Propose Data Mining System to Advance E-Learning Over Online Social Network (Facebook)

Dr. Soukaena Hassan Hashem
Computer Science Department, University of Technology/ Baghdad
Sarraa Mowaffaq Abood
Computer Science Department, University of Technology/ Baghdad
Email:sarra.moafak@yahoo.com.

Received on: 23/3/2014 & Accepted on: 5/3/2015

ABSTRACT

This research presents a proposal to advance e-learning over online social network, facebook, through analyzing the structure of this network and the behavior of their users. This proposal will construct facebook group for Iraqi postgraduate higher education computer sciences students (IPHECSS), this group consist of 300 users.

The Proposal has four consequence steps to advance the e-learning over facebook, these steps are:

1. Constructing a proposed student's facebooks dataset for Iraq students' society called Iraqi postgraduate higher education students (IPHES), which contains self-defined characteristics of a student's facebooks.
2. Applying customized Frequent Pattern (FP-growth) Association Rule (AR) technique to IPHES dataset as a ranker (since it calculates the frequency of attributes) and mining technique (since it extracts knowledge to predict decision making to support e-learning over facebook through analyzing student's behavior).
3. Applying Traditional k-mean and proposed Modified k-mean techniques to IPHES dataset to advance the traditional KM in clustering the students to introduce the structure of network's users; this helps in supporting e-learning over facebok through analyzing students broadcasting and activities. Modification on k-mean is done by injecting a preprocessing substep in traditional KM called attributes weighting depending on ranking results obtained by applying AR as a ranker and modifying Euclidian distance similarity measure to result vectors instead of single value.
4. Analyzing the results of both association rules and clustering using excel2007 and UCINET software.

Keyword: facebook, association rules, k-mean, attributes ranking

مقترح نظام التنقيب لتحسين التعليم خلال الشبكات الاجتماعية (الفيسبوك)

الخلاصة

هذا البحث يقدم مقترح لتطوير التعليم الالكتروني خلال الشبكات الاجتماعية, الفيسبوك, من خلال تحليل هيكلية الشبكة وسلوك مستخدميها.

في هذا المقترح سوف نبني بالفيسبوك مجموعة من التعليم العالي لطلاب الدراسات العليا لعلوم الحاسوب ,هذه المجموعة تتكون من 100 طالب.

المقترح يتكون من اربع خطوات متسلسلة لتطوير التعليم خلال الفيسبوك, هذه الخطوات هي :

اول خطوة للمقترح هي بناء وتجهيز مجموعة البيانات الذي يحتوي على بيانات شخصية للطالب بما في ذلك متغيرات مثل الجامعة, الجنس, سنة التخرج, المتابعة,المتابعين, الفعالية,التواجد وعدم التواجد على الانترنت, الوظيفة, محل الاقامة, المجاميع, اختصاص الطالب, سنة التخرج,الاصدقاء المشتركين.

الخطوة الثانية تقترح لتطبيق قاعدة الربط على مجموعة البيانات باعتباره تقنية ضابط مراتب وتعدين لاستخراج انماط متكررة (السمات الهامة مع النظر في العلاقات المتبادلة مع بعضها البعض) ولاستخراج المعرفة للتنبؤ باتخاذ القرار لدعم التعليم الالكتروني.

لتطوير الخوارزمية التقليدية بتجميع الطلاب بشكل اكفأ k-meanثالث خطوة تقترح استخدام خوارزمية وذلك بضخ خطوة فرعية معالجة مسبقا فيه تسمى اوزان الصفات تعتمد على النتائج التي تم الحصول عليها في نتائج الترتيب التي حصلنا عليها من الخطوة الثانية.

لتطوير التعليم الالكتروني خلال الفيسبوك AR و proposed k_mean واخيرا, الخطوة الرابعة تحلل نتائج كلا للتنبؤ بمعلومات جديدة لم تكتشف من قبل UC INET باتجاهين بنية الشبكة وسلوك الطالب. ثم سوف يستخدم برنامج من خلال AR و KM .

INTRODUCTION

Computer networks are inherently social networks, linking people, organizations, and knowledge. A social network is a social structure of people, related (directly or indirectly) to each other through a common relation or interest. Facebook has become one of the largest social networking sites, comprising the sixth most trafficked site in the US (ComScore, 2008). One of the features of interest in facebook is the ability to create groups. Groups allow facebook account holders to join. An account holder can be a member of multiple groups [1]. A social network is the network of relationships and interactions among social entities such as individuals, groups of individuals, and organizations. Since the rise of Internet and the World Wide Web has enabled us to investigate large-scale social networks, there has been growing interest in social network analysis. A social network is usually formed and constructed by daily and continuous communication between people and therefore includes different relationships, such as the positions, betweenness and closeness among individuals or groups. In order to understand the social structure, social relationships and social behaviors, social network analysis therefore is an essential and important technique. Social network analysis (SNA) is the study of social networks to understand their structure and behavior. Research on social networks could be traced back to sociology, anthropology and epidemiology [2]. E-learning (also referred to as web-based education and e-teaching) is new context for education where large amounts of information describing the continuum of the teaching-learning interactions are endlessly generated and ubiquitously available [3].

Data Mining can be used to extract knowledge from e-learning systems through the analysis of the information available in the form of data generated by their users. In this case, the main objective becomes finding the patterns of system usage by teachers and students and, perhaps most importantly, discovering the students' learning behavior patterns. Several studies have demonstrated that Data Mining techniques could successfully be incorporated into E-learning environments. The application of data mining techniques and concepts in e-Learning systems helps to support educators to improve the e-Learning environment [4].

