



**Revisión sistémica de modelos basados en el análisis de redes y los métodos de la minería de datos para determinar la importancia de un miembro en una red social corporativa.**

**Paula Andrea Solano Pardo  
Heidy Geraldine Murillo Rocha**

**Facultad de Ingeniería  
Programa Ingeniería industrial  
Ibagué, 2021**





**Revisión sistémica de modelos basados en el análisis de redes y los métodos de la minería de datos para determinar la importancia de un miembro en una red social corporativa.**

**Paula Andrea Solano Pardo  
Heidy Geraldine Murillo Rocha**

Trabajo de grado que se presenta como requisito parcial para optar al título de:  
**Ingeniera Industrial**

**Director:**  
Ing. PhD. Carlos Antonio Meisel Donoso  
Profesor Universidad de Ibagué

**Facultad de Ingeniería  
Programa Ingeniería industrial  
Ibagué, 2021**



---

## Resumen

Este documento parte de un estudio de investigación realizado en la Universidad de Ibagué en el cual se busca analizar las relaciones colaborativas entre los miembros que participan en una cadena de suministro, determinando los atributos que generan una colaboración. A raíz de lo anterior, se busca identificar los elementos o variables que contribuyen a la selección de los miembros más influyentes e importantes en una corporación, a través de los métodos existentes del análisis de redes y minería de datos. Para esto, se utiliza la metodología de revisión sistemática de literatura, definiendo criterios de búsqueda e inclusión y criterios para realizar la extracción de los datos. Los resultados de esta investigación contribuyen a ampliar el conocimiento respecto a este tema, consignando toda la información disponible en el campo de la literatura y de la investigación en su solo documento, con el fin de estimular a los investigadores a implementar estos modelos en diferentes escenarios y ampliar su aplicación en pequeñas, medianas y grandes empresas, comparando la precisión de los resultados.

**Palabras claves: redes sociales corporativas, análisis de redes sociales, minería de datos, clasificación de miembros importantes, definición de roles organizacionales.**

## Abstract

This document is based on a research study carried out at the University of Ibagué in which it seeks to analyze the collaborative relationships between the members that participate in a supply chain, determining the attributes that generate collaboration. As a result of the above, it seeks to identify the elements or variables that contribute to the selection of the most influential and important members in a corporation, through the existing methods of network analysis and data mining. For this, the methodology of systematic literature review is used, defining search and inclusion criteria and criteria for data extraction. The results of this research contribute to broadening the knowledge regarding this topic, consigning all the information available in the field of literature and research in its single document, in order to stimulate researchers to implement these models in different settings. and expand its application in small, medium and large companies, comparing the precision of the results.

**Keywords: corporate social networks, social network analysis, data mining, classification of important members, definition of organizational roles.**



## Contenido

<b>Introducción</b> .....	<b>1</b>
<b>1.Planteamiento del problema y justificación</b> .....	<b>2</b>
<b>2.Objetivos</b> .....	<b>3</b>
2.1 Objetivo General .....	3
2.2 Objetivos Específicos .....	3
<b>3. Marco de referencia</b> .....	<b>4</b>
3.1 Análisis de redes sociales corporativas .....	5
3.1.1 Medidas locales .....	6
3.1.2 Medidas globales .....	9
3.2 Minería de datos .....	10
<b>4. Metodología</b> .....	<b>15</b>
4.1 Tipo, Naturaleza y Enfoque del estudio .....	15
4.2 Plan de análisis .....	15
4.2.1 Estrategia y criterios de búsqueda bibliográfica.....	15
4.2.2 Selección y exclusión de literatura .....	27
4.2.3 Extracción de datos .....	29
4.2.4 Análisis de datos e interpretación de los resultados .....	31
<b>5. Resultados</b> .....	<b>32</b>
5.1 Propósito.....	33
5.2 Validación .....	34
5.3 Modelos basados en el análisis de redes sociales corporativas .....	42
5.4 Métodos basados en la minería de datos .....	45
<b>6. Discusión</b> .....	<b>58</b>
<b>7. Bibliografía</b> .....	<b>65</b>



---

## Lista de tablas

<b>Tabla 1.</b> Palabras clave para la búsqueda en las bases de datos académicas definidas en inglés .....	16
<b>Tabla 2.</b> Palabras clave para la búsqueda en las bases de datos académicas definidas en español .....	17
<b>Tabla 3.</b> Combinaciones en inglés .....	18
<b>Tabla 4.</b> Combinaciones en español .....	22
<b>Tabla 5.</b> Descripción de datos extraídos de estudios de modelos de análisis de redes sociales y minería de datos .....	29
<b>Tabla 6.</b> Información sintetizada sobre la validación de los métodos .....	38
<b>Tabla 7.</b> Aspectos y criterios relevantes para determinar la importancia de un miembro en una red social corporativa en inglés .....	51

## Lista de figuras

<b>Figura 1.</b> Diagrama de flujo PRISMA.....	28
--	----



## Introducción

Los entornos organizacionales se han tenido que adaptar a la transformación y evolución digital e informática, que han generado nuevas y mejores formas de gestionar, coordinar, comunicar y desempeñar las actividades diarias, permitiendo el cumplimiento de los objetivos operativos, tácticos y estratégicos. El dinamismo y la resiliencia son pilares fundamentales que garantizan el desarrollo empresarial, ampliando el campo de acción y dando paso a que se exploren diferentes estrategias como: la inclinación al trabajo en equipo, proyectos de mejora continua, creación de joint ventures, alianzas, cadenas de suministro integradas, redes de innovación entre empleados, cambios estructurales, reubicación y/o reasignación de funciones dentro de grupos de trabajo, implementación de herramientas internas para derribar barreras que obstruyen los flujos comunicacionales y detección del talento oculto (de Toni & Nonino, 2010).

La mayoría de las empresas se enfrentan al problema de cómo gestionar el constante cambio organizativo, la comunicación y el aprendizaje. El proceso se puede realizar con la ayuda de miembros claves e influyentes, puesto que difunden información y brindan apoyo a los empleados, están verdaderamente comprometidos con el trabajo que realizan, son líderes, con capacidades para planificar estrategias y proyectos de gran impacto en la organización. Por lo tanto, la selección e identificación de estos actores, que desempeñan roles de agentes de cambio se ha convertido para las organizaciones en una tarea sumamente importante y necesaria. El problema radica en cómo seleccionar a los empleados adecuados. El proceso clásico de selección generalmente se basa en información sobre las posiciones de los empleados en la jerarquía organizacional, en la antigüedad, la experiencia o los roles específicos que desempeñan en los procesos o proyectos (Zbieg et al., 2016). Por esto, el objetivo de esta monografía es realizar una revisión sistémica de estudios que han utilizado métodos de análisis de redes tradicionales y minería de datos en el entorno empresarial, para determinar la importancia de un miembro en una red social corporativa.

Sin embargo, las técnicas ambiguas o clásicas para identificar la importancia de los empleados son muy subjetivas, poco precisas y básicas. Por esta razón, se implementan diferentes métodos en el proceso de selección de actores claves, que permiten clasificar los miembros de la organización y definir roles de forma asertiva. De ahí surgen los modelos pertenecientes al análisis de redes sociales y la minería de datos, los cuales buscan representar y realizar un bosquejo del sistema real, describiendo las interacciones, relaciones y el flujo de información entre los nodos que conforman esa red corporativa. Por otro lado, estos métodos permiten tener una visualización global de la gestión de la empresa. De este modo, el documento expone a cada uno de los modelos, su funcionalidad y aplicabilidad en el campo de la investigación. También, resalta la utilidad que han tenido en el ámbito organizacional para detectar miembros influyentes con estudios reales.



