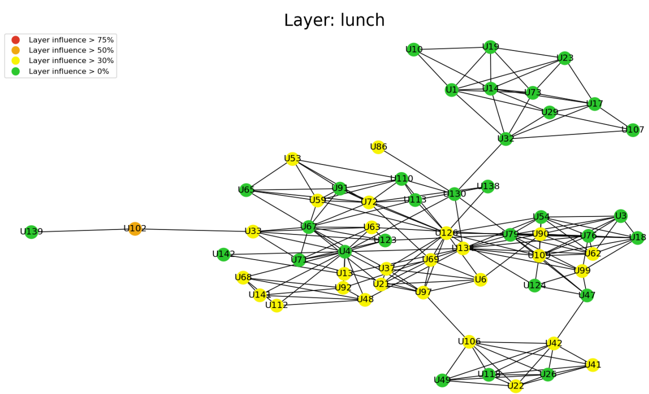


Figure 6.4: AUCS Dataset Lunch Layer Harmonic Centrality Results

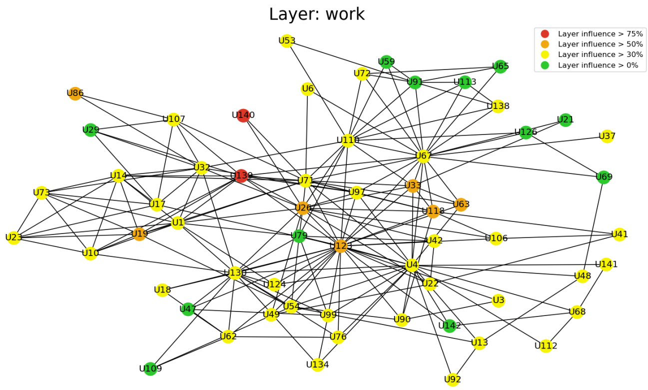


Figure 6.5: AUCS Dataset Work Layer Harmonic Centrality Results

# Chapter 7

# Conclusions and Future Work

Determining the layer centrality in multilayered networks is a challenging task, especially when trying to offer a flexible solution. Our proposed approach is mainly designed for multilayered networks which have undirected graphs as layers. As we have mentioned in Chapter 2, our proposed algorithm uses additive centrality measures and the Shapley Value. Each centrality measure offers a different method of interpreting the influence of a node in a graph, in our case on a layer. Therefore by using different centrality measure with our proposed approach, we have obtained different results which also offered different perspective when interpreting those results.

In the case of the Degree Centrality, Chapter 3, the results of our proposed algorithm, Table 3.1, show that the Lunch and Work layers are in the most cases the most central layers for the nodes in the AUCS dataset network. The Coauthor layer has the least average influence, around 2%, which suggests that it is harder to have connections with other people in the context of scientific work than in other contexts. Compared on the statistics of the AUCS dataset, Table 2.4, the results of the Degree Centrality confirm the intuition that the most central layers are Work and Lunch, followed by Facebook and Leisure.

By applying the Katz Centrality, Chapter 4, we have obtained different results, visible in Table 4.1, where the differences between the layer centrality values is smaller than in the case of the Degree Centrality. We can notice from the results that when we used the Katz Centrality, the layer with the most influence was the Work layer. As we have previously mentioned, the Katz Centrality takes into account the total number of walks between a pair of nodes, thus also considering the centrality of the neighbouring nodes when computing it for a certain node. Interestingly, in the case of nodes which are present only on one or a couple of layers, the layers on which those nodes are not present still have a centrality value greater than 0%. This suggests that these layers, overall, do offer influence to nodes which are not present on them. Such an information might be important, in the sense that the overall contribution might not come only from the layers on which a node is present, but also from the layers which have an indirect contribution for the node. In the case of the Harmonic centrality measure, the minimum influence is 0% but it is still influencing nodes which are not present on the layer itself.

