International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 11, No.1 (Jan-2022)

https://dx.doi.org/10.12785/ijcds/110119

A Spectroscopy-Network-Based Method for Forgery Detection of Documents

Alaa Amged[1], Basim Mahmood[2] and Khalid A.K. Almukhtar[3]

[1]*Department of Computer Science, University of Mosul, Mosul, IRAQ*
[2]*Department of Computer Science, University of Mosul, Mosul, IRAQ / BioComplex Laboratory, Exeter, UK*
[3]*Ministry of Interior, IRAQ*

Received 18 May 2021, Revised 5 Nov. 2021, Accepted 13 Nov. 2021, Published 9 Jan. 2022

Abstract: This article introduces a novel approach for forgery detection of documents based on the spectroscopy of documents' matters. The proposed approach uses concepts from network science to generate a weighted network of spectrums for both the original and questioned documents together. The nodes of the network represent the spectrums and the edges are the correlations among them. The detection method is based on the number of clusters obtained from the tested network using a modified version of the Louvain algorithm. The spectrums of the inks and papers that were used in printing the documents were obtained using Laser-Induced Breakdown Spectroscopy (LIBS) technology. The proposed approach was tested under variety of cases such as inkjet prints, laser prints, and different kinds of printing papers. It was also examined under several clustering algorithms. The findings showed that the approach always successful in distinguishing different kinds of documents with an accuracy of 100%. Moreover, the results of the proposed approach can be visually interpreted, which is more comfortable to investigators. Finally, the proposed approach is considered simple and does not need complex computations compared to the approaches in the literature.

Keywords: Digital Forensics, Document Forgery, Forgery Detection, Complex Networks, LIBS Spectroscopy

1. INTRODUCTION

With the advent of new technologies, our life has become more vulnerable to a variety of risks. For instance, many tools (e.g., software applications and technologies) are available to the public. These tools can be used for producing or editing official documents. This kind of activity may be illegally performed by forgers. For example, performing unauthorized modifications on a document or even imitating an original document. Hundreds of document forgery cases have been reported every day around the world [1]. Preventing this kind of issue is not an easy task to perform due to the high availability of software and hardware tools. Therefore, one of the solutions is to adopt efficient techniques to detect forged documents [2].

Recently, forgery detection has been considered one of the activist fields in forensic science. Researchers around the world work intensively on developing approaches that are efficient in detecting forged documents [3]. However, developing such approaches is a challenging task especially when it comes to simplicity and time consumption [4]. The literature includes a vast number of studies that have been performed in the field of forgery detection. Some approaches use image processing techniques to detect forged documents [5]. However, this kind of technique is considered complex due to the heavy computations needed for each tested document [6]. On the other hand, other approaches use the spectrums of document components (e.g., printing inks, handwriting inks, and printing papers).

The spectrums can be extracted using spectroscopy technologies (e.g., LIBS) and used for forgery detection. This kind is considered more accurate since it deals with the characteristics of the material used to produce documents [7]. Laser-Induced Breakdown Spectroscopy (LIBS) is one of the most common technologies used in extracting the spectroscopy of materials. Spectroscopy approaches study the characteristics of matter's absorption when it is exposed to electromagnetic radiation [8]. Therefore, spectroscopy can be considered as a measurement of the radiation intensity of wavelength for a matter.

A. *Literature Review*

The analysis of spectroscopy has been developed to become crucial in the development of forensic science. The study of Cicconi et al.,[9] used LIBS for testing issues related to the commercial inks. The study performed classification on pen inks for one paper type and various paper types, determination of the deposition order of layered inks. Also, they analyzed a questioned document's

E-mail address: alaa.csp52@student.uomosul.edu.iq, bmahmood@(uomosul.edu.iq / biocomplexlab.org), almukhtar972@gmail.com

signatures and toners. Then, they identified up to seven characteristic metals for the inks examined, which allowed to fully discriminate all eight black inks on one type of printing paper. When the inks were checked on ten various papers, for several reasons, the right classification rates were reduced for some of them. One of the reasons was the presence of the same elements in both the ink and the paper ablated simultaneously with ink. The differential penetration of inks into paper was another reason. The experiments repeated at three crossing points, each involved a pair of blue or black inks were successful in five out of six cases.

Moreover, the literature shows different kinds of techniques when dealing with questioned documents. Ameh and Ozovehe [10] used Fourier Transform Infrared Spectroscopy (FT-IS) to investigate forensic examination of the inks extracted from printed documents. The authors compared and contrasted the extracted inks using two brands of printer cartridges. The findings showed that FT-IS could be used for the examination of inks on documents taking extremely small regions from unimportant areas of the document. They also found that FT-IS was a useful, direct, and reproducible method for distinguishing printing inks. Raman technology has been widely used in forensic analysis of questioned documents with different inks types. Zieba et al.,[11] investigated degraded papers using Infrared and Raman spectroscopy. They investigated the age of documents and their types. They used three types of paper in the experiment and then selected different spots on each paper type. Their approach investigated the possibility of distinguishing the ages of samples. Their proposed approach involved 2D correlation analysis on Node's method using ATR FT-IR spectra as input data for producing the correlation maps. The results showed that the pattern of 2D maps gave insight into the samples degradation mechanism, which is of interest for many researchers in the field of forensics.

Buzzini et al.,[12] studied the criteria of the discrimination of inkjet printer inks based on micro-Raman spectroscopy. Inkjet printed documents produce microscopic colored dots that can be detected using a microscopical approach coupled with Raman spectroscopy. The general aim of the study was to determine whether the Raman data collected from the three cyan, magenta, and yellow microscopic colored dots constitute, coupled, a chemical signature of enough discriminating quality to provide trustworthy investigative leads in a time effective and manner non-destructive. The criteria were necessary to achieve discriminations between inkjet printer inks from different brands, the same brands, and even the same models have been evaluated on a set of samples. Although Raman spectroscopy is already a relatively well-established method for the characterization of colorants (both dyes and pigments), the study showed that the contribution of minor peaks within Raman spectra improves the discriminating power of the technique. In another study performed by Borba et al.,[13], they proposed an approach to distinguish blue ballpoint pen inks using Raman spectroscopy and chemometrics. The research aimed to assess whether the combination of Raman spectroscopy and chemometric tools was good enough to be used in differentiating between blue ballpoint pen inks. The used methodology included 2 phases: 1) preprocessing phase that involved selecting the appropriate derivative that would prohibit baseline and fluorescence contributions, signal normalization to facilitate comparison among ink spectra, building up dataset table, and spectral channels autoscaling. 2) Data Analysis that included exploratory analysis using PCA (Principal Component Analysis) and HCA (Hierarchal Component Analysis). Besides, Ink classification analysis using PLS-DA (Partial Least Squares- Discriminant Analysis. PLS-DA derived models were able to achieve ¿97% classification rate. Baseline and fluorescence contribution can affect the spectra quality and spectral band shape; therefore, a preprocessing step was necessary to prohibit these effects.