## 1. Planteamiento del problema y justificación

Debido a la transformación digital, las empresas han presentado problemas relacionados con la organización del trabajo y el flujo de comunicación e información. Puesto que, las corporaciones solo analizan a miembros con cargos calificados, tales como, directivos, colaboradores o socios, sin tener en cuenta a todos los actores (Bar et al., 2000), e impidiendo la relación directa entre algunos nodos. Por lo tanto, la transmisión de la información se ve afectada por la falta de comunicación entre todos los trabajadores. Esto a su vez provoca la existencia de miembros que no poseen una intervención significativa en los diferentes proyectos o actividades que se realizan dentro de la empresa.

En este mismo orden de ideas, la centralidad de una red corporativa indica que tan importante es este nodo en dicha red (Sparrowe et al., 2001). Los individuos centrales, debido a sus conexiones más numerosas con los demás, tienen más relaciones a las que recurrir para obtener recursos (conocimiento específico de la tarea e información confidencial sobre temas relacionados con el trabajo) y, por lo tanto, dependen menos de un solo individuo (Cook & Emerson, 1978; Sparrowe et al., 2001). Por otro lado, la centralidad está relacionada con el poder, la influencia en la toma de decisiones, la facilidad para resolver problemas y la innovación que tiene el actor dentro de la red organizacional (Weissman et al., 1990; Yamadori et al., 1990). Por el contrario, aquellos que están en posiciones periféricas en la red tienen más dificultad para desarrollar experiencias sobre problemas y soluciones relacionados con las tareas y, por lo tanto, es menos probable que desarrollen las competencias y la experiencia necesarias para alcanzar altos niveles de rendimiento (Sparrowe et al., 2001).

Esta investigación permitirá encontrar los diferentes métodos que se han desarrollado a lo largo de la literatura, pertenecientes al análisis de redes y minería de datos, que buscan determinar el grado de importancia y participación de un miembro en la red social corporativa (Arif et al., 2015; Gaur, 2012; Shetty & Adibi, 2005) con el objetivo de mejorar la estructura organizacional de la empresa, realizar un mejor aprovechamiento del conocimiento organizacional, obtener de forma más rápida y eficiente flujos de información, aumentar la coordinación de la comunicación grupal, lograr una mayor participación de los empleados, una programación rápida, asignación de tareas e informes eficientes, mejor comunicación a través de niveles jerárquicos y restablecer la coordinación de la comunicación dentro de grupos dispersos (Bailyn et al., 1992; Bar et al., 2000).



## **2. Objetivos**

### **2.1 Objetivo general**

Realizar una revisión sistémica de estudios que han utilizado métodos de análisis de redes tradicionales y minería de datos en el entorno empresarial, para determinar la importancia de un miembro en una red social corporativa.

### **2.2 Objetivos específicos**

- Identificar los artículos que han empleado métodos de análisis de redes tradicionales y de minería de datos en el entorno empresarial para hacer la revisión de la literatura.
- Definir los aspectos y criterios relevantes usados en el campo de la investigación teórica para determinar el relacionamiento efectivo y predecir la importancia de un miembro en una red social corporativa.
- Realizar una taxonomía de los métodos de análisis de redes tradicionales y de minería de datos que han sido usados en la literatura científica para determinar el relacionamiento efectivo y predecir la importancia de un miembro en una red social corporativa.



### 3. Marco de referencia

Para que una organización funcione de forma correcta, esta debe estar compuesta por diferentes grados de jerarquía, los cuales garantizan la existencia de miembros encargados de liderar y gestionar el direccionamiento de la compañía. Los actores que ocupan cargos calificados tienen la responsabilidad de dar ejemplo e involucrarse activamente en los procesos y propiciar la credibilidad, productividad, eficiencia, compromiso, comunicación y el trabajo colaborativo, de lo contrario, cualquier proceso de desarrollo con la gente no es válido ni alcanzarán sus objetivos propuestos (Carrión, 2010). Sin embargo, existen situaciones donde no siempre los directivos son los miembros con mayor importancia, debido a que no asumen un comportamiento de liderazgo e influencia.

Las características que determinan la importancia de un nodo en una red corporativa son: el liderazgo y la inteligencia emocional. No es que el coeficiente intelectual y las destrezas técnicas sean irrelevantes. Son importantes, pero como “aptitudes de umbral”; es decir, son los requisitos básicos para puestos ejecutivos. Pero estudios recientes, muestran claramente que la inteligencia emocional es una condición crucial que debe tener un empleado que se considere clave (TORRES, 2011). Sin ella, una persona puede tener la mejor preparación del mundo, una mente incisiva y analítica, y un infinito surtido de ideas inteligentes, pero aun así no logrará influenciar, controlar, planificar, guiar, dirigir, ayudar, trazar metas, formular valores, crear y transformar la cultura (TORRES, 2011).

Un nodo para ser considerado clave en una red o sistema debe tener competencias informativas, persuasivas, participativas, y delegatorias, y generar potencial en los demás. Las competencias más significativas que hacen que un miembro sea importante son: la experticia, que significa dominio, maestría en el hacer y uso instrumental de ese hacer. La maestría implica concentración absoluta, entrega total de la persona y eficiencia. Por otro lado, debe adquirir un alto grado de sinergia completa de sus habilidades técnicas, para llevar a feliz término el desarrollo de su cargo (Koch, 1993).

Según García et al. (2018) los criterios y destrezas generales que convierten a un individuo en alguien productivo, significativo, participativo e indispensable son:

**Habilidades técnicas:** Integran los conocimientos acerca de los métodos, procesos, procedimientos y técnicas para la conducción de las actividades de la unidad de trabajo del líder.

**Habilidades interpersonales:** Corresponden a la conducta y procesos interpersonales, habilidad para comprender los sentimientos, actitudes y motivos de otros a partir de lo que dicen o hacen, es la habilidad para comunicarse de manera clara y persuasiva, para establecer



relaciones cooperativas (tacto, diplomacia, encanto, empatía, sensibilidad social, fluidez en el lenguaje, etc.).

**Habilidades conceptuales:** Se refieren a la habilidad analítica general, pensamiento lógico, eficiencia en la formación de conceptos y conceptualización de relaciones ambiguas y complejas, creatividad en la generación de ideas y solución de problemas, habilidad para analizar los hechos, percibir las tendencias, anticipar los cambios y reconocer las oportunidades y problemas potenciales.

Partiendo de lo anterior, puede presentarse el caso en donde los miembros de la organización con menor nivel jerárquico o que aparentemente son pocos calificados para participar en el desarrollo estratégico y aportar ideas constructivas, sean los que más se destaquen en las actividades diarias y generen mayor valor e impacto interno (Carrión, 2010). Por lo tanto, es importante que se implementen modelos basados en el análisis de redes y métodos de la minería de datos que permitan de forma objetiva identificar los nodos claves y la importancia de estos en la red social corporativa.

### 3.1 Análisis de redes sociales corporativas

El análisis de las redes sociales corporativas (ACSN) es una analogía de cómo el mundo está interconectado e interrelacionado, estas redes están formadas por nodos y enlaces. Los nodos son los agentes, las personas, las organizaciones o el objeto de estudio y los enlaces son las interacciones y el intercambio de información entre los nodos (Snow & Fjeldstad, 2015). El análisis de las redes sociales corporativas es ampliamente utilizado en las organizaciones para evaluar el desempeño, mejorar la gestión empresarial, el rendimiento, el crecimiento de recursos, la innovación (Teixeira et al., 2019), investigar patrones de influencia en redes interorganizacionales, estudiar el poder o la competencia en las organizaciones (Scott, 1988), descubrir la estructura organizativa oculta y la selección de miembros destacados. Un ejemplo interesante de tal desafío es descubrir grupos ocultos y la importancia de un miembro de la red social corporativa mediante el análisis de sus registros de correo electrónico (Shetty & Adibi, 2005).

