

APLICACIÓN SE SOCIAL NETWORK ANALYSIS EN ENTORNOS EDUCATIVOS VIRTUALES.

Luis Fernando Salgado Durango

1.035.873.037

Proyecto integrador, Dpto Ing. Telecomunicaciones

Universidad de Antioquia

Email: fernando.salgado.edu.co

Resumen—En este artículo se pretende utilizar las técnicas que ofrece la minería de datos en la educación para encontrar una red que contenga todas interacciones sociales de un grupo de estudiantes en un entorno de educación virtual, para obtener datos que sumados con la información obtenida de los cursos virtuales, determinar la relación que hay entre el desempeño y la deserción académica con las interacciones sociales de los estudiantes, todo esto para ofrecer un análisis analítico que facilite a los tutores de este modelo educativo conocer los factores importantes en el proceso de aprendizaje.

Palabras Claves—social network analysis (SNA), Gephi, educational data mining (EDM) .

I. INTRODUCCIÓN

Estamos en un tiempo donde la importancia de los datos es muy grande, por esto muchas entidades dedican grandes esfuerzos para hacer el mejor uso de la información a la que estas pueden acceder, lo que se refleja en una creciente relevancia de las técnicas que permiten realizar el análisis de datos, entre ellas hay una rama de esta disciplina que se desempeña en el área de la educación, ya que la cantidad de empresas que prestan este servicio va en aumento y con ella la manera de brindar información a un grupo de estudiantes en un entorno virtual. La minería de datos en la educación ofrece un buen número de herramientas para abordar la información que aportan estos entornos de educación virtual y al implementarlas obtener información que no es evidente, para con esta mejorar la manera en que se enseña en este tipo de plataformas.

Con esto en mente, en este proyecto se pretende demostrar que es posible encontrar una red en la que se pueda ver las interacciones sociales en un aula virtual utilizando técnicas de minería de datos en la educación y con ella hallar información para saber si esta tiene alguna correlación con el desempeño de los estudiantes o identificar los estudiantes en riesgo dentro del aula. La importancia de este estudio radica en que se ayuda a dar una idea analítica de cómo influyen las relaciones intrapersonales y sociales en el proceso de

aprendizaje de un estudiante que utiliza una plataforma virtual, ya que habrá unos resultados basados en análisis de los datos consignados por los estudiantes en los servidores de la plataforma. Con resultados de estudios similares a este se puede tomar decisiones para mejorar la forma en la que se le enseña en estos espacios.

los datos que se utilizarán para realizar este estudio serán proporcionados por parte ingenia que es una unidad de la facultad de ingenia de la universidad de Antioquia, la investigación se centrará en tres de los cursos virtuales que ofrece esta plataforma.

II. MARCO TEÓRICO Y ESTADO DEL ARTE

Actualmente con el crecimiento de las plataformas que brindan cursos en línea, crece la necesidad de aprovechar de mejor manera los recursos que estas tienen y mejorar los servicios que ofrecen, La minería de datos en la educación (EDM) como se indican en [1] EDM nace como un paradigma orientado al diseño de modelos, tareas, métodos o algoritmos para explorar datos del ámbito educativo y de [2] se puede concluir que EDM permite resolver incógnitas sobre como aprovecha el estudiante los elementos que le brinda la plataforma, que tanto influye estos elementos en el proceso de aprendizaje y que está aprendiendo realmente el estudiante, para después de realizar una serie de análisis entregar información que sea de utilidad para basados en ellas se mejore la forma en la que la información es suministrada al estudiante, y con esto el docente podrá tomar decisiones sobre cuál es la mejor manera de interaccionar con el estudiante o dar importancia a las variables que más repercuten dentro del proceso de aprendizaje.

EDM cuenta con varias maneras de aportar a los dos sistemas de educación (tradicional y a distancia) debe saber a cuál de los dos se quiere realizar el análisis para con esto tener claridad de cuales métodos tienen una mayor efectividad en el análisis de los datos. encontramos en [3]

unas definiciones y claras diferenciaciones de estos sistemas de educación. En el ambiente tradicional que es el más ampliamente usado y necesita de un lugar físico para la reunión de los integrantes del grupo, se debe tener en cuenta que los datos que se almacenan para un posterior análisis, son menores y requiere de un mayor esfuerzo por parte de los tutores o coordinadores el almacenar estos. Lo que de cierta manera limita los alcances que pueda aportar un análisis con EDM. Como para este caso se trabajará con ambientes a distancia este documento se enfocará más en estos.