Related works

This section will outlines related works done in social network analysis, most research focused on both of structures and user’s behaviors. In [5] Krpan D., et. al., study the popularity and the use of e-learning systems which increased through last decades. Where students produce a lot of data through their interactions with the system, this data is often not exploited. They show practical experience with specific e-learning system and applied data mining technique for the analysis which served as a tool for grouping students with similar characteristics. In [6] Nancy P. et. al., they focused on the formulation of association rules using which decisions can be made for future Endeavour. And applied Apriori algorithm to 100 universities datasets of facebook has originated from Adam D’Angelo. They apply association rules between the attributes or variables and explore the association rule between a course and gender, and discover the influence of gender in studying a course. The previous research with this dataset has applied only regression models and this is the first time to apply association rules. In [7] Bonchi F., Study the success of online social networks and microblogs such as Facebook, Flickr and Twitter, the phenomenon of influence exerted by users of such platforms on other users, and how it propagates in the network, has recently attracted the interest of computer scientists, information technologists, and marketing specialists. One of the key problems in this area is the identification of influential users, by targeting whom certain desirable marketing outcomes can be achieved. In this paper, they take a data mining perspective and discuss what (and how) can be learned from the available traces of past propagations. While doing this we provide a brief overview of some recent progresses in this area and discuss some open problems. By no means, this paper must be intended as an exhaustive survey: it is instead (admittedly) a rather biased and personal perspective of the author on the topic of influence propagation in social networks. In [8] Raju E., et. al., Study of online social network structures lies on the intersection of different areas of research: sociology, graph theory and data mining. This paper studies issues around analysis of social networks using web mining techniques. Techniques and concepts of web mining and social networks analysis will be introduced and reviewed along with a discussion about how to use web mining techniques for social networks analysis. This paper also sets out a process for social network analysis using web mining.

Comparison of social networks with other networks is also studied. Discussions of the challenges and future research are also included. In [9] Falakmasir M. H., et. al., They aimed to investigate the impact of a number of e-learning activities on the students’ learning development. The results show that participation in virtual classroom sessions has the most substantial impact on the students’ final grades. This paper presents the process of applying data mining methods to the web usage records of students’ activities in a virtual learning environment. The main idea is to rank the learning activities based on their importance in order to improve students’ performance by focusing on the most important ones. In this work described the process of applying data mining methods in order to rank the students activities based on their impact on the performance of students in final exams. They used a number of ‘Feature Selection’ and ‘Attribute Evaluation’ methods together with real usage data of students picked up from the Moodle LMS in order to perform the case study.

In [10] Sael N., et. al., they benefit from a new preprocessing approach applied to Moodle platform in order to apply clustering and association rule mining techniques to analyze learners’ behaviors, to help in learning evaluation, and to enhance the structure of a given SCORM content. They adopted the feature selection process and multilevel clustering that allowed us to confirm the importance of these new data preprocessing methods and to validate the usefulness of the attributes describing the learners' interactions with the SCORM content pertaining to learners’ profiles detection. They also benefited from this approach as we sought to find possible relationships between the different parts of the relevant content and to help the teacher/ tutor to evaluate the structure of such content. In [11] Hu H. W., et.al., Study researches focus on the diffusion effect in social network. Most of these studies concentrate on identifying the key students, detecting potential groups and the predicting the users’ behaviors, but they seldom consider the impact of time.

In fact, like the prime time in TV network, the influence power of time is an important factor to determine the diffusion effect in social network. For selecting the appropriate time to spread message, not only the number of online users, but the structure of relationship and attributes of users need to be taken into account. In this study, they define three types of users according to their profiles, and propose an approach to extract the influence power patterns of time, so we can predict the proper time to announce information from historical data.

The Proposal of Data Mining System over Facebook’s E-Learning

The significance of this research is primarily of academic interest. How to advance the e-learning over facebook, to explain the proposal in details see algorithm (1) and figure (1) where both of them give the general strategy of the proposal.

Algorithm (1): The Proposed IPHE Mining System

Input: Facebook's accounts IPHES of (100-300) users.
Output: Recommendations to strength e-learning over facebook

Process

Step1: Construct the dataset of the Facebook's group IPHES.

Step2: Apply the customized Association Rule (Fp-growth) technique to the constructed dataset to extract the following:

1. 1-itemset attribute frequency which will present single attributes ranking. So these attributes will be ranked according their frequency in dataset.
2. n-itemsets attributes frequency which will present correlated attributes ranking. So these correlated attributes will be rank according their frequency in dataset.
3. Extracted association rules will be analyzed to present the prediction of student's behavior to aid in decision making for e-learning.

Step3: Apply the traditional k-mean and the modified k-mean, the modified k-mean will be implemented in the following steps:

1. Preprocessing the IPHES dataset by weighting the attributes in consistence to their importance which is known ranking and users' activities of Facebook.
2. Depending on ranking resulting from association rules, determine the important attributes that are employed to clustering IPHES in to clusters.
3. Applying k-mean as in traditional way using Euclidian Distance but with proposed vision. That is by considering each student's attributes as a weighted vector compared with others in a manner of vector (attribute by attribute), not as a single value.

Step4: After completing Rule mining (to study student's behavior) and clustering (to study student's structures), an analysis will be done to advance the e-learning over Facebook. The advances are done by predictions and recommendations. The tools used in the analysis are Microsoft Excel 2007 and UCINET.