Según Scott (1988), en las redes sociales corporativas, las personas, organizaciones o nodos están vinculados entre sí mediante lazos invisibles que se entrelazan en una malla de conexiones entrecruzadas, de la misma manera que una red de pesca o un tramo de tela que está hecho de entrelazados y tejidos. Se distinguen dos formas principales de ACSN: la red egocéntrica y la red completa. En los estudios de redes egocéntricas se analiza la red de una persona mientras que, en las redes completas, se intenta encontrar todas las relaciones entre los participantes en la red (Otte & Rousseau, 2002).



Encontrar nodos clave o predecir enlaces entre nodos son problemas fundamentales en el análisis de redes sociales corporativas. Por ello, existen varios métodos que permiten que la red reconstruida tenga una funcionalidad muy cercana a la red real (Lü & Zhou, 2011). En la literatura actual se identificaron dos tipos de algoritmos tradicionales de análisis de redes: las medidas de centralidad o medidas locales (centralidad de grado, intermediación, cercanía, centralidad de Eigenvector, centralidad de Katz y centralidad de Pagerank) (Ghali et al., 2012; Michalski et al., 2012) y las medidas globales (diámetro y radio, distancia promedio, grado promedio, densidad, modularidad) (Kantarci & Labatut, 2013; Michalski et al., 2012). Estas medidas han sido de gran utilidad en diferentes estudios organizacionales para mejorar el desempeño de la empresa.

### 3.1.1 Medidas Locales

Son todas aquellas que están basadas en el concepto de centralidad (redes no dirigidas) y prestigio (redes dirigidas), que corresponden a medidas generales de la posición de un actor en la estructura global de la red social. La principal función está fundamentada en identificar nodos claves y permitir observar cómo las relaciones que conforman la red se concentran en unos pocos individuos, dando una idea de su influencia o poder social. Existen varias medidas de centralidad tales como: grado, intermediación, cercanía, vector propio, centralidad de Katz, centralidad de PageRank. Las tres primeras fueron propuestas por Freeman (LC. 1979).

- **Grado de centralidad:** El grado de centralidad de un nodo se define como el número de vínculos que tiene este nodo (en terminología gráfica-teórica, el número de aristas adyacentes a este nodo). En términos matemáticos, la centralidad de grados,  $d(i)$ , del nodo  $i$  se define como (Otte & Rousseau, 2002):

$$d(i) = \sum_j m_{ij}$$

Donde  $m_{ij} = 1$  si hay un vínculo entre los nodos  $i$  y  $j$ , y  $m_{ij} = 0$  si no existe tal vínculo. El grado de centralidad en una red de  $N$ -nodos se puede estandarizar dividiendo por  $N-1$ :  $d_s(i) = \frac{d(i)}{N(1)}$  (Otte & Rousseau, 2002).

- **Centralidad de la inmediatez (intermediación):** Se puede definir como el número de veces que un nodo necesita un nodo determinado para llegar a otro nodo. Dicho de otra manera, es el número de rutas más cortas que pasan por un nodo determinado. Como expresión matemática, la centralidad de intermediación del nodo  $i$ , denotado como  $b(i)$  se obtiene como (Otte & Rousseau, 2002):



$$b(i) = \sum_{j,k} \frac{g_{jik}}{g_{jk}}$$

Donde  $g_{jk}$  es el número de rutas más cortas desde el nodo  $j$  al nodo  $k$  ( $j, k \neq i$ ), y  $g_{jik}$  es el número de rutas más cortas desde el nodo  $j$  al nodo  $k$  a través del nodo  $i$ . La intermediación mide la medida en que un nodo facilita el flujo en la red. Se puede demostrar que para una red de  $N$ -nodos, el valor máximo de  $b(i)$  es  $(N^2 - 3N + 2)/2$ . Por lo tanto, la intermediación de centralización estandarizada es (Otte & Rousseau, 2002):

$$b_s(i) = \frac{2b(i)}{N^2 - 3N + 2}$$

- **Centralidad de cercanía:** La centralidad de cercanía de un nodo es igual a la distancia total (en el gráfico) de este nodo de todos los demás nodos. Como fórmula matemática, la centralidad de proximidad,  $c(i)$ , del nodo  $i$  se puede escribir como (Otte & Rousseau, 2002):

$$c(i) = \sum_j d_{ij}$$

Donde  $d_{ij}$  es el número de enlaces en una ruta más corta desde el nodo  $i$  al nodo  $j$ . La cercanía es una medida inversa de centralidad en la que un valor más grande indica un actor menos central mientras que un valor más pequeño indica un actor más central. Por esta razón, la cercanía estandarizada se define como  $c_s(i) = \frac{N-1}{c(i)}$ , lo que la convierte nuevamente en una medida directa de centralidad (Otte & Rousseau, 2002).

- **Centralidad del vector propio:** Se define como el vector propio principal de la matriz de adyacencia. Incluso si un vértice tiene algunos vínculos, si esos pocos vértices influyen en muchos otros (quienes ellos mismos influyen aún más en otros), entonces el primer vértice de esa cadena es muy influyente. En una red de enfermedades, si una persona tiene el potencial de contraer una enfermedad de unos pocos vecinos, y esos vecinos tienen un alto riesgo de enfermedad, el riesgo potencial de contraer la enfermedad para la primera persona sigue siendo muy alto (Grassi et al., 2010).

$$\forall v \in V \quad Ceig(v) = x_v$$



Donde  $x_v$  es el  $v$ , el componente del vector propio principal  $x$ , es decir,  $x_v = \frac{1}{p} \sum_{i=1}^n a_{iv} x_i$ ,  $i = 1, \dots, n$ . El vector propio normalizado es  $Ceig(v) / \|x\|_2$ , donde  $\|\cdot\|_2$  es la norma euclidiana (Grassi et al., 2010).

- **Centralidad de Katz:** Métodos de recurrencia que permiten la correlación entre los nodos que participan en la red. La esencia de la centralidad de Katz es calcular la centralidad de un nodo en función de las centralidades del vecino, de estas centralidades se derivan la de Bonacich y Hubbell (1965) (Katz, 1953).

$$x_i = \alpha \sum_j A_{ij} x_j + \beta$$

- **Centralidad de PageRank:** Presentado por Sergey Brin y Larry Page en abril de 1998 como un algoritmo básico utilizado principalmente para asignar numéricamente, la relevancia de documentos o páginas web, poco después fundaron el buscador más grande del mundo Google. El algoritmo PageRank, es uno de los algoritmos de clasificación de páginas más utilizados, establece que si una página tiene enlaces importantes dirigidos a ella (conexiones de entrada), sus enlaces a otras páginas también se vuelven importantes (conexiones de salida). Una versión ligeramente simplificada de PageRank se define como (Chen et al., 2012; Wu et al., 2008):

$$PR(u) = c \sum_{v \in B(u)} \frac{PR(v)}{N_v}$$

Donde  $u$  representa una página web.  $B(u)$  es el conjunto de páginas que apuntan a  $u$ .  $PR(u)$  y  $PR(v)$  son puntuaciones de clasificación de la página  $u$  y  $v$ , respectivamente.  $N_v$  denota el número de enlaces salientes de la página  $v$ .  $c$  es un factor utilizado para la normalización. En PageRank, la puntuación de clasificación de una página,  $p$ , se divide uniformemente entre sus enlaces salientes. Los valores asignados a los enlaces salientes de la página  $p$  se utilizan a su vez para calcular los rangos de las páginas a las que apunta la página  $p$ . Las puntuaciones de clasificación de las páginas de un sitio web se pueden calcular de forma iterativa a partir de cualquier página web. Dentro de un sitio web, dos o más páginas pueden conectarse entre sí para formar un bucle. Si estas páginas no se refirieran a otras páginas web fuera del bucle, sino que se refieran a ellas, acumularían rango, pero nunca distribuirían ningún rango. Este escenario se denomina sumidero de rango (Chen et al., 2012; Wu et al., 2008).