Los ambientes a distancia son los que utilizan unas herramientas o técnicas que permiten al estudiante acceder a la información suministrada por el tutor sin necesidad de que exista un contacto físico entre ellos, como estos están soportados en su mayoría en un servidor web la información que pueden ofrecer es más amplia y con mayores facilidades para su captura, documentos como los:

Log file: contiene la información de las interacciones entre agentes del sistema de educación, estas interacciones pueden ser realizadas por el servidor, la cual contendrá información del tiempo, respuestas y entradas, por el cliente el cual será el estudiante y dará información de su interacción con la plataforma y por último por el proxí la cual tendrá información de la comunicación entre cliente y servidor.

Contenido: información dentro de la plataforma que fue dispuesta al estudiante.

Estructura: información que describa la organización que tiene el contenido.

Uso: datos que contienen los patrones de uso que se le da a la información que existe en la plataforma.

Perfil del usuario: información personal del estudiante.

Al igual que una buena fuente de información y comprensión de los datos que esta puede ofrecer, es importante saber el tratamiento que se le debe dar a la información, dado que por lo general la información viene dada de forma o en formatos que no son adecuados para aplicar las herramientas que brindan en EDM, para esto se recomienda realizar un procesamiento previo a la información para adecuarla en [4] indican una variedad de formatos posibles o herramientas para tratar información, este caso se centrará en la adecuación de los datos para el uso de lenguajes de programación como Python que permite realizar análisis estadístico de datos científicos o R que es una herramienta que permite el análisis estadístico de datos y aporta análisis de redes sociales dada la cantidad de paquetes que tiene incluidos para el análisis de social networks adapting (SNA).

Por la facilidad que ofrece para el ordenamiento de datos y comprensión a la lectura de estos por su interfaz Microsoft

Excel una opción bastante tentativa ya que también permite el análisis de datos científicos. La manera en que esta herramienta representa los datos permite identificar problemas estructurales y semánticos en las variables ya representa estas de una manera clara además que es fácil de usar en Python o R. Para el estudio de las técnicas que brinda la EDM se considera correcto ir ilustrando con casos de investigación realizadas con anterioridad.

En [5] se puede ver como se estudia las interacciones sociales en un aula de clase por medios de análisis cualitativo y SNA, donde esta última describe patrones de relaciones interpersonales, lo que en si es una técnica de reconocimiento de patrones, pero también se utilizan técnicas de agrupamiento. La técnica de agrupamiento está presente en una gran cantidad de soluciones implementadas para solucionar problemas, en [6] donde se pretende la detección temprana de estudiantes en riesgo, en este estudio también se utiliza una técnica de predicción. Otro estudio donde se utiliza la técnica de agrupamiento es [7] que intenta proponer agrupamientos y dar un perfil a estudiantes de acuerdo a las interacciones que tengan con un tutor inteligente. Como último ejemplo para esta técnica se tiene [8] donde se pretende utilizar métodos supervisados y no supervisados para la creación de modelos de usuario para la exploración de ambientes de aprendizaje, en este se usan no solo agrupamiento si no también clasificación como técnicas para la solución del problema.

Como se puede notar las técnicas son variadas dentro de EDM, observando que se puede clasificar, agrupar, predecir, reconocer patrones, asociación y otras que no han sido utilizadas en los anteriores estudios, pero que [1] da a conocer brindando ejemplo de métodos o técnicas que se pueden agrupar en la I.

Como en este caso se plantea el uso de SNA se indaga mas en este, este método consiste en el estudio de redes sociales, generando una serie de patrones que permitan identificar las interacciones entre sujetos que hacen parte del grupo de estudio, esta técnica proporciona herramientas para el estudio de mercado o áreas en las que sea importante conocer la manera en que se relacionan las personas, como en el caso del marketing.

SNA aplicado en el entorno educativo permite conocer como interaccionan los estudiantes con las plataformas virtuales que permiten el desarrollo de los cursos en líneas, con esto se puede ver la que herramientas de la plataforma no están siendo utilizadas y al estudiar las interacciones entre estudiantes se puede dar una idea de cuales factores son de relevancia en el proceso de aprendizaje, SNA permite obtener una serie de atributos en forma de datos para realizar los análisis que se consideren pertinentes. Estos nuevos atributos son:

Nodo: Este representa a un integrante de la red.