End Process

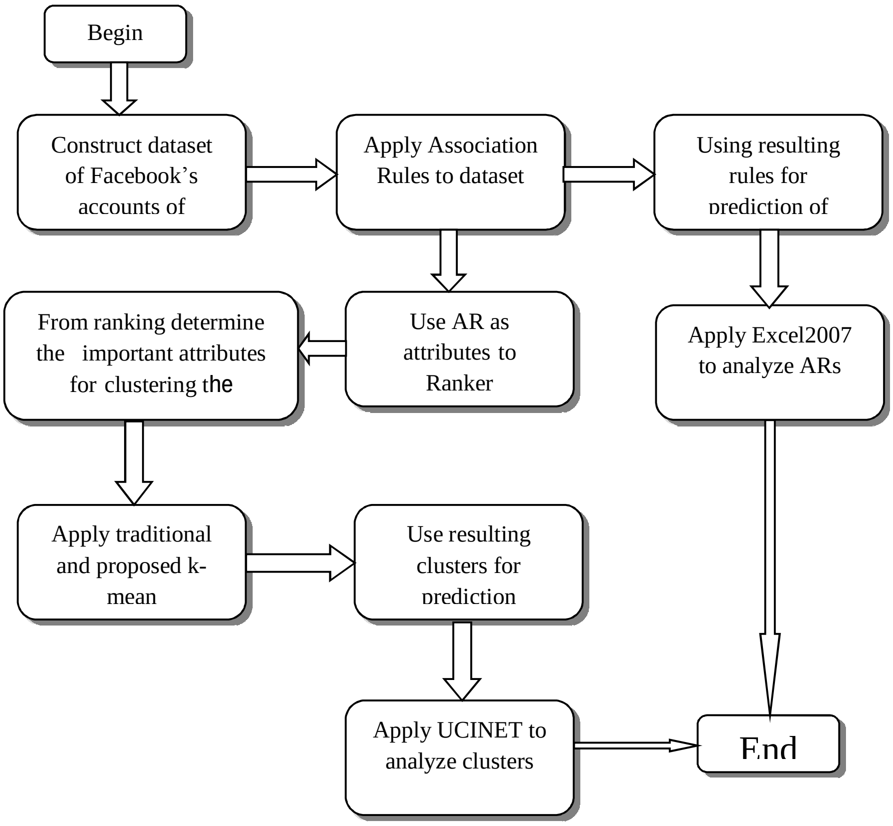
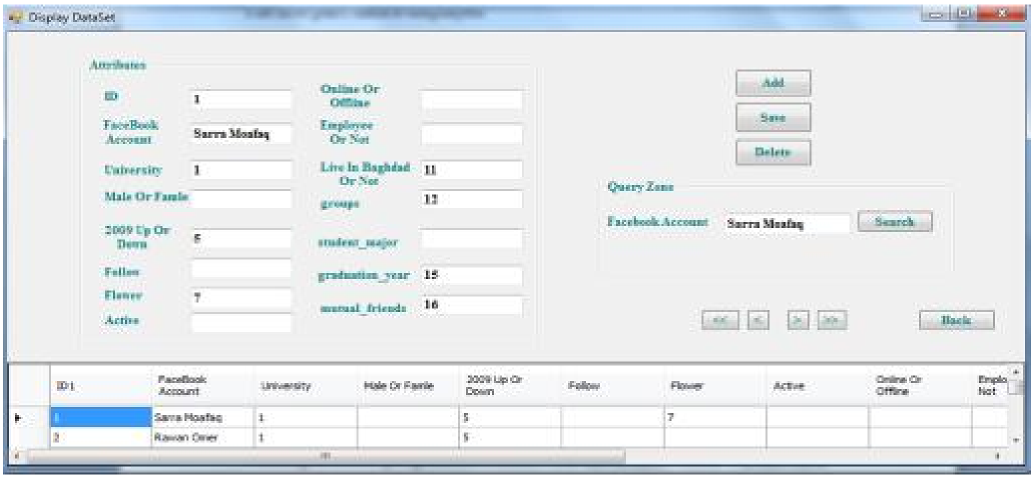

Begin
Construct dataset of Facebook's accounts of
Apply Association Rules to dataset
Using resulting rules for prediction of
Use AR as attributes to Ranker
From ranking determine the important attributes for clustering the
Apply Excel2007 to analyze ARs
Apply traditional and proposed k-mean
Use resulting clusters for prediction
Apply UCINET to analyze clusters
End

Figure (1): general diagram of the proposal

Dataset Construction

In proposal a dataset built from construct Facebook group for Iraqi postgraduate higher education computer sciences students (IPHECSS), this group consist of 300 users. It contains self-defined characteristics of a person including: facebook ID, facebook account, university, gender (female/ male), year of grad (2009 up or down), follow, follower, active, online/offline, employ, live in Baghdad or not, group, student major, graduation year, mutual friend. For reasons of Privacy (and because the data is simply not available, depending on an account's privacy settings) only the user's ID will be collected. Figure (2), explain how to build the dataset of students of IPHECSS group by entering the information of student as in proposed encoding to ensure registration the information in right manner.



Attributes

ID	1	Online Or Offline	
FaceBook Account	Sarra Moafaq	Employee Or Not	
University	1	Live In Baghdad Or Not	11
Male Or Famle		groups	12
2009 Up Or Down	5	student_major	
Follow		graduation_year	15
Flower	7	mutual_friends	16
Active			

Query Zone

Facebook Account Sarra Moafaq

ID1	FaceBook Account	University	Male Or Famle	2009 Up Or Down	Follow	Flower	Active	Online Or Offline	Emplo Not
1	Sarra Moafaq	1		5		7			
2	Rawan Omer	1		5					

Figure (2): Student information entering.

Apply Data Mining Techniques

Data mining and E-learning Aims provide taxonomy of e-learning problems to which Data Mining techniques have been applied, including, for instance: Students' classification based on their learning subjects, year of gender; detection of Irregular learning behaviors ; e-learning system navigation and interaction optimization; clustering according to similar e-learning system usage; and systems' adaptability to students' requirements and capacities. The proposal aim to:

1. Predict students' future learning behavior by creating student's models that incorporate detailed information such as students' knowledge and their Facebook's attributes.
2. Discover and improve domain models that characterize the majors to be learned and how it communicates with other attributes such as gender.
3. Study the effects of different attributes such as employment, location of living and year of graduation.
4. Analyze students' map (structure of their connectivity) for discovering point of diffusions in Facebook which aids in broadcasting the information over the group and other groups the students are member of. Such attributes are flow, flower, group, active and online/offline.

The application of data mining in e-learning systems is an iterative cycle in which the mined knowledge should enter the loop of the system and guide, facilitate and enhance learning as a whole, not only turning data into knowledge, but also filtering mined knowledge for decision making. The e-learning data mining process of the proposal consists of the following two basics steps as follows:

1. Apply customized FP-Growth AR to ranking and knowledge extraction and identification.
2. Apply traditional k-mean and the modified k-mean according ranked attributes to cluster the students of IPHES; because these clusters present the student's structure over Facebook.