### 3.1.2 Medidas globales

Aportan información más compacta que la que permite evaluar la estructura global de la red, proporcionando información sobre las propiedades importantes de los fenómenos sociales subyacentes. Entre las medidas globales se encuentran: diámetro y radio, distancia promedio, calificación promedio, densidad y modularidad (Kantarci & Labatut, 2013).

- **Diámetro y radio:** El diámetro  $D(G)$  es la excentricidad máxima sobre la red, es decir, la distancia máxima entre dos nodos en  $G$  (Newman & Girvan, 2004):

$$D(G) = \max_{u \in G} e(u)$$

Por el contrario, el radio  $R(G)$  es la excentricidad mínima de la red, es decir (Newman & Girvan, 2004):

$$R(G) = \min_{u \in G} e(u)$$

- **Distancia promedio:** La distancia geodésica  $d(u, v)$  entre dos nodos  $u$  y  $v$  corresponde a la longitud del camino más corto entre ellos. La distribución de la distancia se ha estudiado especialmente en el contexto de redes informáticas como Internet. Cuando aumenta logarítmicamente con el tamaño de la red, el sistema tiene la llamada propiedad de mundo pequeño (Newman, 2003).
- **Calificación promedio:** Además de las medidas mencionadas, que son globales por construcción, también se considera como medidas globales los promedios de las medidas locales, las cuales son: grado de centralidad, centralidad de intermediación, centralidad de cercanía, centralidad del vector propio, centralidad de Katz y centralidad de PageRank (Kantarci & Labatut, 2013).
- **Densidad:** Esta medida global indicada como  $\delta(G)$  corresponde a la relación entre enlaces existentes y posibles en la red  $G$ . Va desde 0 (ningún enlace) a 1 (todos los nodos están conectados). Las redes del mundo real generalmente se consideran muy escasas, con densidades cercanas a 0 y 1 (Dehmer, 2011).

$$\delta(G) = \frac{m}{n(n-1)}$$



- **Modularidad:** Es una medida de la estructura de redes o gráficos. Fue diseñado para medir la fuerza de la división de una red en módulos (también llamados grupos o comunidades). Las redes con alta modularidad tienen conexiones sólidas entre los nodos dentro de los módulos, pero pocas conexiones entre nodos en diferentes módulos (Kantarci & Labatut, 2013).

### 3.2 Minería de datos.

La revolución digital ha hecho posible que la información digitalizada sea fácil de capturar, procesar, almacenar, distribuir, y transmitir. Con el importante progreso en informática, en las tecnologías relacionadas y la expansión de su uso en diferentes aspectos de la vida, se continúa recogiendo y almacenando en bases de datos gran cantidad de información. Descubrir conocimiento de este enorme volumen de datos es un reto en sí mismo. La minería de datos (DM por sus siglas en inglés) es un intento de buscarle sentido a la explosión de información que actualmente puede ser almacenada (Riquelme et al., 2006).

La minería de datos es una subdisciplina de apoyo en las demás disciplinas de las ciencias. Su fortaleza radica en el hecho de que forma parte del proceso de descubrimiento del conocimiento, cuyo objetivo es la búsqueda de patrones de datos que sean válidos, novedosos, potencialmente útiles y comprensibles (Riquelme et al., 2006).

La DM es el descubrimiento de estructuras y patrones en conjuntos de datos grandes y complejos. Hay dos aspectos de la minería de datos: creación de modelos y detección de patrones. La construcción de modelos en la minería de datos es muy similar al modelado estadístico, aunque surgen nuevos problemas debido al gran tamaño de los conjuntos de datos y al hecho de que la minería de datos es a menudo un análisis de datos secundarios. La detección de patrones busca anomalías o pequeñas estructuras locales en los datos, siendo la gran masa de datos irrelevante (Hand & Adams, 2015). De hecho, un punto de vista de muchas actividades de minería de datos a gran escala es que constituyen principalmente filtrado y reducción de datos. Aunque algunas subdisciplinas de estadística han examinado casos especiales de este problema, la mayor parte del trabajo sobre detección de patrones hasta la fecha ha sido computacional, con énfasis en algoritmos (Hand & Adams, 2015).

En redes complejas, es necesario implementar algoritmos que puedan descubrir y analizar patrones en grandes volúmenes de datos. Es por ello que se han creado un sinnúmero de métodos de minería de datos, que facilitan a las organizaciones e investigaciones procesar la información y permiten realizar el modelo respectivo para el análisis de los resultados. Desde finales de la década de 1980, investigadores de comunidades, algoritmos y bases de datos de inteligencia artificial han desarrollado varios algoritmos de minería de datos. En la conferencia internacional IEEE de 2006 sobre minería de datos (ICDM), se identificaron los



10 algoritmos de minería de datos más influyentes, los cuales son: algoritmos genéticos, búsqueda tabú, recocido simulado, redes neuronales artificiales, árboles de decisión, K-means, A priori, EM, AdaBoost y el Naive Baye (Chen et al., 2012; Wu et al., 2008).

Estos métodos son muy influyentes y de tendencia en el estudio de la clasificación en organizaciones, células sociales y entornos turbulentos; todas estas herramientas han sido utilizadas por la estadística para el análisis de datos, en la clasificación de patrones y la estimación de variables, que aún se encuentran en un desarrollo constante para la explotación de la sociedad (Mitchell, 1997; G. P. Zhang, 2000).

- **Algoritmos genéticos:** Los algoritmos genéticos (GAs) fueron inventados por John Holland, quien desarrolló esta idea en su libro “Adaptación en sistemas naturales y artificiales” en el año 1975. Holland propuso GA como un método heurístico basado en “Supervivencia del más apto. GA se descubrió como una herramienta útil para problemas de búsqueda y optimización (Sivanandam & Deepa, 2008).
- **Búsqueda tabú (TS):** Surge la búsqueda tabú, en un intento de proporcionar "inteligencia" a los algoritmos de búsqueda locales. Según Glover & Laguna (1998), su primer definidor, “la búsqueda de Tabú guía es un procedimiento de búsqueda local para explorar el espacio de soluciones más allá del óptimo local”. La búsqueda tabú toma el concepto de memoria de la Inteligencia Artificial y lo implementa a través de estructuras simples con el objetivo de dirigir la búsqueda teniendo en cuenta su historia, es decir, el procedimiento intenta extraer información de lo sucedido y actuar en consecuencia. En este sentido se puede decir que hay algún aprendizaje y que la búsqueda es inteligente (Glover & Laguna, 1998). Este método ha logrado impresionantes éxitos prácticos en aplicaciones que van desde la programación y el equilibrio de canales informáticos hasta el análisis de conglomerados y la planificación espacial, y más recientemente ha demostrado su valor en el tratamiento de problemas clásicos (Glover, 1989).
- **Recocido simulado:** El recocido simulado es un algoritmo de búsqueda metaheurística para problemas de optimización global; el objetivo general de este tipo de algoritmo es encontrar una buena aproximación al valor óptimo de una función en un gran espacio de búsqueda. Este valor óptimo se denomina "óptimo global" (Kirkpatrick et al., 1983). Según Carley & Svoboda (1996), la adaptación organizacional se produce a través de la reestructuración y el aprendizaje. Las organizaciones se pueden modelar utilizando un método de doble nivel en el que la reestructuración se modela como un proceso de recocido simulado y el aprendizaje individual se basa en técnicas de aprendizaje estocástico y agentes con límites racionales (Mitchell, 1997).