Zieba-Palus and Kunicki [14] proposed a method to integrate the micro-FTIR spectroscopy, Raman spectroscopy, and XRF method for examining the inks of documents. The research aimed to study the ability to achieve a high distinguishing rate between ink samples by mixing destructive and nondestructive spectrometric methods. They performed this integration as follows: a) IR spectra were extracted from samples and compared to those extracted from standard dyes from Polish ink producers. The comparison showed qualitative similarity. b) Raman spectra were performed on the samples. Ten measurements were performed for several (5–7) points on a line formed by the ink. c) Three X-Ray fluorescence measurements were carried out for each sample of the examined ink to obtain elemental composition information for each sample. The results showed that approximately 95% of blue and black inks were distinguished based on IR and Raman Spectra, whereas 90% of discrimination was achieved for the gel inks samples examined with IR & Raman methods only. The discrimination power can be increased if elemental composition information were considered.

B. Problem Statement

According to the literature, it can be inferred that even with the use of sophisticated technologies (e.g., LIBS and Raman), forgery detection is not an easy task. Most of the proposed methods need complex computations and advanced statistical analysis tools to achieve a high forgery detection rate. Therefore, the literature has limitations when it comes to the complexity of the developed approaches. Table I summarizes the aforementioned works along with their limitations.

The contribution in this work is developing a novel network-based approach for forgery detection of documents. The proposed approach is considered simple and easy to implement compared to the literature and does not require complex computations. Also, the approach does not need experts to perform the detection since it is developed to be semi-automated.

The article is divided into the following sections; the rest of this section of this section includes the literature review and the limitations of the literature as well as the contribution of this work. Section 2 presents the proposed method including the dataset used, the configurations of LIBS technology, and the details of the proposed method. Section 3 includes the experimental results and their discussions as well as the limitations. Finally, Section 4 concludes the work.

2. Research Method

This section provides a description about the documents that were used in testing the proposed approach as well as the configurations of the LIBS technology. Also, it provides the details of the proposed method.

A. *Tested Documents and LIBS Configurations*

The samples we used in this work consisted of two sets. The first set represented the original documents and the second represented the questioned documents. The first had 9 samples that were printed using inkjet, laser printers, and photocopiers. All documents in this set were printed using the same type of paper. For every printer considered, three square boxes (5cm x 5cm) filled with black ink were printed on the white A4 office paper (COPY & LASER, 80 gsm). Table 2 presents a description of the brands of the printers involved, their models, and their types. The second set consisted of prints (in the form of official letters) that were printed using the same printers mentioned in Table 2 but on different types of papers (see Table 3).

Then, we extracted the spectrums of documents' matters aiming to detect forged documents. There are many technologies that can be used for this specific purpose such as IR, Raman, and LIBS. In fact, we used LIBS due to its accuracy and efficiency in extracting matters' spectroscopy under variety of settings. The setting of the LIBS device was configured as follows: The plasma was generated by a Q-switch Nd: YAG laser, emitting $1064nm$ with $10ns$ pulse duration. The measurements were performed using laser pulse energies of $80mJ$. A converging lens of a focal length of $100mm$ was used to focus the laser beam onto the sample surface. The target was placed in the sample stand, the distance between the focusing lens and the sample was $10cm$. The optical fiber was adjusted at an angle of 45 with the beam axis positioned at a distance of $5cm$ from the sample. The emitted light from the laser-induced plasma was collected by a collimator lens and focuses on the optical fiber aperture, diameter $(200m/0.22NA)$. The LIBS was supported by a Visual Spectra 2.1 program to record the spectrum in the PC. The spectrum was recorded between 200 and $900nm$ wavelengths range with resolution $0.8nm$. Now, for each sample, 5 independent LIBS Spectra were acquired. Each independent spectrum was recorded in an Excel file, which consisted of two columns. The first was dedicated to the wavelengths (within a range of 200 to$900nm$) and the second column was for the intensity (a range of 2048 spectral points).

B. *The Proposed Approach*

After obtaining the spectrums of the samples using LIBS technology, a correlation matrix was constructed among the spectrums of the original and questioned documents and also among the spectrums of each document. The reason behind this process was to measure the strength of the correlations between the obtained spectra for both documents. The correlation matrix was calculated using the following formula [21]:

$$C_{a,b} = \frac{\sum_{i=1}^{n} a_i b_i}{\sqrt{(\sum_{i=1}^{n} a_i^2)(\sum_{i=1}^{n} a_i^2)}} \quad (1)$$

where a and b refer to the LIBS spectra of samples, a_i and b_i are their i^th spectral component and represent the number of points spectral in the LIBS spectra of samples. Their values ranged from 1 to n, wherein our experiments were $2,048$. Each element in the generated matrix represented the correlation of a pair of spectrums. In our case, the correlation values were greater than zero, which meant that the correlations of the spectrums of the samples were $non-zero$. The resulting correlation matrix was symmetric, where the values above the diagonal were the same as under the diagonal values and the values of diagonal were always 1. The dimensions of the resulting matrix were 10 x 10 since 5 spectrums were obtained from both the original and questioned documents. After that, we transformed the correlation matrix into an adjacency matrix under the concepts of graph theory. In this context, the individual LIBS spectrums represented the vertices (V) of the network, and the correlation value between two vertices represented the weight (w) of the edge (E) between both. In practice, a graph $G_w(V, E)$ was constructed based on the values under the diagonal in the correlation matrix. The generated graph was considered to be a Completed Graph since all the nodes are interconnected. The resulting graph was undirected and weighted for all the networks generated in this work. The characteristics of such a graph can be summarized as follows [22][23]:

- The number of edges is $(\frac{n(n-1)}{2})$ where n is the number of nodes.
- The average degree of nodes is $n-1$.
- The diameter of the network is 0 when $n \leq 1$ and 1, Otherwise
- The average clustering coefficient of the network is 1.
- The density of the network is 1.
- The average path length of the network is 1.
- - An edge weight varies based on the correlation between two spectrums.