Apply Customized FP-Growth Association Rules Technique

A formal statement of the association rule problem is describe in the following: Let I = $\{I_1, I_2, \ldots, I_m\}$ be a set of m distinct attributes of facebook's students, also called literals. Let D be a database, which represent the constructed dataset, where each record (student's attributes values) T has a unique identifier, and contains a set of items such that $T \subseteq I$. An association rule is an implication of the form $X \Rightarrow Y$, where X, $Y \subset I$, are sets of items called correlated attributes, and $X \cap Y = \Phi$. Here, X is called antecedent, and Y consequent. Two important measures for association rules support (s) and confidence (c), can be defined as follows. The support (s) of an association rule is the ratio (in percent) of the records that contain $X \cup Y$ to the total number of records in the database. For a given number of records, confidence (c) is the ratio (in percent) of the number of records that contain $X \cup Y$ to the number of records that contain X. The confidence of a rule indicates the degree of correlation in the dataset between X and Y. Confidence is a measure of a rule's strength. Often a large confidence is required for association rules. Apply customized Fp-growth algorithm in two directions:

1. **Attributes ranker:** Through the first part of association rules you will obtain the frequency of itemsets especially single itemsets (each attribute separately). The frequency of attributes will determine the importance of the attributes among the rest, and thereby take advantage of the ranker process in clustering student's structure, as will be explained later in proposed k-mean algorithm. Ranker process is based on the frequent items and the number of appearances of specific item in the IPHES dataset compared to the rest of the items.
2. **Predicting student behavior:** By association rule resulting from the implementation of customized Fp-growth algorithm you will support the learning process through finding relationships that relate the attributes with each other's for IPHES students through mined rules.

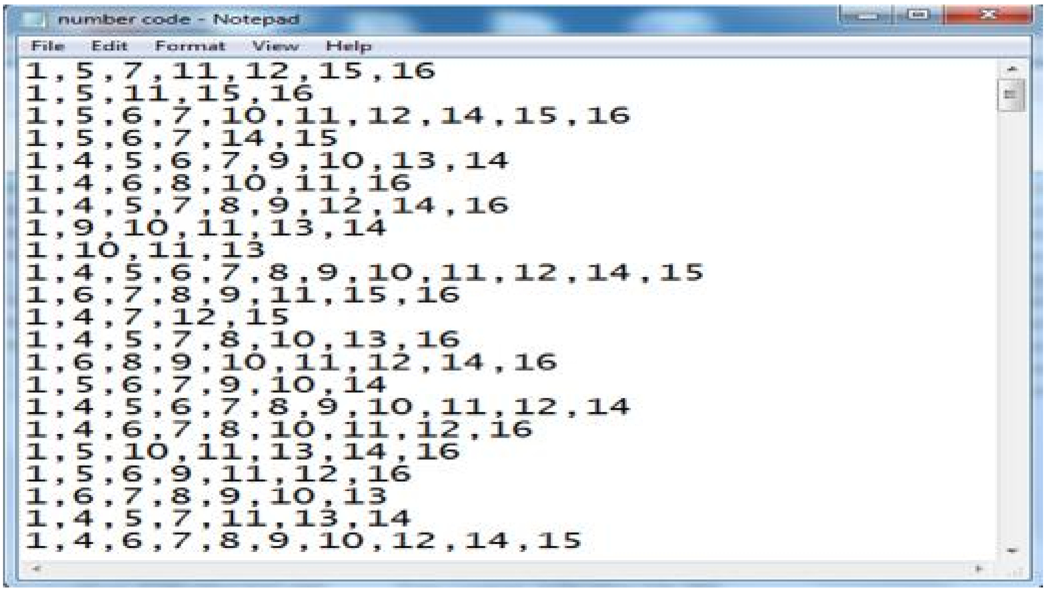
The FP Growth algorithm is presented in Algorithm (2), this algorithm will explain the method of extracting the frequent itemsets. The procedure of generating the association rules is not presented since it as in traditional Apriori Algorithm.

Algorithm (2): Customized FP-Growth Algorithm

Input: A transactional database DB, minimum support threshold
Output: Its frequent pattern tree, tree structure

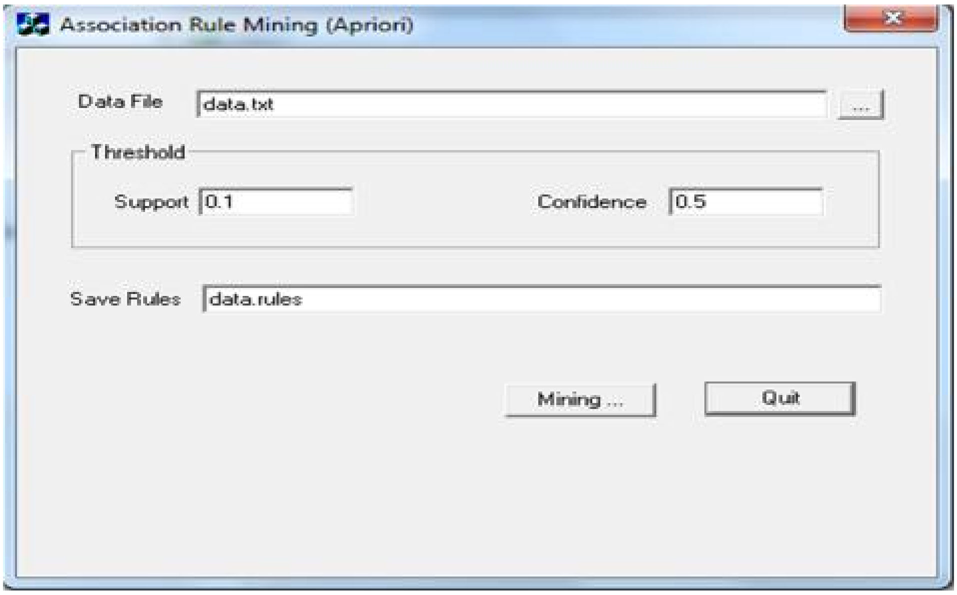
```
Process
 1. The FP-tree is constructed in the following steps:
(a) Scan the transaction database DB once. Collect the set of frequent items F and their supports. Sort F in support descending order as L, the list of frequent items.
(b) Create the root of an FP-tree, T, and label it as "root". For each transaction Trans in DB do the following:
•       Select and sort the frequent items in Trans according to the order of L.
•       Let the sorted frequent item list in Trans be [p | P], where p is the first element and Pis the remaining list.
•       Call insert_tree function ([p | P], T).

Insert_ tree function ([p | P], T) :
//The function insert_tree ([p | P], T) is performed as follows:
1.      If T has a child N such that (N.item-name = p.item-name) then
Increment N's count by 1.
Else
Create a new node N, and let its count be 1, its parent link be linked to T, and its node-link be linked to the nodes with the same item-name via the node-link structure
2.      If P is nonempty, call insert_tree (P, N) recursively.
3.      The FP_tree is mined by calling FP-growth(FP-tree, null)
Call FP-growth (FP-tree, null).
// Procedure FP-growth(Tree, A) {
•       if Tree contains a single path P then
•        for each combination (denoted as B) of the nodes in the path P
•       generate pattern B $\cup$ A with support=minimum support of nodes in B
•       else for each ai in the header of the Tree do
{
•       generate pattern B = ai $\cup$ A with support = ai.support; construct B's conditional pattern base by traverse the FP-tree by following the link of each frequent item and accumulate all of transformed prefix paths of that item and construct B's conditional FP_tree (TreeB) by
•       accumulate the count for each item in the conditional pattern base >= minimum support
•       if Tree B $\neq \emptyset$ then
•       call FP-growth(Tree B, B)
}
}
End Process
```