- **Redes neuronales Artificiales (ANN)** : La red neuronal es muy apropiada para resolver problemas de minería de datos por sus características de buena robustez, adaptación auto organizada, procesamiento en paralelo y tolerancia a fallas (Arif et al., 2015). El método de la red neuronal se utiliza para la clasificación, agrupación, extracción de características, predicción y reconocimiento de patrones (Gaur, 2012).

El estudio de las redes neuronales artificiales (ANN) se ha inspirado en parte en la observación de los sistemas de aprendizaje biológico que están formados por redes muy complejas de neuronas interconectadas. En una analogía aproximada, las redes neuronales artificiales se construyen a partir de un conjunto de unidades simples densamente interconectadas, donde cada unidad toma una cantidad de entradas de valor real (posiblemente las salidas de otras unidades) y produce una única salida de valor real (que puede convertirse en la entrada a muchas otras unidades) (Mitchell, 1997). El objetivo de las redes neuronales artificiales es hacer un modelo muy simplificado del cerebro humano. De esta manera, las redes neuronales artificiales intentan aprender tareas (para resolver problemas) que imitan el comportamiento del cerebro (Arif et al., 2015).

Las características especiales de los sistemas de computación neuronal permiten que esta nueva técnica de cálculo sea utilizada en una extensa variedad de aplicaciones. La computación neuronal provee un acercamiento mayor al reconocimiento y percepción humana que los métodos tradicionales de cálculo (Mitchell, 1997). Las redes neuronales artificiales presentan resultados razonables en aplicaciones donde las entradas presentan ruido o las entradas están incompletas (Basogain, 2015). Una de las principales características de las ANN es su capacidad de aprendizaje. El entrenamiento de las ANN muestra algunos paralelismos con el desarrollo intelectual de los seres humanos. No obstante, aun cuando parece que se ha conseguido entender el proceso de aprendizaje conviene ser moderado porque el aprendizaje de las ANN está limitado (Basogain, 2015).

Gaur (2012) por su parte menciona la subclasificación que puede tener el modelo de red neuronal.

- **Redes de avance:** Considera el modelo de retropropagación, de percepción y la red de funciones como representantes, y se utiliza principalmente en áreas como la predicción y el reconocimiento de patrones.
- **Red de realimentación o comentarios:** Considera el modelo discreto Hopfield y el modelo continuo como representantes, y se utiliza principalmente para la memoria asociativa y el cálculo de optimización.
- **Redes de autoorganización:** considera el modelo de teoría de la resonancia adaptativa (ART) y el modelo Kohonen como representantes, y se utiliza principalmente para el análisis de conglomerados.



Uno de los algoritmos para abordar las redes neuronales es el SVM (Support Vector Machine), que constituyen un método basado en aprendizaje para la resolución de problemas de clasificación y regresión. En ambos casos, esta resolución se basa en una primera fase de entrenamiento (donde se informa con múltiples ejemplos ya resueltos, en forma de pares: problema, solución) y una segunda fase de uso para la resolución de problemas. En ella, las SVM se convierten en una “caja negra” que proporciona una respuesta (salida) a un problema dado (entrada) (Mukkamala et al., 2002). La teoría SVM se basa en la minimización del riesgo estructural y en la última década ha sido ampliamente utilizada para explicar los comportamientos de redes complejas, como las neuronales o las redes sociales corporativas (Vapnik, 2010), dado que sus escenarios muestran claros rasgos de análisis de comportamiento y clasificación de estructuras complejas, donde la información en los datos de entrada es muy robusta y permite una gran cantidad de análisis de datos. Las SVM forman parte de los 10 algoritmos de minería de datos más influyentes (Chen et al., 2012; Wu et al., 2008).

Las máquinas de vectores de soporte (SVM) se consideran recientemente muy exitosas en el campo del aprendizaje automático y la clasificación de patrones (Vapnik, 2010). Una de las principales ventajas del SVM sobre otras redes es que su entrenamiento se realiza mediante la solución de un problema de programación cuadrática convexa con restricciones lineales: por lo tanto, solo existe un mínimo global (no necesariamente específico) y, dada una tolerancia fija, estos algoritmos eficientes pueden encontrar una solución aproximada en un número finito de pasos (Anguita & Boni, 2002; Keerthi & Gilbert, 2002).

Por otra parte, El K-NN (K-Nearest Neighbor) es otro algoritmo perteneciente a la familia de redes neuronales. Es un método no paramétrico para clasificar las observaciones en uno o más grupos basándose en una o más variables cuantitativas. La asignación grupal de una observación se decide mediante las asignaciones de grupo de su primer vecino más cercano. Este método tiene la ventaja de aproximar mejor la distribución de la muestra al dividir el espacio variable en cualquier número arbitrario de regiones de decisión con el máximo limitado por el número total de observaciones (Tam & Kiang, 1992). El K-NN es un método de clasificación poderoso capaz de lidiar con problemas arbitrariamente complejos, siempre que haya un gran conjunto de datos de entrenamiento (Jiang & Zhou, 2004). Este algoritmo es muy eficiente cuando existe poco o ningún conocimiento previo sobre la distribución de datos, el método K-NN debe ser una de las primeras opciones de clasificación y forma parte del grupo de algoritmos más eficientes dentro de la minería de datos (Chen et al., 2012; Parvin et al., 2008; Wu et al., 2008).

La era de la información ha permitido a muchas organizaciones recopilar grandes cantidades de datos. La clasificación de escenarios motivacionales juega un papel importante en la minería de datos y la clasificación K-NN es un tipo de algoritmo de clasificación diferida que



ofrece muchas ventajas (Xiong et al., 2007), tales como: fácil de usar, datos de entrenamiento robustos a ruidosos, especialmente si el cuadrado inverso de la distancia ponderada se usa como medida de "distancia", y efectivo si los datos de entrenamiento son grandes (Parvin et al., 2008).

- Los **métodos pertenecientes a la familia de árboles de decisión** son: C4.5, ID3, CHAID, MARS, Árboles de inferencia condicional y el CART. Los árboles de decisión son clasificadores efectivos en una variedad de dominios. La elección de una prueba "mejor" es lo que hace que este algoritmo sea codicioso (Murthy & Salzberg, 1995). La optimización de un árbol de decisiones se puede medir en términos de predicción, precisión, tamaño o profundidad. Debe quedar claro que es deseable construir árboles óptimos en términos de uno o más de estos criterios (Murthy & Salzberg, 1995).
- El **K-means**: es un algoritmo de agrupamiento utilizado con una amplia gama de aplicaciones (C. Zhang & Fang, 2012). Se utiliza en minería de datos y reconocimiento de patrones con el fin de minimizar el índice de desempeño del clúster, el error cuadrático y el criterio de error que son la base de este algoritmo (Y. Li & Wu, 2012).
- El algoritmo **A priori**: es aplicado para extraer patrones repetidos del conjunto de datos de transacciones para encontrar conjuntos de elementos frecuentes y asociaciones entre varios conjuntos de elementos (Nafie Ali & Mohamed Hamed, 2018).
- El algoritmo **EM**: es una técnica de agrupación basada en modelos bien establecidos que es simple y directo de implementar e intenta optimizar el ajuste entre los datos dados y algunos modelos matemáticos (Fraley & Raftery, 2002; Masangcap et al., 2018).
- El algoritmo **AdaBoost**: propuesto por Yoav Freund y Robert Schapire, es uno de los métodos de conjuntos más importantes, ya que tiene una base teóricamente sólida, una predicción muy precisa, una gran simplicidad (Freund & Schapire, 1997). Corrige las clasificaciones erróneas hechas por clasificadores débiles y es menos susceptible al sobreajuste que la mayoría de los algoritmos de aprendizaje (Hu et al., 2008).