The variance between the information obtained by using different centrality values can also be noticed by comparing the results of the Subgraph and Harmonic centrality measures. Overall, by using the Harmonic Centrality, the Work layer obtains high centrality values for most of the nodes in comparison with the other layers, the Lunch layer having the second highest average influence and number of highest influenced nodes. On the other hand, in the case of the Subgraph Centrality, the Work layer had the highest centrality value for most of the nodes, but the second highest average influence and number of highest influenced nodes is obtained by the Facebook layer.

Our results show that the Work and Lunch layers have the most influence over the nodes in the AUCS dataset. In the case of the Subgraph Centrality, had a higher average influence than the Lunch layer, see Table 5.1. Although the dataset represents the different types of connections between a group of people at a university, the Coauthor layer has the smallest average influence for any of the used centrality measures. Thus, our approach manages to detect that because the number of nodes present on the Coauthor layer and the number of connections between those nodes, even in a university department, are smaller on the Coauthor layer than on the other layers, the overall contribution at multilayered network is small. In the case of other layers, such as the Work, Lunch and Facebook layer we can notice, that depending on the centrality measure which we use, we can obtain different results. More specifically, in the case of the Work and Lunch layers, the total amount of nodes and total amount of edges is equal or very close, there are also other factors which can lead to different results, for example the difference between the maximum number of edges of a node on the layers which we can notice in Table 2.4.

The layer centrality values can offer valuable information about the multilayered network itself, but also about the nodes present on the layers. Therefore, by analyzing the results which we obtained, we can suggest multiple ideas of continuing this work. One of those would be to analyze other additive centrality measures, study the results and also analyze the centrality of the layers in different cases, for example what we have observed in the case of the Katz Centrality regarding the influence of layers on nodes which are not present on these layers. Another idea would be to analyze if we can use this information to influence the nodes themselves, by changing the influence of the layers. In the latter proposed idea, one of the questions would be on which layers and through which means could we achieve a positive or negative influence manipulation for the targeted node. Last but not least, different clustering techniques could be used in order to analyze the obtained node clusters based on the influence of the layers.

# Bibliography


[BCMR17] Piotr Bródka, Anna Chmiel, Matteo Magnani, and Giancarlo Ragozini. Quantifying layer similarity in multiplex networks: a systematic study. *CoRR*, abs/1711.11335, 2017.

[Bon72] Phillip Bonacich. Factoring and weighting approaches to status scores and clique identification. *The Journal of Mathematical Sociology*, 2(1):113–120, 1972.

[ERV05] Ernesto Estrada and Juan A. Rodríguez-Velázquez. Subgraph centrality in complex networks. *Phys. Rev. E*, 71:056103, May 2005.

[Fre77] Linton C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, 1977.

[Fre78] Linton C. Freeman. Centrality in social networks conceptual clarification. *Social Networks*, 1(3):215–239, 1978.

[Kat53] Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, March 1953.

[KNR19] Tarun Kumar, Manikandan Narayanan, and Balaraman Ravindran. Effect of inter-layer coupling on multilayer network centrality measures. *Journal of the Indian Institute of Science*, 99, 03 2019.

[ML00] Massimo Marchiori and Vito Latora. Harmony in the small-world. *Physica A: Statistical Mechanics and its Applications*, 285(3):539–546, 2000.

[PBMW99] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November 1999. Previous number = SIDL-WP-1999-0120.

[RIAB18] C. Rahmede, Jacopo Iacovacci, A. Arenas, and G. Bianconi. Centralities of nodes and influences of layers in large multiplex networks. *ArXiv*, abs/1703.05833, 2018.

[RM15] Luca Rossi and Matteo Magnani. Towards effective visual analytics on multiplex and multilayer networks. *Chaos, Solitons and Fractals*, 72:68–76, 2015.

[Sha52] Lloyd S. Shapley. *A Value for n-Person Games.* RAND Corporation, Santa Monica, CA, 1952.

[TVY21] Leandro Tortosa, Jose F. Vicent, and Gevorg Yeghikyan. An algorithm for ranking the nodes of multiplex networks with data based on the pagerank concept. *Applied Mathematics and Computation*, 392:125676, 2021.