Figure (3), present the implementation of the customized FP-Growth algorithm. In (a) shows input file of customized FP-Growth which presents dataset values as coded data, as explained in section. In (b) will see the interface of AR apriori implementation. Finally in (c) the resulted association rules are saved in .doc file.



```
1,5,7,11,12,15,16
1,5,11,15,16
1,5,6,7,10,11,12,14,15,16
1,5,6,7,14,15
1,4,5,6,7,9,10,13,14
1,4,6,8,10,11,16
1,4,5,7,8,9,12,14,16
1,9,10,11,13,14
1,10,11,13
1,4,5,6,7,8,9,10,11,12,14,15
1,6,7,8,9,11,15,16
1,4,7,12,15
1,4,5,7,8,10,13,16
1,6,8,9,10,11,12,14,16
1,5,6,7,9,10,14
1,4,5,6,7,8,9,10,11,12,14
1,4,6,7,8,10,11,12,16
1,5,10,11,13,14,16
1,5,6,9,11,12,16
1,6,7,8,9,10,13
1,4,5,7,11,13,14
1,4,6,7,8,9,10,12,14,15
```

Figure (3-a) Input text file in the Fp-Growth



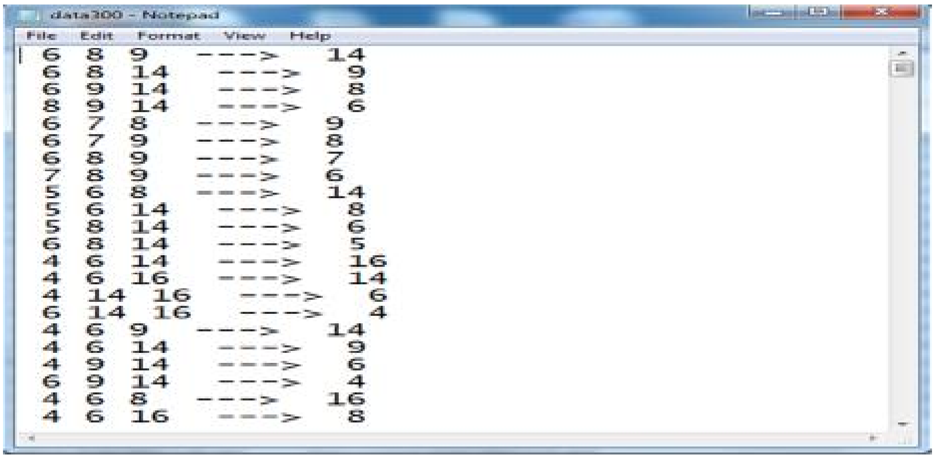
Data File: data.txt

Support: 0.1

Confidence: 0.5

Save Rules: data.rules

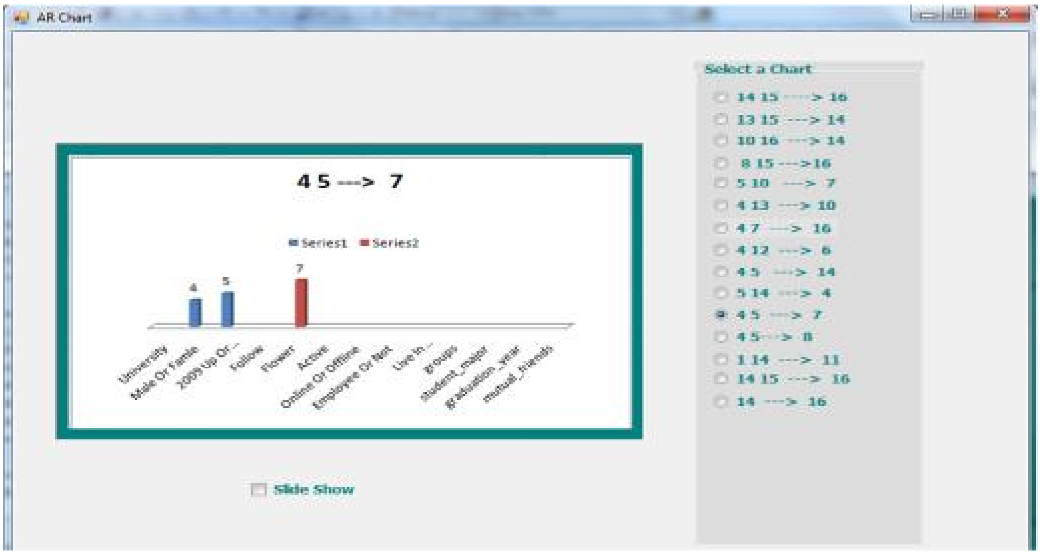
Figure (3-b) the interface of AR apriori implementation