El algoritmo **Naive Baye (NB)**: es una familia de clasificadores probabilísticos simples basados en un supuesto común de que todas las características son independientes entre sí, dada la variable de categoría, y se usa a menudo como línea de base en la clasificación de textos (Xu, 2018). Sin embargo, el NB no es completamente bayesiano, al menos no en el sentido de que se estima una distribución posterior sobre los parámetros de los documentos de entrenamiento y luego se utiliza para la inferencia predictiva de un nuevo documento (Xu, 2018).



## **4. Metodología**

### **4.1 Tipo, Naturaleza y Enfoque del estudio**

En esta investigación se va a realizar un estudio transversal descriptivo, de naturaleza cualitativa con un enfoque deductivo.

Los estudios transversales descriptivos buscan especificar las propiedades, las características y los perfiles de personas, grupos, comunidades, procesos, objetos o cualquier otro fenómeno que se someta a un análisis (Hernández Sampieri et al., 2010). Este proyecto es transversal descriptivo, puesto que, describe cómo los métodos tradicionales de análisis de redes y minería de datos han sido aplicados a nivel organizacional con el fin de evaluar los miembros que conforman una cadena de suministros según su importancia, participación, centralidad, habilidades y capacidad para tomar decisiones.

Por otro lado, esta monografía es de naturaleza cualitativa, debido a que la recolección de datos no tiene medición numérica para descubrir o afinar preguntas de investigación en el proceso de interpretación (Hernández Sampieri et al., 2010). Además, posee un enfoque deductivo dado que, el razonamiento parte de un marco general de referencia hacia algo en particular, es decir, al momento de realizar la revisión sistémica se seleccionará únicamente los artículos que contribuyan a lograr el objetivo general.

### **4.2 Plan de análisis**

#### **4.2.1 Estrategia y criterios de búsqueda bibliográfica**

En este trabajo se lleva cabo una estrategia de revisión sistémica de estudios que han utilizado métodos de análisis de redes tradicionales y minería de datos en el entorno empresarial, para determinar la importancia de un miembro en una red social corporativa, mediante las bases de datos electrónicas, tales como: Scopus, Science Direct y EBSCO. Los criterios de búsqueda (términos clave tanto en inglés como en español) utilizados para realizar la búsqueda bibliográfica se muestran en las tablas 1 y 2.



Tabla 1. Palabras clave para la búsqueda en las bases de datos académicas definidas en inglés.

<b>keywords for search in academic databases</b>	
<b>Topic</b>	<b>keywords</b>
Social Network Analysis	Importance in organizations
	Key members
	Global measures
	Local measures
	Application to companies
	Importance of nodes
	Centrality of PageRank
	Traditional methods
	Centrality of actors
	Corporate social network
AND	
Data Mining	Classification of nodes
	Learning machine
	Members important
	Digital transformation
	Advanced algorithms
	Organizations transformation
	Supply chain
	Data management development
	Management addressing
AND	
Artificial Neural Networks	Classification of members
	SVM
	KNN
	News algorithms
	Business importance
	Application in corporations
	Artificial intelligence
	Data storage
	Business analysis

Fuente: las autoras



Tabla 2. Palabras clave para la búsqueda en las bases de datos académicas definidas en español.

<b>Palabras clave para la búsqueda en las bases de datos académicas</b>	
<b>Tema</b>	<b>Palabras Clave</b>
Análisis de Redes Sociales	Importancia en las organizaciones
	Miembros clave
	Medidas globales
	Medidas locales
	Aplicación a empresas
	Importancia de los nodos
	Centralidad de PageRank
	Métodos tradicionales
	Centralidad de los actores
	Red social corporativa
Y	
Minería de Datos	Clasificación de nodos
	Aprendizaje automático
	Miembros importantes
	Transformación digital
	Algoritmos avanzados
	Transformación de las organizaciones
	Cadena de suministro
	Desarrollo de gestión de datos
	Direccionamiento gerencial
Y	
Redes Neuronales Artificiales	Clasificación de miembros
	SVM
	KNN
	Nuevos algoritmos
	importancia empresarial
	Aplicación en corporaciones
	Inteligencia artificial
	Almacenamiento de datos
	Análisis empresarial

Fuente: las autoras



Usando una combinación de palabras claves, se emplea una estrategia de búsqueda para garantizar que se identificaran todos los estudios relevantes publicados desde el año 1995 hasta la actualidad. Las combinaciones que se emplearon para la búsqueda tanto en inglés como en español se muestran en las Tablas 3 y 4:

*Tabla 3. Combinaciones en inglés*

<b>Domain</b>	<b>Social Network Analysis</b>
<b>Keywords</b>	<p>"Degree of centrality" AND ("important member in the organization" OR "Identification of a key node in an organization" OR "Corporate network" OR "Links between members in the organization")</p> <p>OR</p> <p>"Centrality of intermediation" AND ("important member in the organization" OR "Identification of a key node in an organization" OR "Corporate network" OR "Links between members in the organization")</p> <p>OR</p> <p>"Closeness centrality" AND ("important member in the organization" OR "Identification of a key node in an organization" OR "Corporate network" OR "Links between members in the organization")</p> <p>OR</p> <p>"Centrality of the eigenvector" AND ("important member in the organization" OR "Identification of a key node in an organization" OR "Corporate network" OR "Links between members in the organization")</p> <p>OR</p>



	<p>“Katz centrality" AND ("important member in the organization" OR "Identification of a key node in an organization" OR "Corporate network" OR "Links between members in the organization")</p> <p>OR</p> <p>“PageRank centrality" AND ("important member in the organization" OR "Identification of a key node in an organization" OR "Corporate network" OR " Links between members in the organization")</p> <p>OR</p> <p>“SNA" AND ("important member in the organization" OR "Identification of a key node in an organization" OR "Corporate network" OR "Links between members in the organization")</p>
<b>Domain</b>	<b>Data Mining</b>
<b>Keywords</b>	<p>"Genetic Algorithms" AND ("important member in the organization" OR "Identification of a key node in an organization" OR "Corporate network" OR "Links between members in the organization")</p> <p>OR</p> <p>"Taboo search" AND ("important member in the organization" OR "Identification of a key node in an organization" OR "Corporate network" OR "Links between members in the organization")</p> <p>OR</p>



	<p>"Simulated annealing" AND ("important member in the organization" OR "Identification of a key node in an organization" OR "Corporate network" OR "Links between members in the organization")</p> <p>OR</p> <p>"Artificial neural networks " AND ("important member in the organization" OR "Identification of a key node in an organization" OR "Corporate network" OR "Links between members in the organization")</p> <p>OR</p> <p>"K-Nearest Neighbor" AND ("important member in the organization" OR "Identification of a key node in an organization" OR "Corporate network" OR "Links between members in the organization")</p> <p>OR</p> <p>"Support vector machine" AND ("important member in the organization" OR "Identification of a key node in an organization" OR "Corporate network" OR "Links between members in the organization")</p> <p>OR</p> <p>"Decision trees" AND ("important member in the organization" OR "Identification of a key node in an organization" OR "Corporate network" OR "Links between members in the organization")</p> <p>OR</p>
--	---



	<p>“CART” AND ("important member in the organization" OR "Identification of a key node in an organization" OR "Corporate network" OR "Links between members in the organization")</p> <p>OR</p> <p>“K-means AND ("important member in the organization" OR "Identification of a key node in an organization" OR "Corporate network" OR "Links between members in the organization")</p> <p>OR</p> <p>“Apriori” AND ("important member in the organization" OR "Identification of a key node in an organization" OR "Corporate network" OR "Links between members in the organization")</p> <p>OR</p> <p>“EM” AND ("important member in the organization" OR "Identification of a key node in an organization" OR "Corporate network" OR "Links between members in the organization")</p> <p>OR</p> <p>“AdaBoost” AND ("important member in the organization" OR "Identification of a key node in an organization" OR "Corporate network" OR "Links between members in the organization")</p>
--	--