Based on the aforementioned, we generated two files;

TABLE I. Summarizing the spectroscopy-based Forgery Detections Techniques

Author	Document	Methodology	Limitations
Cicconi et al.,[9]	Documents with Signatures	LIBS spectroscopy	Needs complex computations
Ameh and Ozovehe [10]	Printed Documents	FT-IS	Needs Complex Computations
Zieba et al.,[11]	Degraded Papers	Infrared and Raman	Complex to implement due to 2D correlations
Buzzini et al.,[12]	Inkjet Printer Inks	Micro-Raman	High computational cost due to several technologies involved.
Borba et al.,[13]	Blue Ballpoint Pen Inks	PCA, HCA, and PLS-DA	Needs complex computations
Zieba and Kunicki [14]	Several Types of Inks	Micro-FTIR, Raman, and XRF	Lead to the destruction of the examined document.
Other similar methods in the literature			
Lennard et al.,[15]	Laser and Inkjet Prints	Analysis of LIBS Spectroscopy	Complex computations required.
Elsherbiny Nassef [16]	Laser and Inkjet Prints	Analysis of LIBS Spectroscopy	
Verma et al.,[17]	Laser and Inkjet Prints	Analysis of LIBS Spectroscopy	
Gal et al.,[18]	Laser and Inkjet Prints	Analysis of LIBS Spectroscopy	
Udristioiu et al.,[19]	Different Types of Papers	Micro-Raman and FT-IR	
Raza and Saha [20]	Stamp Pad Inks	Raman and HPTLC	

TABLE II. Printers used

#	Printer Type	Brand	Model
1	Laser	Canon	i-SENSYS MF231
2	Laser	Canon	i-SENSYS MF4010
3	Laser	Canon	i-SENSYS LBP6000
4	Laser	Canon	Image CLASS MF264dw
5	Inkjet	Epson	EcoTank ITS L3070
6	Inkjet	Canon	Pixma TS6020
7	Inkjet	H P	Page Wide Pro 577dw
8	Laser	Ricoh	Aficio MP 4001
9	Laser	Ricoh	Aficio MP C2051

TABLE III. Paper types used

#	Paper Brand	Model
1	COPY & LASER	Indonesia
2	Ballet Universal	China
3	PAPEROne	India
4	PAPERLine	Indonesia
5	PPLITE	India

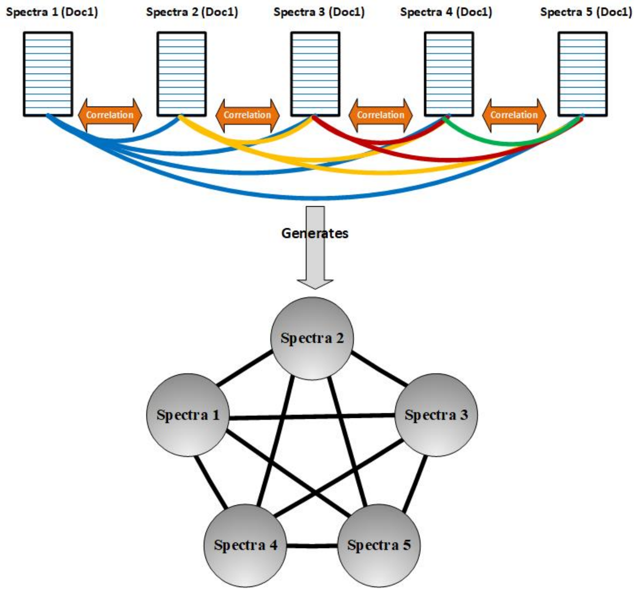

Figure 1. Network generation

node file and edge file using MATLAB programming. In practice, an Adjacency matrix is converted into two spreadsheets (.csv) files, one for the node table, where each node is assigned a unique id. The other file is for the edge table, where all relations among nodes are expressed as relations between Ids.

- **ID:** The identifiers of the spectrums that are extracted using LIBS.
- **Label:** The label of the spectrums, which was used in the visualization.

The edges table mainly contains the following columns:

- **Source, Target:** The IDs of the nodes that are connected to each other.
- **Weight:** The value of the weights of the edges in the network.

Thereafter, we used Gephi software to generate and visualize the network of each test considered in this work. Gephi is a common software tool that can be used in visualizing data in a form of nodes and edges. The tool also enables us to perform network calculations in nodes level and network level (e.g., modularity). Moreover, Gephi can be used to implement different clustering algorithms and incorporate the results within the network model. This feature support producing profisional visualization of data. Therefore, these features are useful in this work to be aceived. The input of Gephi is a dataset of nodes and edges. Figure 1 demonstrates how the spectrums are represented as nodes and edges and how a network is generated.

Our next step was to extract the clusters in the generated network. The detection strategy we followed was based on the number of extracted clusters from each of the generated networks. In this regard, we selected two fast algorithms for clustering namely, Louvain and Leiden clustering algorithms. The Louvain Algorithm was proposed by Blondel et al.,[24] and has the ability to find high modularity partitions of a weighted network efficiently. It is based on two phases, the first one is the local moving of nodes, while the second phase is the aggregation of the network (inserting nodes to communities). The algorithm is a greedy-based optimization that appears to run-in-time $O(v.logv)$ where v is the number of network vertices [25]. The Louvain algorithm detects communities based on the optimization of the modularity (Q) level that is in the range of $(-1, 1)$. Since the generated networks of samples are weighted, the modularity is formulated as follows [26]:

$$Q = \frac{1}{2m} + \sum_{ij} [W_{ij} - \frac{k_i k_j}{2m}] \delta(c_i, c_j) \tag{2}$$

Where W_{ij} denotes the weight of the edge between the nodes i and j. The sum of the weights of edges linked to the nodes i and j are denoted by k_i and k_j, respectively. m represents the sum of all weights of the edges within the network. The communities are represented by c_i and c_j, and finally, δ is a non-negative such that $\delta(c_i, c_j) = 1$ when $c_i = c_j$ and 0, Otherwise.

The other algorithm we tested in this work was the Leiden Clustering Algorithm, which was proposed by Traag et al.,[27] and considered as an updated version of the Louvain algorithm. It is based on three phases, the first one is the local moving of nodes, then, refinement of the partition, and the third phase is the aggregation of the network based on the refined partition. However, during the experiments, we observed that the same results can be obtained using both algorithms due to the small size of our networks. Therefore, we decided to use the Louvain algorithm due to its performance in terms of speed.