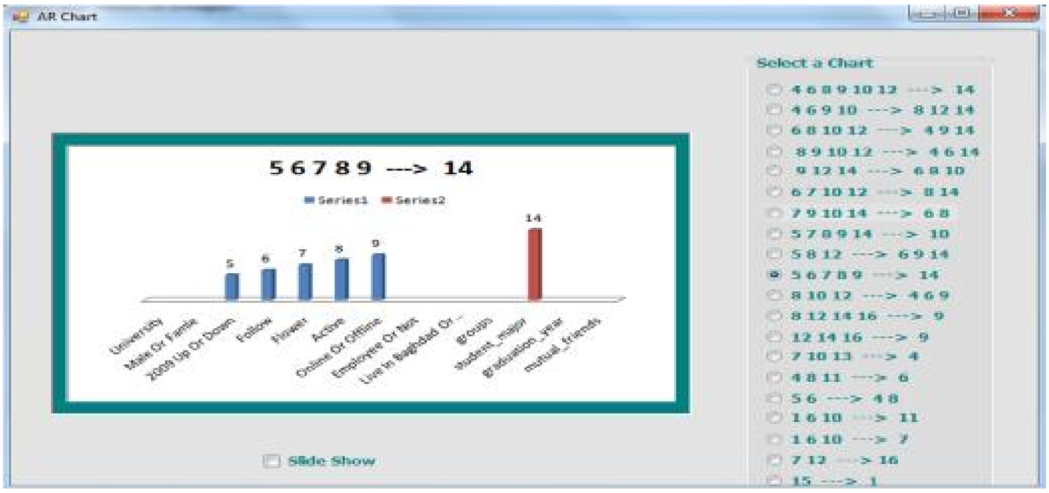
```
6  8  9    --->   14
6  8  14    --->   9
6  9  14    --->   8
8  9  14    --->   6
6  7  8    --->   9
6  7  9    --->   8
6  8  9    --->   7
7  8  9    --->   6
5  6  8    --->   14
5  6  14    --->   8
5  8  14    --->   6
6  8  14    --->   5
4  6  14    --->   16
4  6  16    --->   14
4  14  16    --->   6
6  14  16    --->   4
4  6  9    --->   14
4  6  14    --->   9
4  9  14    --->   6
6  9  14    --->   4
4  6  8    --->   16
4  6  16    --->   8
```

Figure (3-c) Output text file of the Fp_Growth
Figure (3) customized FP-Growth.

Figure (4), presents the analysis of the customized FP-Growth algorithm. Will present a sample of analysis two extracted association rules and how they are used in advance e-learning over facebook. In (a) the rule will predict a student behavior among gender, online states and student major attributes, since student follows other students, active, employee and member in other groups. In (b) the rule will predict a student behavior among follow and active attributes, since student is the follower, online, employee and has a specific major.



(a)



(b)

Figure (4) Two analyzed association rules.

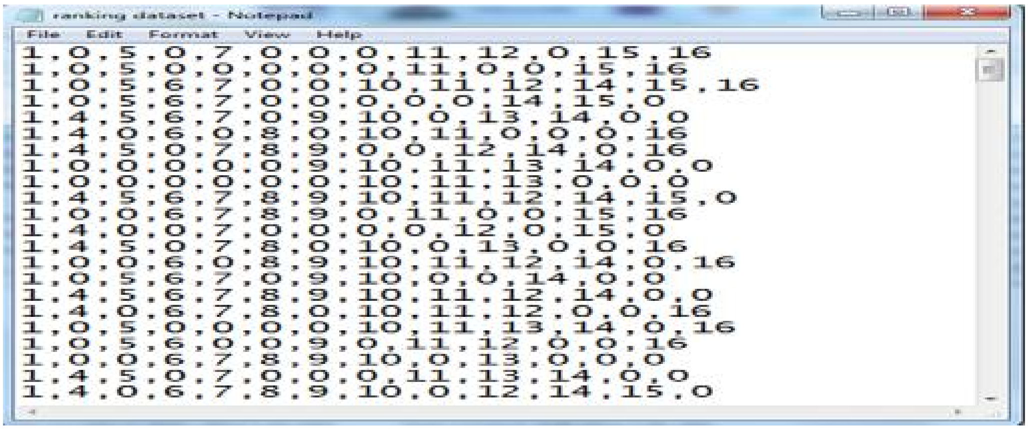
Apply K-mean Technique

The K means algorithm will do the three steps until convergence, these steps are explained in algorithm (3) below.

Algorithm (3): Traditional K-Mean
Input: Dataset as vectors, n present no. of clusters. **Output:** n clusters
Process 1. Iterate until *stable* (= no object move group): 2. Determine the centroid coordinate. 3. Determine the distance of each object to the centroids. 4. Group the object based on minimum distance (find the closest centroid). **End Process**