	<p>OR</p> <p>“Naive Baye” AND                  ("important member in the organization"                  OR "Identification of a key node in an                  organization" OR "Corporate network" OR                  "Links between members in the                  organization")</p>
--	--

Fuente: las autoras

*Tabla 4. Combinaciones en español*

<b>Dominio</b>	<b>Análisis de redes sociales</b>
<b>Palabras Claves</b>	<p>"Grado de centralidad " AND                  ("Miembro importante en la organización"                  OR "Identificación de un nodo clave en una                  organización" OR "Red corporativa" OR                  "Vínculos entre los miembros de la                  organización")</p> <p>OR</p> <p>"Centralidad de intermediación " AND                  ("miembro importante de la organización"                  OR "Identificación de un nodo clave en una                  organización" OR "Red corporativa" OR                  "Vínculos entre los miembros de la                  organización")</p> <p>OR</p> <p>"Centralidad de cercanía" AND                  ("miembro importante de la organización"                  OR "Identificación de un nodo clave en una                  organización" OR "Red corporativa" OR                  "Vínculos entre los miembros de la                  organización")</p>



	<p>OR</p> <p>"Centralidad del vector propio" AND ("miembro importante de la organización" OR "Identificación de un nodo clave en una organización" OR "Red corporativa" OR "Vínculos entre los miembros de la organización")</p> <p>OR</p> <p>“Centralidad de Katz” AND ("miembro importante de la organización" OR "Identificación de un nodo clave en una organización" OR "Red corporativa" OR "Vínculos entre los miembros de la organización")</p> <p>OR</p> <p>"Centralidad De PageRank" AND ("miembro importante de la organización" OR "Identificación de un nodo clave en una organización" OR "Red corporativa" OR "Vínculos entre los miembros de la organización")</p> <p>OR</p> <p>"SNA" AND ("miembro importante de la organización" OR "Identificación de un nodo clave en una organización" OR "Red corporativa" OR "Vínculos entre los miembros de la organización")</p>
--	---



Dominio	Minería de Datos
<p><b>Palabras Claves</b></p>	<p>"Algoritmos Genéticos " AND                      ("Miembro importante en la organización"                      OR "Identificación de un nodo clave en una organización" OR "Red corporativa" OR                      "Vínculos entre los miembros de la organización")</p> <p>OR</p> <p>"Búsqueda tabú " AND                      ("miembro importante de la organización"                      OR "Identificación de un nodo clave en una organización" OR "Red corporativa" OR                      "Vínculos entre los miembros de la organización")</p> <p>OR</p> <p>"Recocido simulado" AND                      ("miembro importante de la organización"                      OR "Identificación de un nodo clave en una organización" O "Red corporativa" OR                      "Vínculos entre los miembros de la organización")</p> <p>OR</p> <p>"Redes neuronales artificiales" AND                      ("miembro importante de la organización"                      OR "Identificación de un nodo clave en una organización" OR "Red corporativa" OR                      "Vínculos entre los miembros de la organización")</p> <p>OR</p> <p>"K-NN" AND                      ("miembro importante de la organización"                      OR "Identificación de un nodo clave en una organización" OR "Red corporativa" OR                      "Vínculos entre los miembros de la organización")</p>



	<p>OR</p> <p>"SVM" AND ("miembro importante de la organización" OR "Identificación de un nodo clave en una organización" OR "Red corporativa" OR "Vínculos entre los miembros de la organización")</p> <p>OR</p> <p>"Arboles de decisión " AND ("miembro importante de la organización" OR "Identificación de un nodo clave en una organización" OR "Red corporativa" OR "Vínculos entre los miembros de la organización")</p> <p>OR</p> <p>"CART" AND ("miembro importante de la organización" OR "Identificación de un nodo clave en una organización" OR "Red corporativa" OR "Vínculos entre los miembros de la organización")</p> <p>OR</p> <p>"K-means" AND ("miembro importante de la organización" OR "Identificación de un nodo clave en una organización" OR "Red corporativa" OR "Vínculos entre los miembros de la organización")</p> <p>OR</p> <p>"Apriori" AND ("miembro importante de la organización" OR "Identificación de un nodo clave en una organización" OR "Red corporativa" OR "Vínculos entre los miembros de la organización")</p>
--	--



	<p>OR</p> <p>"EM" AND ("miembro importante de la organización" OR "Identificación de un nodo clave en una organización" OR "Red corporativa" OR "Vínculos entre los miembros de la organización")</p> <p>OR</p> <p>"AdaBoost " AND ("miembro importante de la organización" OR "Identificación de un nodo clave en una organización" OR "Red corporativa" OR "Vínculos entre los miembros de la organización")</p> <p>OR</p> <p>"Arboles de decisión " AND ("miembro importante de la organización" OR "Identificación de un nodo clave en una organización" OR "Red corporativa" OR "Vínculos entre los miembros de la organización")</p>
--	--

Fuente: las autoras



#### 4.2.2 Selección y exclusión de literatura

Para llevar a cabo el proceso de selección de la literatura, en primer lugar, se realiza una búsqueda tanto en inglés como en español que abarca todas las palabras clave y criterios estipulados; esta revisión se efectúa en todas las bases de datos definidas. Adicionalmente, se considera fuentes externas como apoyo para lograr una recopilación de la información más completa y precisa que contribuya a alcanzar el objetivo general.

Posteriormente, se excluye registros que no estén relacionados con el tema, no sean confiables y se desvíen totalmente de los parámetros de búsqueda. Lo anterior permite obtener una revisión sistemática potencialmente relevante. Sin embargo, es pertinente limitar aún más los artículos que serán seleccionados. Para llegar a esto, se rechaza los estudios que no estén orientados a organizaciones y a las relaciones colaborativas entre nodos. Además, dichos estudios tienen que exponer de forma contundente la aplicabilidad de alguno de los métodos relacionados con el análisis de redes sociales o la minería de datos para que sea información de gran utilidad.

Por consiguiente, se verifica de cada uno de los registros: el título, resumen, metodología y resultados, con el fin de aminorar la información insignificante y de poco aporte para el estudio. Por otro lado, se efectúa una revisión de las referencias bibliográficas pertenecientes a los artículos seleccionados. Estos artículos son analizados para determinar si cumplen con los criterios de búsqueda.

Todo el procedimiento descrito anteriormente, se visualiza en la figura 1. Además, este es la base fundamental para realizar un barrido pertinente de la literatura y selección de los artículos que permitirán alcanzar el fin buscado.