Practically, the Louvain algorithm has an issue related to the resolution limit of Modularity [27]. This issue may lead to gathering the smaller groups into one big community. More precisely, modularity may hide some communities in the network, which is insufficient. Therefore, we proposed to use a combination of the Louvain and Leiden algorithms (modified version of the Louvain). In the Louvain algorithm, instead of using modularity as a quality function, we involved an alternative quality function called *Constant Potts Model (CPM)*, which was proposed in [27] and was introduced in [28]. The CPM (H) is abile to overcome the limitations in modularity and can be defined as follows [28]:

$$H = \sum_{c} [e_c - \gamma(\binom{n_c}{k})] \tag{3}$$

where c is a community whose number of nodes is n_c. γ is the resolution parameter that is quite straightforward. The density of communities should be at least γ, while among communities should be lower than γ. At this point, as proved in [27], the Louvain algorithm guarantees that the communities are well-separated and no communities can be merged. Also, the size of our networks was small in terms of the number of nodes. Also, based on the simulations performed, the updated form of the Louvain (with CPM) was a little faster than Leiden. Accordingly, we used the updated Louvain as our forgery detection algorithm in this work. The results of the Louvain algorithm were used to determine whether the questioned document was forged based on the number of clusters obtained. Hence, in our case, *if the number of clusters equals 1*, the questioned document is detected as *original*, otherwise, the document is *forged*. The general workflow of our approach is presented in Algorithm **??** and Figure 2.

It should be mentioned that our proposed approach can be applied to a single document and detect whether it is counterfeited in some parts. In this case, our approach can produce more than one cluster if any of one (or more) part of the document were counterfeited.

Algorithm 1: General workflow of the proposed approach. Given that we have two documents: Original (D_O) and Questioned (D_Q)

```
START
1....... SET LIBS Configurations.
2.......ACQUIRE 5 LIBS tests for each of $D_O$ and $D_Q$.
3.......CREATE the Correlation Matrix (CM) among the acquired spectra.
4.......CONVERT the $CM$ into Adjacency Matrix ($AM$).
5.......FORMALIZE $AM$ into a dataset of nodes and edges and create the network.
6.......APPLY some clustering algorithms on the generated network.
7.......IF No. of Clusters = 1 THEN
............. $D_Q$ is also ORIGINAL
.........ELSE $D_Q$ is FORGED
END
```

3. Results and Discussions

Many experiments were performed on the sample documents using the proposed approach for forgery and counterfeit detection. For clarity purposes, Table IV presents information (e.g., printer types, paper types, and ink types) about the experiments performed under the proposed approach.

A. *Group 1: Testing different ink types with same paper types*

Two documents with different types of inks and the same paper type were tested. The first was the original document (O_D) and the second was the questioned document (Q_D). The original and questioned document were printed with two different printers (Laser printer Canon mf231 and inkjet printer Epson 3070 respectively) on the same type of paper (A4 COPY & LASER). Figure 3 shows the visualization of the spectrums of the inks used in both documents. The visualization shows that two clusters were obtained after applying the modified Louvain algorithm, which was able

TABLE IV. Appendix: Summary of the experiments and tests performed in this work.

Group	Samples	Printer Type	Paper Type
Group 1	Original and Questioned	Laser Canon mf231 and Inkjet Epson 3070	COPY& LASER
Group 1	Original and Questioned	Laser Canon mf231 and Inkjet Ricoh 4001	COPY& LASER
Group 2	Original and Questioned	Laser Canon mf264 and Inkjet Canon 6020	COPY& LASER, PAPERLINE
Group 3	Original and Questioned	Inkjet Canon 6020	COPY& LASER
Group 4	Questioned Document	Laser Canon mf264 and Inkjet HP577	COPY& LASER
Group 5	1 laser print and 2 inkjet prints	see Table 4	COPY& LASER
Group 6	3 laser prints and 2 inkjet prints	see Table 4 + Laser Canon 264	COPY& LASER
Group 6	3 laser prints and 2 inkjet prints	see Table 4 + inkjet HP577	COPY& LASER
Group 7	2 original and 2 questioned documents	...	PaperOne, COPY& LASER, Ballet Universal

to distinguish two different documents. As a result, the questioned document is forged with an accuracy of 100%. All the original document spectra appeared in the first cluster marked in red, wholly separated from the second cluster marked in green, which included all the spectra of the questioned document.

The second test in Group 1 was performed on two documents (original and questioned). They were tested using two different printers (Laser printer Canon mf231 and Photocopier Ricoh 4001) and the same paper type (paper A4 COPY & LASER). Figure 4 shows the visualization of the ink spectrums used in producing the two documents. Two clusters were obtained after applying the modified Louvain algorithm. The proposed approach could distinguish between different ink types used for printing two-document and detect a forgery case with an accuracy of 100 %.

B. Group 2: Documents with different ink types and different paper types

The proposed approach was also tested using another two samples, the first sample representing the original document was printed using a laser printer Canon mf264 on paper from type COPY & LASER. The second sample representing the suspect document was printed using an inkjet printer Canon 6020, on paper from type PAPERLINE. The result showed that the ink spectrums of the two samples were separated into two clusters as shown in Figure 5. This result also confirms the robustness of the proposed approach in detecting the forgery with an accuracy of 100% on paper from the same type and a different type.

C. Group 3: The questioned document is original

When applying the proposed method to the inks spectrums of the original document (O_D) and the questioned document (Q_D), and the appearance of a single cluster that combines the spectra of the two samples, In this case, that the questioned document is original and not forged. This means the questioned document was printed from the same printer that appears with it in the same cluster. in this test, the tested documents printed using an inkjet printer (Canon 6020) on paper type (COPY & LASER). Figure 6 depicts the appearance of the ink spectrums of two documents in one cluster with close levels of weights.

TABLE V. The printers used in the Group 5 test.

Type	Brand	Model
Laser	Canon	i-SENSYS MF231
Laser	Canon	i-SENSYS MF4010
Inkjet	Epson	EcoTank ITS L3070
Inkjet	Canon	Pixma TS6020
Laser	Ricoh	Aficio MP 4001

D. Group 4: Single document test

In this kind of test, we use a single document that is questioned in some parts (e.g., manipulation performed on the original document). The proposed approach should generate a network of one cluster if and only if the document has not been manipulated, otherwise the document is counterfeited and the network should include more than one cluster. The test was performed using a sample that was printed using laser printer canon264 on a paper type of COPY & LASER. This sample has been manipulated in a particular place through adding sensitive information by an inkjet printer HP577. Figure 7 shows the resulted network with two clusters. The first cluster includes four red nodes, which represent the original documents. The second cluster includes one green node that reflects the part of the document that has been counterfeited. This means the original document has been forged in some parts.

E. Group 5: Laser vs. inkjet samples

Laser against inkjet samples was tested using the proposed approach. In this test, the ink spectrums of three laser printers and two inkjet printers were collected in one network. The resulted network showed two clusters (see Figure 8), the first included all the ink spectrums of the laser printers and the second included all the ink spectrums of the inkjet printers. Table V presents the printer brands and models used in this test.