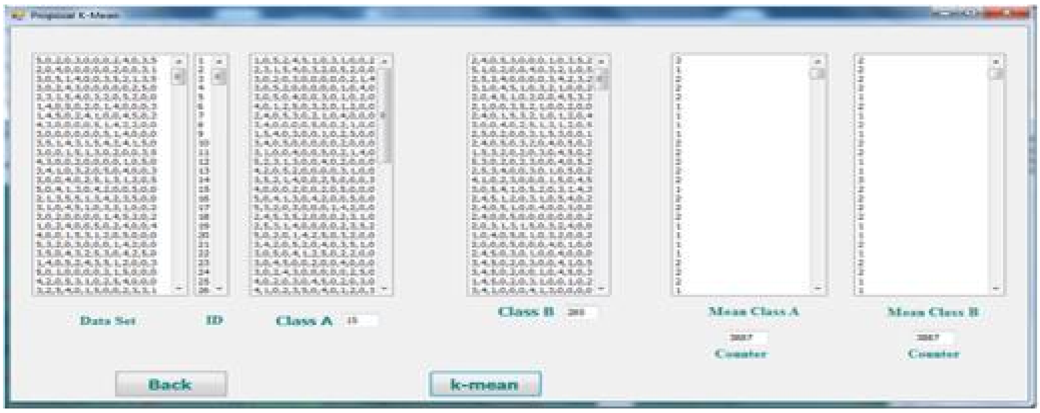
After implementing the apriori on IPHECSS will apply the clustering technique using k-mean algorithm, a suggestion to preprocess the attributes by weighting those accords their importance. The importance of attributes resulted from applying AR techniques as ranker of attributes. The AR ranker, rank attributes according their frequency. So will give heavy weight for important attributes then reduce the weights of attributes with decreasing their importance. Then apply proposed k-mean on weighted dataset. The proposed k-mean comes from the fail of traditional k-mean in clustering the students into separated clusters and because the traditional present each student as a vector of attributes values. Using traditional Euclidian distance similarity will give similarity among students' vectors, so the clustering will be wrong since it takes values of attributes in general. Where the proposed k-mean method will concentrates on the similarity of weighted attributes not as traditional which depends on similarity of attributes values.

Figure (5) will show in (a) dataset after weighting and (b) after apply proposed k-mean of the weighted dataset.



```
1.0.5.0.7.0.0.0.11,12,0,15,16
1.0.5.0.0.0.0.0.11,0,0,15,16
1.0.5.6.7.0.0.10,11,12,14,15,16
1.0.5.6.7.0.0.0.0.0.14,15,0
1.4.5.6.7.0.9.10.0.13.14.0.0
1.4.0.6.0.8.0.10.11,0.0.0.16
1.4.5.0.7.8.9.0.0.12.14.0.16
1.0.0.0.0.0.9.10.11.13.14.0.0
1.0.0.0.0.0.0.10.11.13.0.0.0
1.4.5.6.7.8.9.10,11,12.14.15.0
1.0.0.6.7.8.9.0.11,0.0.15.16
1.4.0.0.7.0.0.0.0.12.0.15.0
1.4.5.0.7.8.0.10.0.13.0.0.16
1.0.0.6.0.8.9.10.11,12.14.0.16
1.0.5.6.7.0.9.10.0.0.14.0.0
1.4.5.6.7.8.9.10.11.12.14.0.0
1.4.0.6.7.8.0.10.11.12.0.0.16
1.0.5.0.0.0.0.10.11.13.14.0.16
1.0.5.6.0.0.9.0.11.12.0.0.16
1.0.0.6.7.8.9.10.0.13.0.0.0
1.4.5.0.7.0.0.0.11.13.14.0.0
1.4.0.6.7.8.9.10.0.12.14.15.0
```

(a)



(b)

Figure (5) Proposed k-mean implementation.

Figure (6) shows the (a) analysis of traditional vs. (b) proposed k-mean by using UCINET program on 100 records. UCINET is a famous Social Network Analysis (SNA) used to analyze the interpersonal relationships within a Facebook group and can provide rich and systematic descriptions and interpretation of complex social relationships. UCINET focuses on the interconnections of the actors, instead of on the peculiarities of the actors themselves.

From analysis of traditional k-mean the following information is obtained:

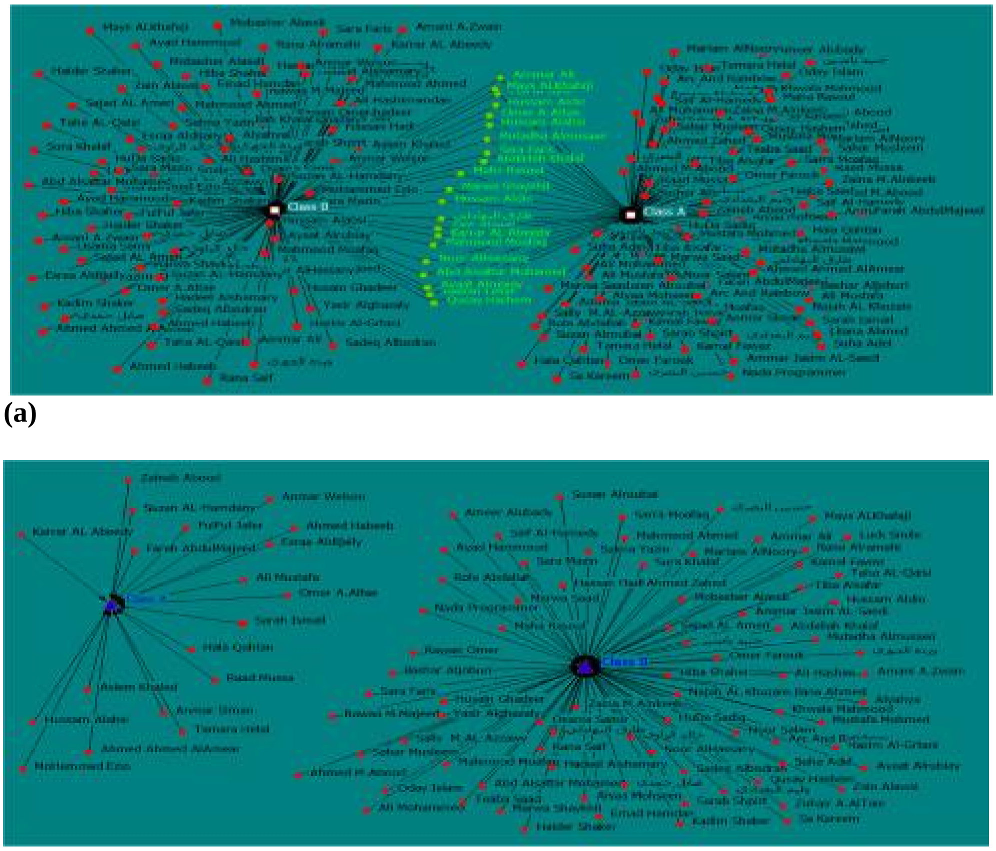
- Number of Class: 2 classes A and B
- Number of Overlap: 19 Elements
- Number of Redundancies: 62 in class A and 57 in class B
- Number of student in class A: 43 Without Redundancy
- Number of student in class B: 38 Without Redundancy

- Selected centeriod: Randomly

From analysis of proposal modified k-mean the following information is obtained:

- Number of overlap: 0
- Number of redundancy: 0
- Number of student in class A: 18
- Number of student in class B: 82

Selected centeriod: All students are taken as centeriod each time with mean of each class.