Modelos	Objetivos	Técnicas	Métodos y Tareas
Predicción	Pretende inferir características dentro de una base de datos a partir del diseño de un modelo que se lo permita.	Regresión, clasificación, Agrupación, categorización.	Social network análisis (SNA), logistic regression, multiple Regression, decision tree, Pearson correlation, mean squared error, modelo oculto de markov.
Descriptivo	Dar a conocer los aspectos más significativos dentro de un grupo de datos	Regresión, Agrupación, Patrones secuenciales, reglas de asociación, análisis de correlación, análisis numérico	decision tree, interactions network, cluster analysis Association rules, performance sequence Analysis, Bayes theorem.

Cuadro I
RESUMEN TÉCNICAS Y TAREAS EDM

Enlace: representa la interacción entre dos integrantes de la red o entre dos nodos, estos pueden ser direccionados.

Grado: es el número de enlaces conectados a un nodo, sin tener en cuenta la dirección de los enlaces.

Grados de entrada y de salida: que son la cantidad de enlaces entrantes y salientes de un nodo respectivamente.

También se tienen otras medidas que son mas complejas como lo son:

Intermediación: que es igual al número de caminos más cortos de todos los nodos a todos los demás que pasan por dicho nodo.

Autoridad: se considera que un nodo es de autoridad cuando sus enlaces entrantes lo conectan a nodos que tienen una gran cantidad de enlaces salientes.

Hubs: se considera que un nodo es hub si los enlaces salientes lo conectan a nodos que tienen una gran cantidad de enlaces entrantes.

Existen otros atributos que están dedicados a conocer características de la red como un todo, estas características son:

Densidad de la red: esta mide la cantidad de enlaces que existen en la red en comparación con el número máximo posible de enlaces que pueden existir en ella.

diámetro de la red: esta es definida como la mayor cantidad de nodos que deben atravesarse para viajar desde un nodo a

otro.

cantidad de componentes de la red: es el número de subredes en las cuales dos nodos están conectados entre sí por caminos sin tener en cuenta la dirección de sus enlaces.

Reciprocidad: se produce cuando la existencia de un enlace de un nodo a otro activa la creación del enlace inverso.

Estos nuevos atributos se pueden utilizar para realizar estudios como en el caso de [9] donde se utilizan los atributos grado, grado de entrada, grado de salida, autoridad, intermediación y hub para realizar estudios de un curso en línea, de igual manera se tiene [10] donde se realiza un estudio similar solo agregando otras variables para el análisis, pero en ambos casos se llega a que estos atributos son los que mejor representan la red social.