F. Group 6: Identifying ink types of the questioned documents

Knowing the type of ink that was used for printing a questioned document can provide important forensic evidence for investigators. A test was carried out by adding ink spectra of a questioned document was printed using a laser printer (Canon 264) to inks spectra of printers mentioned in Table V in one network. The test result was the appearance

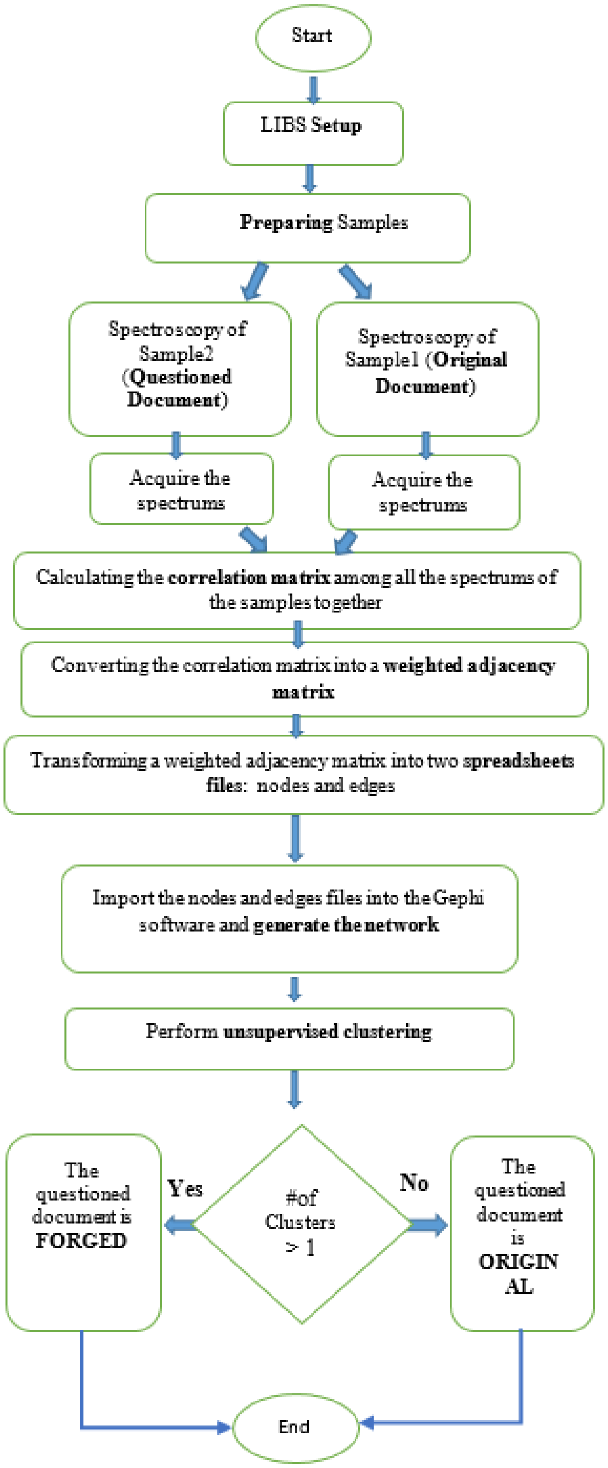

Figure 2. Workflow of the proposed approach when having two documents original and questioned.

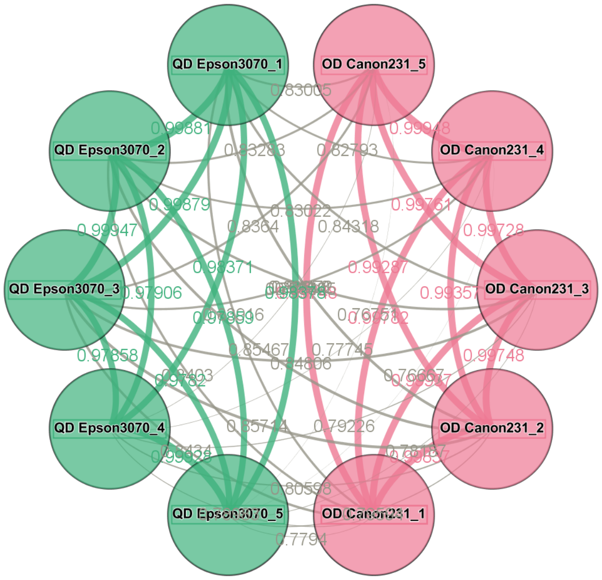

Figure 3. Group 1 Test 1: Visualization of ink spectrums of the original and questioned document produced using two different printers (Laser printer Canon mf231 and inkjet printer Epson 3070) on papers of the same type (paper A4 COPY & LASER).

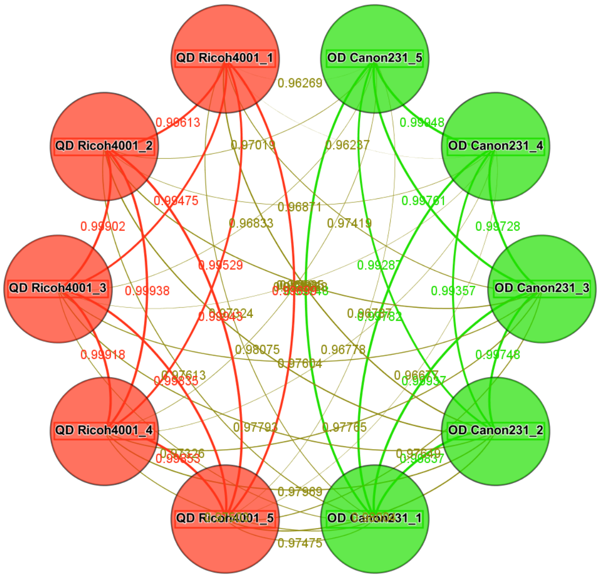

Figure 4. Group 1 Test 2: Visualization of ink spectrums of the original and questioned documents that produced using two different printers (Laser printer Canon mf231 and Photocopier Ricoh 4001) and same paper type (paper A4 COPY & LASER).

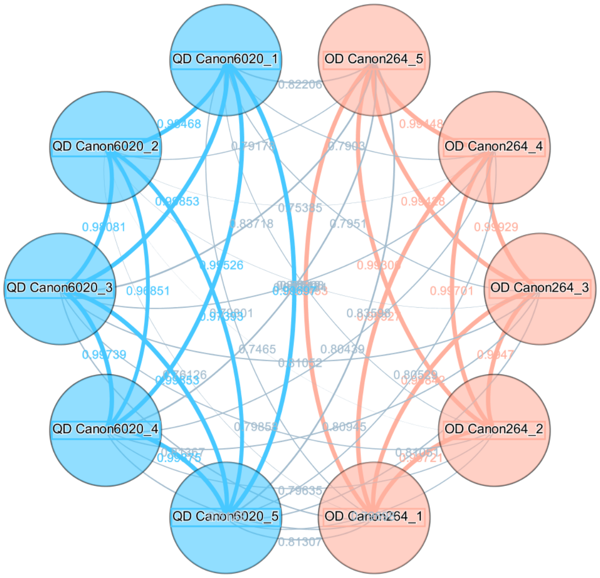

Figure 5. Group 2: Visualization of ink spectrums of the original and questioned documents that printed using two different printers; Laser printer Canon mf264 and inkjet printer Canon 6020, and papers of a different type; COPY & LASER and PAPERLINE respectively.