(a)

(b)

Figure (6) Visualization of traditional k-mean and proposed k-mean.

From analyzing the results of proposal k-mean, there are many points of predictions of students' structure, these are:

1. The proposal introduces some of suggested terms such as positive student and positive time. Positive student means a student that has the ability to influence other students in the group. That is, the positive student can easily influence other students through direct or indirect connects. Influence is the capability of information diffusion in social networks.
2. With the greater influence, the information can be spread to other student more quickly and widely. We divide influence feasibility of the student into two parameters,

influence range and influence degree. Influence range means the spreading scope of message or behavior, and influence degree represents how deep the student can affect the neighbor's students.

3. Positive time means the time which has the greatest influence. That is, a positive time point is a specific time for a facebook that has a large number of positive students who are actually online.
4. The proposal can infer that the capability of information diffusion may be more powerful because the number of online students more than offline and the number of students which have medium influence power is more than low influences.
5. The proposal can infer that the influence power is stronger with high number of high influential students.
6. The proposal can infer that the information diffusion may be more effective by selecting the right time,
7. The proposal needs to consider the distribution of students and the relation structure of students.
8. There are two different relation structures of students; quality and quantity. The amount of online users is more high quantity, but the quantity of high and medium influence students is much more quantity and quality.
9. If we only use the number of online students to evaluate the influence power, the results are difficult to convince people. So we need to consider the distribution of students. Besides, not only the amount of online students and their distribution, but the relation structure of students must be taken into account.
10. Here come some examples: if there are two students, A and B, and they both have many neighbors' students which are in other groups. So the information can be spread to many students or more.

CONCLUSIONS

From design, implementing and results analysis we conclude the following points:

1. Customized Fp-growth is linked with proposed K-Mean to build an integrated model, from which customized Fp-growth is used in ranking attributes according their importance. They are used it in proposed K-Mean to identify Facebook members or users who could be relied upon in diffusion process to the rest of users taking into consideration the attributes of the missing and non-integrated to create relationships with other attributes that could be relied on.
2. The FP-growth method transforms the problem of searching for long frequent patterns into shorter ones in much smaller conditional databases recursively and then connecting the suffix, it uses the least frequent items as a suffix, and this is done by building Fp-tree, offering good selectivity in term of timing.
3. The Fp-Growth provides efficiency in terms of information storage this is useful in dealing with the large actual data base that will be dealt with, especially the increasing demand for the use of Facebook by users and spread everywhere.
4. The proposed k-mean comes from the fail of traditional k-mean in clustering the students into separate clusters. In traditional k-mean it is possible to repeat in the same cluster and the same student appears in two clusters, whereas in proposed k-mean there is

no repeat in the clusters and cluster are completely separated from some. This is what matters in information diffusion in Facebook because.

5. Diffusion focuses on delivery of information to the largest number of students on Facebook in shortest time, the repetition prevents this aim.

6. The traditional system presents each student as a single value of attributes. Using traditional Euclidian distance similarity will give similarity to student's vectors, so the clustering will be wrong since it takes values of attributes in general. On the other hand the proposed k-mean method will concentrate on the similarity of weighted attributes not as in traditional system which depends on similarity of attributes as a vector for each one.

7. The traditional k-mean selects, the centroid randomly, this means it is possible that the problem which lies the selection is not the best, while in the proposal k-mean algorithm this problem is solved because it takes all student attributes as a centeriod.

REFERENCES

[1]. Adamic L. A., and Adar E., “Friends and Neighbors on the Web”, Social Networks, Vol. 25, 2007, pp. 211-230.

[2]. Jamali M. and Ester M., “Mining Social Networks for Recommendation”, at ICDM 2011, December 12th 2011.

[3]. Romero C. and Ventura S.," Data Mining in E-Learning", University of Cordoba, Spain, University of Cordoba, Spain, (2006).

[4]. Lile A., "Analyzing E-Learning Systems Using Educational Data Mining Techniques", Mediterranean Journal of Social Sciences, (2011).

[5]. Krpan D. and Stankov S., “Educational Data Mining for Grouping Students in E-learning System”, 34th Int. Conf. on Information Technology Interface, 2012.

[6]. Nancy.P, Ramani R. G., “Discovery of Patterns and evaluation of Clustering Alg'orithms in Social Network Data (Face book 100 Universities) through Data Mining Techniques and Methods”, International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.2, No.5, September 2012.

[7]. Bonchi F., “Influence Propagation in Social Networks: A Data Mining Perspective”, IEEE Intelligent Informatics Bulletin, Vol.12, No.1, December 2011.

[8]. Raju E. and Sravanthi K., “Analysis of Social Networks Using the Techniques of Web Mining”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 10, October 2012 ISSN: 2277 128X, www.ijarcsse.com

[9]. Falakmasir M. H. and Habibi J., “Using Educational Data Mining Methods to Study the Impact of Virtual Classroom in E-Learning”,

[10]. Sael N. , Marzak A. and Behja H., “Multilevel clustering and association rule mining for learners' profiles analysis”, IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 3, No 1, May 2013

[11]. Hu H. W. and Lee S. Y., “Study on Influence Diffusion in Social Network”, International Journal of Computer Science and Electronics Engineering (IJCSEE) Volume 1, Issue 2 (2013) ISSN 2320–4028 (Online).