III. DESARROLLO Y ANÁLISIS

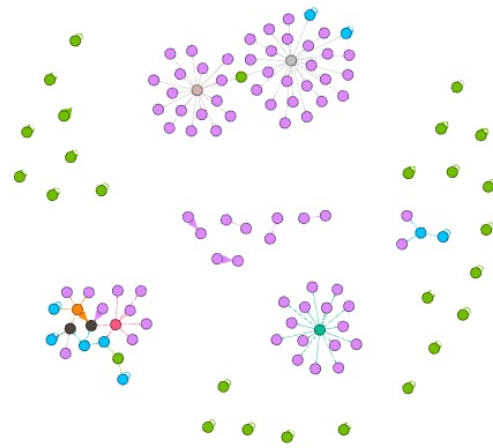


Figura 1. red de interacciones sociales

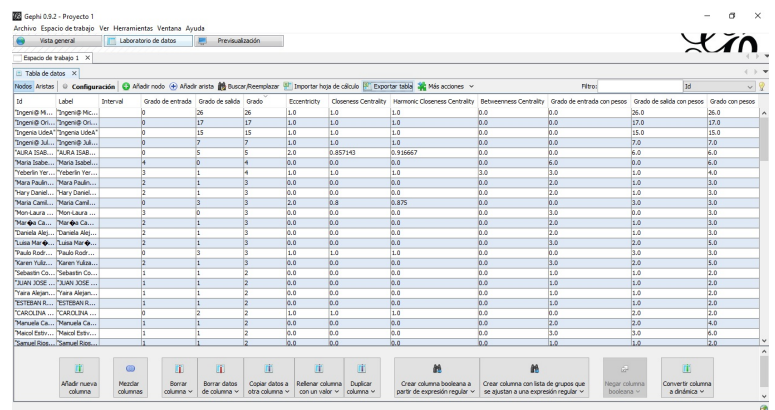


Figura 2. red de interacciones sociales

Para obtener la red de interacciones sociales se utilizó la herramienta gephi que permite el estudio de este tipo de sistemas y calcula los atributos que se mencionaron anteriormente, utilizando datos proporcionados por ingenia, se obtuvo 1 que es la red correspondiente al curso de comprensión que dicta la universidad de Antioquia por medio de ingenia. El resultado es a lo sumo interesando viéndose como hay una gran cantidad de nodos aislados y unas dub redes cuyo nodo central corresponde a un tutor que se encarga de responder las preguntas que van realizando los estudiantes, también se puede decir que la interacción entre los estudiantes es pobre y esto lo corroboramos en 2 donde se puede observar que los 3 nodos con mayores interacciones son los centros de las subredes que se visualizan en 1 y estos son tutores, también se puede decir esto ya que podemos ver en 2 que estos nodos solo se dedican a dar información ya que su grado es igual a su grado de salida, son nodos dedicados a brindar información, mientras que los otros nodos son en su mayoría receptores de esta, y no hay casi interacción entre ellos, , esto toma más relevancia cuando encontramos que de los 1200 estudiantes iniciales solo 121 hicieron al menos un uso de los foros.

IV. CONCLUSIONES

La minería de datos en la educación toma cada vez más importancia dado que las herramientas que ofrece nos permiten realizar análisis detallados de datos y con estos se mejoran los ambientes de aprendizaje que se ofrecen a los estudiantes, un claro ejemplo de estas seria SNA que para este caso se podría decir que nos permite decir que los foros no están siendo muy utilizados por los estudiantes, lo que abre una posibilidad de mejora para estos cursos que ofrece la universidad de Antioquia aunque este tendría más peso si se realiza un estudio de muchos de los cursos que ofrece así si se encuentra que la falta de interacciones en foros es una constante dentro de estos cursos, se podría pasar a planear una solución para este problema.

El análisis solo de como interaccionan los estudiantes puede verse sesgado ya que no se puede dar un análisis de en cuantos temas del curso participa el estudiante solo cuantas veces lo hace en total, si se agrega esta opcion los resultados no solo pueden darnos una mejor idea de cuales temas son en los que los estudiantes presentan mas falencias si no que puede dar una idea que aspectos del método en que se brinda la información a los estudiantes puede ser mejorada.

REFERENCIAS

- [1] A. Peña-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works," *Expert systems with applications*, vol. 41, no. 4, pp. 1432–1462, 2014.
- [2] A. M. Lince, J. F. V. Bonilla, and L. Y. C. Olivares, "Educational data mining (edm) para la determinación de comportamientos en estudiantes de ingeniería del modelo ude@."
- [3] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," *Expert systems with applications*, vol. 33, no. 1, pp. 135–146, 2007.
- [4] S. Slater, S. Joksimović, V. Kovanovic, R. S. Baker, and D. Gasevic, "Tools for educational data mining: A review," *Journal of Educational and Behavioral Statistics*, vol. 42, no. 1, pp. 85–106, 2017.
- [5] A. Martinez, Y. Dimitriadis, B. Rubia, E. Gómez, and P. De La Fuente, "Combining qualitative evaluation and social network analysis for the study of classroom social interactions," *Computers & Education*, vol. 41, no. 4, pp. 353–368, 2003.
- [6] J. Berens, K. Schneider, S. Gortz, S. Oster, and J. Burghoff, "Early detection of students at risk-predicting student dropouts using administrative student data from german universities and machine learning methods," *JEDM— Journal of Educational Data Mining*, vol. 11, no. 3, pp. 1–41, 2019.
- [7] F. Bouchet, J. M. Harley, G. J. Trevors, and R. Azevedo, "Clustering and profiling students according to their interactions with an intelligent tutoring system fostering self-regulated learning," *JEDM— Journal of Educational Data Mining*, vol. 5, no. 1, pp. 104–146, 2013.
- [8] S. Amershi and C. Conati, "Combining unsupervised and supervised classification to build user models for exploratory," *JEDM— Journal of Educational Data Mining*, vol. 1, no. 1, pp. 18–71, 2009.
- [9] C. Palazuelos, D. García-Saiz, and M. Zorrilla, "Social network analysis and data mining: an application to the e-learning context," in *International Conference on Computational Collective Intelligence*. Springer, 2013, pp. 651–660.
- [10] R. Rabbany, M. Takaffoli, and O. R. Zaiane, "Analyzing participation of students in online courses using social network analysis techniques," in *Proceedings of educational data mining*, 2011.