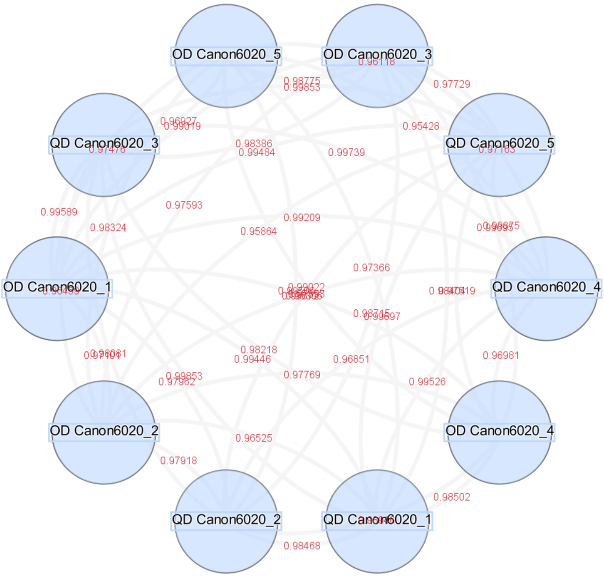

Figure 6. Group 3: Visualization of ink spectrums of the original and questioned document, that printed using a laser printer (Canon MF 231) on paper (COPY & LASER).

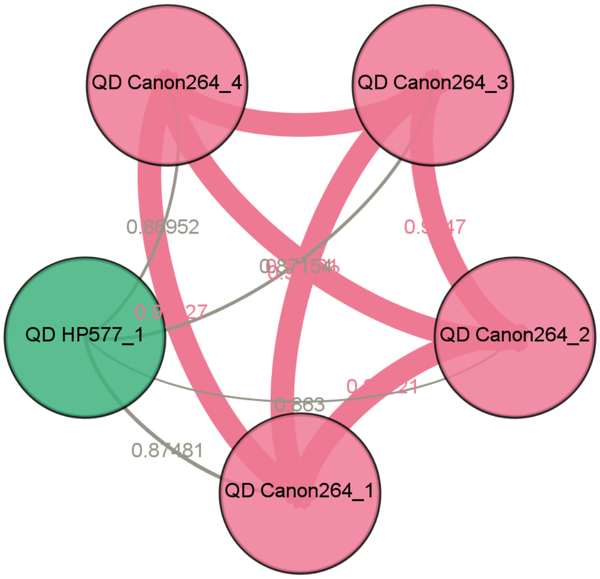

Figure 7. Group 4: Visualization of the ink spectrums network for a document with questionable parts.

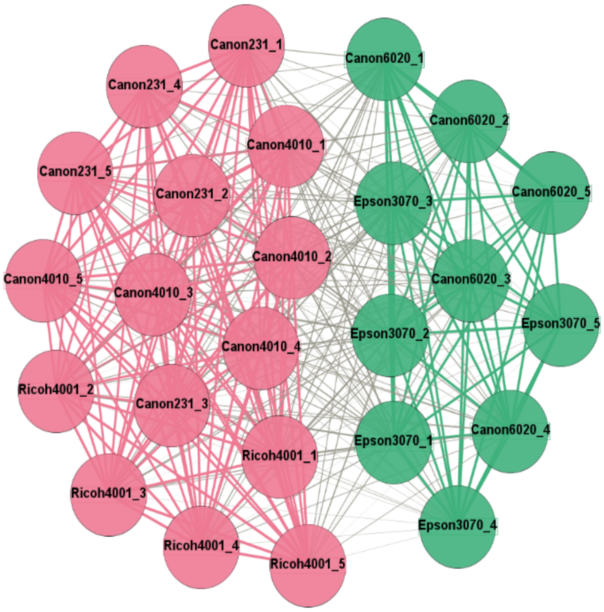

Figure 8. Group 5: Visualization of ink spectrums of five printers. The network is separated printers into two clusters. The first includes inks for laser printers and the second includes inks for inkjet printers.

Figure 9. Group 6 Test 1: Visualization of the questioned document's spectrums that was printed using the laser printer (Canon 264). The spectrums were clustered with the cluster of the same type of printing.

of the spectra of the laser printer Canon 264 within the cluster of laser printers, as shown in Figure 9. Another test was performed using a different sample (inkjet). Figure 10 shows the visualization of the spectrums that were printed using the inkjet printer (HP 577), which belongs to the inkjet cluster.

G. Group 7: Testing different paper types

In addition to the previous tests, we examined documents with different paper types. The goal of this test was to test whether the questioned document was printed using the same paper type as the original document. Our approach was successfully distinguished between papers of different brands. The pair types of paper considered in this test were (PAPEROne and COPY & LASER), (Ballet Universal and COPY & LASER). The results are shown in Figures 11 and 12.

H. Group 8: Forgery detection using other clustering algorithms

The proposed approach was examined using other clustering algorithms such as K-Means clustering (K-Means) [29], Gaussian Mixture Model (GMM) [30], Affinity Propagation (AP) [31], Agglomerative Hierarchy (AH) [32], Spectral Clustering (SC) [33]. The strategy followed in performing the benchmarking was based on using all the algorithms with all the tests considered in this thesis and calculate the average for each algorithm. This average is considered the overall performance of the algorithms involved.

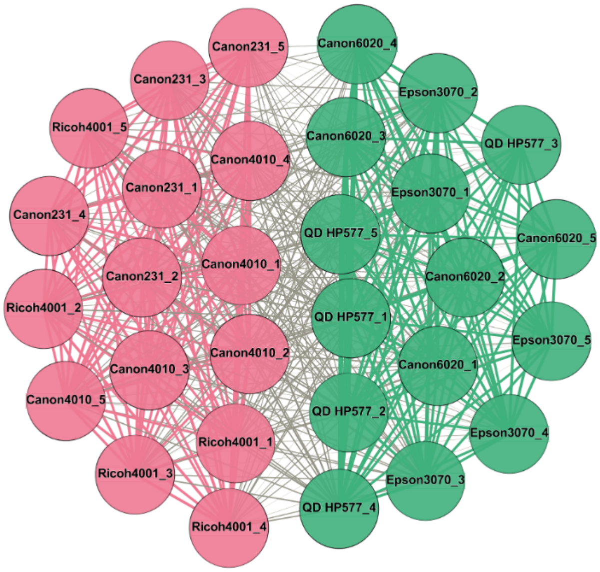

Figure 10. Group 6 Test 2: Visualization of the questioned document's spectrums that was printed using the inkjet printer (HP 577). The spectrums were clustered with the cluster of the same type of printing.

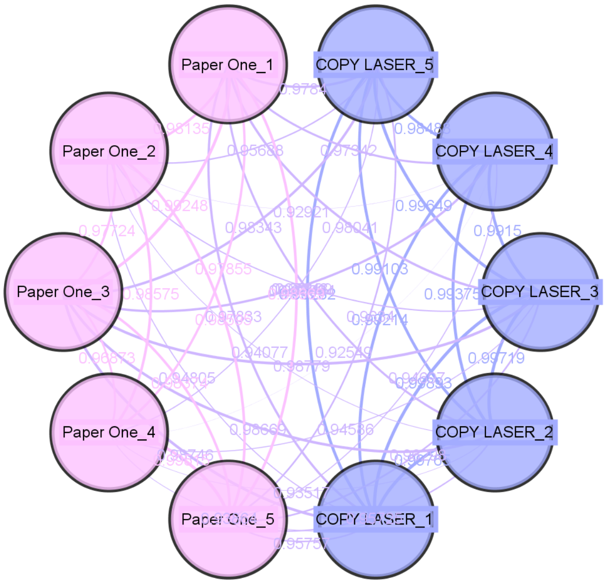

Figure 11. Group 7 Test 1: The proposed approach distinguished between two types of paper (PAPEROne and COPY & LASER), which appeared in two clusters.

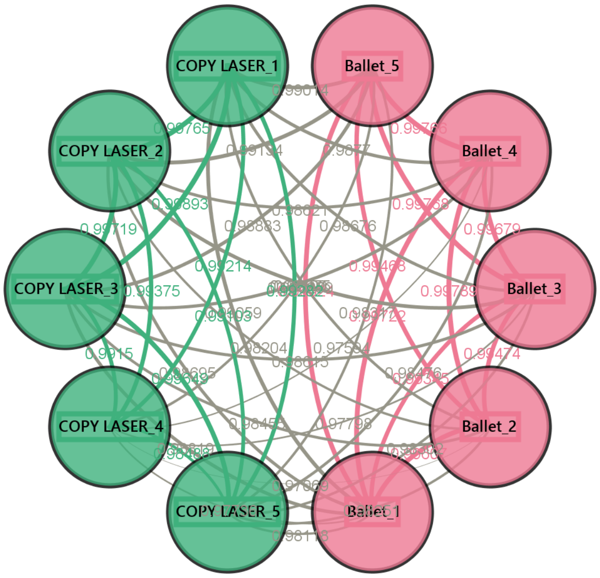

Figure 12. Group 7 Test 2: The proposed approach distinguished between two types of paper (Ballet Universal and COPY & LASER), which appear in two clusters.

TABLE VI. the performance of the proposed algorithm compared to the benchmarking algorithms.

Algorithm	Detection Rate (%)
K-Means [29]	67.53%
GMM [30]	80.61%
AP []31t	82.18%
AH [32]	87.75%
SC [33]	90.42%
LV [24]	95.8%
LD [28]	94.2%
Updated Lovain	100.00%

As can be observed in Table VI, the performance of the modified version of the Louvain algorithm was always successful in providing 100% of detection. This means the proposed approach has a strong ability to distinguish forged documents. These results confirm the efficiency, reliability, and stability of the proposed algorithm.

I. The limitations of the proposed approach

The proposed approach has few limitations that should be taken into considerations as follows:

- The number of LIBS tests (number of scans) should not be less than five. According to the experiments, in our work, five scans were considered as a threshold. Therefore, the accuracy of our approach is expected to decrease when the number of scans becomes below the threshold.
- The configurations of LIBS should be as described in Section 2-A. According to the pre-experiments performed, changing these configurations will affect the accuracy of our approach.

4. CONCLUSIONS

In this work, we propose a simple, fast, and efficient approach for the forgery detection of documents. The proposed approach uses concepts inspired from complex networks. For each sample test considered in this work, we generated a network based on the spectrums of that document. In all the generated networks, the nodes were the spectrums of the documents that were extracted using LIBS technology. The edges represented the correlations among the spectrums. Our approach was examined using several sample tests including documents printed by laser printers and inkjet printers. The approach was also tested under different kinds of printing papers. Our approach depended on the number of generated clusters within the network using a modified Louvain Clustering Algorithm. The proposed approach always provided 100% of accuracy in detecting forged documents.

As future works, it is planned to include more samples that are different from the ones considered in this work (e.g., special official documents). In addition to detecting forgery of documents containing writing inks and stamp ink.

ACKNOWLEDGMENT

We are grateful to the Software and Computer Science departments/University of Mosul for making this work achieved. We also would like to thank the departments of Software and Computer Science for providing us with all the possible support in performing this research.

NOMENCLATURE

LIBS Laser-Induced Breakdown Spectroscopy
FT-IS Fourier-transform infrared spectroscopy
QD Questioned documents
OD Original Document

REFERENCES

[1] R. A. Khan and S. A. Lone, "A comprehensive study of document security system, open issues and challenges," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 7039–7061, 2021.

[2] A. G. Dyer, B. Found, and D. Rogers, "An insight into forensic document examiner expertise for discriminating between forged and disguised signatures," *Journal of Forensic Sciences*, vol. 53, no. 5, pp. 1154–1159, 2008.

[3] A. Parkinson, M. Colella, and T. Evans, "The development and evaluation of radiological decontamination procedures for documents, document inks, and latent fingermarks on porous surfaces," *Journal of Forensic Sciences*, vol. 55, no. 3, pp. 728–734, 2010.

[4] L. Valderrama, P. H. Marcco, and P. Valderrama, "Model precision in partial least squares with discriminant analysis: A case study in document forgery through crossing lines," *Journal of Chemometrics*, vol. 34, no. 12, p. e3265, 2020.

[5] N. B. Abd Warif, A. W. A. Wahab, M. Y. I. Idris, R. Ramli, R. Salleh, S. Shamshirband, and K.-K. R. Choo, “Copy-move forgery detection: survey, challenges and future directions,” *Journal of Network and Computer Applications*, vol. 75, pp. 259–278, 2016.

[6] P. Niu, C. Wang, W. Chen, H. Yang, and X. Wang, “Fast and effective keypoint-based image copy-move forgery detection using complex-valued moment invariants,” *Journal of Visual Communication and Image Representation*, vol. 77, p. 103068, 2021.

[7] J. Laserna, J. M. Vadillo, and P. Purohit, “Laser-induced breakdown spectroscopy (libs): fast, effective, and agile leading edge analytical technology,” *Applied spectroscopy*, vol. 72, no. 1_suppl, pp. 35–50, 2018.

[8] M. Markiewicz-Keszycka, X. Cama-Moncunill, M. P. Casado-Gavalda, Y. Dixit, R. Cama-Moncunill, P. J. Cullen, and C. Sullivan, “Laser-induced breakdown spectroscopy (libs) for food analysis: A review,” *Trends in food science & technology*, vol. 65, pp. 80–93, 2017.

[9] F. Cicconi, V. Lazic, A. Palucci, A. C. Almeida Assis, and F. Saverio Romolo, “Forensic analysis of commercial inks by laser-induced breakdown spectroscopy (libs),” *Sensors*, vol. 20, no. 13, p. 3744, 2020.

[10] P. O. Ameh and M. S. Ozovehe, “Forensic examination of inks extracted from printed documents using fourier transform infrared spectroscopy,” *Edelweiss Applied Science and Technology*, vol. 2, no. 1, pp. 10–17, 2018.

[11] J. Zikeba-Palus, A. Weselucha-Birczynska, B. Trzcinska, R. Kowalski, and P. Moskal, “Analysis of degraded papers by infrared and raman spectroscopy for forensic purposes,” *Journal of molecular structure*, vol. 1140, pp. 154–162, 2017.

[12] P. Buzzini, C. Polston, and M. Schackmuth, “On the criteria for the discrimination of inkjet printer inks using micro-r aman spectroscopy,” *Journal of Raman Spectroscopy*, vol. 49, no. 11, pp. 1791–1801, 2018.

[13] F. d. S. L. Borba, R. S. Honorato, and A. de Juan, “Use of raman spectroscopy and chemometrics to distinguish blue ballpoint pen inks,” *Forensic science international*, vol. 249, pp. 73–82, 2015.

[14] J. Zikeba-Palus and M. Kunicki, “Application of the micro-ftir spectroscopy, raman spectroscopy and xrf method examination of inks,” *Forensic science international*, vol. 158, no. 2-3, pp. 164–172, 2006.

[15] C. Lennard, M. M. El-Deftar, and J. Robertson, “Forensic application of laser-induced breakdown spectroscopy for the discrimination of questioned documents,” *Forensic science international*, vol. 254, pp. 68–79, 2015.

[16] N. Elsherbiny and O. A. Nassef, “Wavelength dependence of laser induced breakdown spectroscopy (libs) on questioned document investigation,” *Science & Justice*, vol. 55, no. 4, pp. 254–263, 2015.

[17] N. Verma, R. Kumar, and V. Sharma, “Analysis of laser printer and photocopier toners by spectral properties and chemometrics,” *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 196, pp. 40–48, 2018.

[18] L. Gál, M. Belovičová, M. Oravec, M. Palková, and M. Čeppan, “Analysis of laser and inkjet prints using spectroscopic methods for forensic identification of questioned documents,” 2013.

[19] Y. Aboul-Enein, İ. G. TANASE, A. A. BUNACIU, and F. M. UDRIŞTIOIU, “Application of micro-raman and ft-ir spectroscopy in forensic analysis of questioned documents,” *Gazi University Journal of Science*, vol. 25, no. 2, pp. 371–375, 2012.

[20] A. Raza and B. Saha, “Application of raman spectroscopy in forensic investigation of questioned documents involving stamp inks,” *Science & Justice*, vol. 53, no. 3, pp. 332–338, 2013.

[21] J. H. Steiger, “Tests for comparing elements of a correlation matrix.” *Psychological bulletin*, vol. 87, no. 2, p. 245, 1980.

[22] R. Albert and A.-L. Barabási, “Statistical mechanics of complex networks,” *Reviews of modern physics*, vol. 74, no. 1, p. 47, 2002.

[23] A.-L. Barabási, “Network science,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 371, no. 1987, p. 20120375, 2013.

[24] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.

[25] A. Lancichinetti and S. Fortunato, “Community detection algorithms: a comparative analysis,” *Physical review E*, vol. 80, no. 5, p. 056117, 2009.

[26] M. E. Newman, “Analysis of weighted networks,” *Physical review E*, vol. 70, no. 5, p. 056131, 2004.

[27] V. A. Traag, L. Waltman, and N. J. Van Eck, “From louvain to leiden: guaranteeing well-connected communities,” *Scientific reports*, vol. 9, no. 1, pp. 1–12, 2019.

[28] V. A. Traag, P. Van Dooren, and Y. Nesterov, “Narrow scope for resolution-limit-free community detection,” *Physical Review E*, vol. 84, no. 1, p. 016114, 2011.

[29] A. Likas, N. Vlassis, and J. J. Verbeek, “The global k-means clustering algorithm,” *Pattern recognition*, vol. 36, no. 2, pp. 451–461, 2003.

[30] D. A. Reynolds, “Gaussian mixture models.” *Encyclopedia of biometrics*, vol. 741, pp. 659–663, 2009.

[31] K. Wang, J. Zhang, D. Li, X. Zhang, and T. Guo, “Adaptive affinity propagation clustering,” *arXiv preprint arXiv:0805.1096*, 2008.

[32] A. Bouguettaya, Q. Yu, X. Liu, X. Zhou, and A. Song, “Efficient agglomerative hierarchical clustering,” *Expert Systems with Applications*, vol. 42, no. 5, pp. 2785–2797, 2015.

[33] J. Liu and J. Han, “Spectral clustering,” in *Data Clustering*. Chapman and Hall/CRC, 2018, pp. 177–200.

Alaa Amged received his Bachelor degree in Computer Science in 2005 from the University of Mosul. He is currently a Masters students at the department of Computer Science/College of Computer Science and Mathematics/ University of Mosul/IRAQ.

Basim Mahmood received his M.Sc. degree in Computer Science from the University of Mosul in 2009. His Ph.D. degree was in 2015 from the college of Engineering at Florida Institute of Technology, USA. He is working as a faculty member at the Computer Science Dept., University of Mosul, Iraq. He also works as a researcher in the BioComplex Laboratory, Exeter, UK. His current research interests include Complex Networks, Big Data Analysis and Data Mining.

Khalid A.K. Almukhtar received his Ph.D. degree Criminal Law from the College of Law/ University of Mosul in 2013. His Masters degree was was also in Law from the same college. He is currently a researchers in the field of forensics and its integration with technology. His current research interests include Digital Forensics and Crime Science.