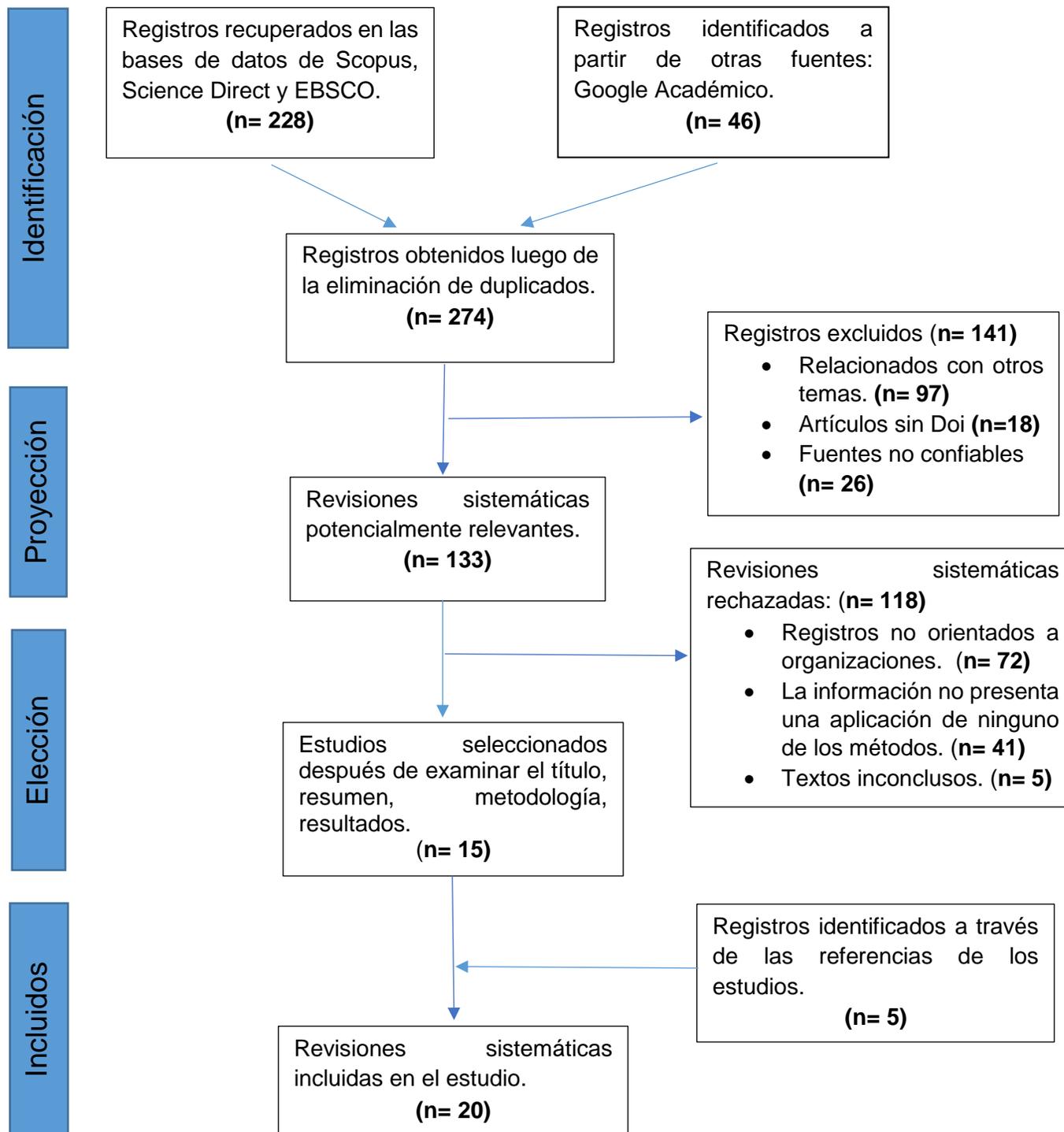


Figura 2. Diagrama de flujo PRISMA

Fuente: las autoras



### 4.2.3 Extracción de datos

Una vez definida la estrategia de búsqueda se determinó un mecanismo de extracción de datos para la selección de aquella literatura que satisfaga los objetivos principales del proyecto. Ese mecanismo utilizó una tabla que sintetiza los aspectos más relevantes de cada investigación a través de los cuales se puede conocer ampliamente la temática de estudio de cada uno y compararla con alguno de los criterios definidos en la estrategia de búsqueda.

En la *Tabla 5* se puede observar la revisión y búsqueda bibliográfica que se realizó a profundidad de todos los estudios incluidos y extracciones de acuerdo con la herramienta y criterios descritos. A partir de la información consolidada se comparó la extracción de la información, se clasificó y, posteriormente, se analizó según el propósito planteado y resultados alcanzados.

*Tabla 5. Descripción de datos extraídos de estudios de modelos de análisis de redes sociales y minería de datos.*

Datos extraídos	Descripción
<b>SNA Y DM</b>	
Tipo de modelo	Análisis de redes sociales y/o Minería de datos
Clase de modelo	Medidas locales: medidas de centralidad, medidas globales, algoritmos genéticos, búsqueda tabú, recocido simulado, redes neuronales artificiales (SVM Y K-NN), árboles de decisión, K-means, algoritmo Apiori, EM, AdaBoost, Naive Baye, combinados, otros, poco claro.
Enfoque organizacional	Si o No
Base para la construcción del modelo	Marco conceptual, opinión de expertos, construcción de modelos grupales, literatura, combinados, otros, poco claro.
Tipo de organización	Definir la organización y describir el objeto social.
VARIABLES CLAVES	Definición de variables claves.
Recolección de datos	Describir el método de recolección de los datos utilizado. Estos pueden ser: bases de datos, encuestas, entrevistas, métodos de observación, revisión de documentos históricos.
Modelo matemático	Si o No. En caso afirmativo, ¿se proporcionó una justificación para el (los) modelo (s) matemático (s) / forma (s) funcional (es) utilizada (s)?



Datos extraídos	Descripción
<b>SNA Y DM</b>	
Software utilizado	Si o No. En caso afirmativo, nombrar el software utilizado y su aporte a la investigación.
Construcción de la red	Si o No
Validación	Si o No. En caso afirmativo, seleccione el proceso utilizado para validar el comportamiento del modelo: medidas estadísticas utilizadas para comparar modelos y datos, comparación cuantitativa o cualitativa entre modelos y resultados, confiabilidad de los datos recolectados y métodos empíricos.
Número de redes construidas	Número total de redes construidas.
Descripción de las redes construidas	Describir las redes construidas: cantidad de nodos, relaciones entre nodos, grados de jerarquía, interconexiones, nodos claves y cantidad de enlaces de entrada y salida.
Resultados	Describir los resultados relevantes para la revisión y su contribución en la red corporativa.
Reestructuración del sistema organizacional	Si o No. En caso afirmativo, consiste en plantear propuestas a partir del análisis de resultados, que cambien el funcionamiento de la red corporativa para mejorar la productividad, rendimiento y ambiente laboral.
Limitaciones del estudio y recomendaciones para futuras investigaciones	Limitaciones claves de la discusión como se describe por los autores.
Referencias Bibliográficas relacionadas con el tema	Si o No. En caso afirmativo, consiste en realizar la revisión de las referencias citadas y determinar si cumplen con los parámetros de búsqueda.

Fuente: las autoras



#### **4.2.4 Análisis de datos e interpretación de los resultados.**

Al momento de haber llevado a cabo la depuración y selección de los artículos base para la estructuración del proyecto, se procedió a realizar una tabla que sintetiza cada uno de los registros elegidos con el fin de clasificarlos según el grado de aporte que tuvieron. Por otro lado, con esta tabla se realizó un bosquejo de toda literatura encontrada en las diferentes bases de datos acerca de los nuevos métodos que las empresas están adoptando para el mejoramiento del flujo de comunicación e información dentro de las relaciones colaborativas que surgen en una organización tanto a nivel interno como externo.

A partir de la interpretación de los resultados y la clasificación de la información recopilada, se efectuó una descripción de cada uno de los métodos y su importancia en el campo empresarial, así mismo, se planteó la evolución que han tenido las empresas por causa de la transformación digital y que ha dado lugar a que estas puedan tener mayor capacidad en la gestión y almacenamiento de datos.

La evaluación de los proyectos seleccionados se centra en entregar una visión general de lo que arrojaron los criterios de inclusión y la búsqueda de literatura con los artículos escogidos. Para esto se elaboró un análisis de las variables escogidas, se describieron aspectos compartidos y conclusiones definidas según la similitud y los enfoques prácticos de cada investigación consultada, las variables, aspectos, características y criterios que determinan la importancia de un miembro en una red social corporativa, propósitos, patrones validados, construcción de redes basadas en el análisis de redes sociales y construcción de redes con los métodos de la minería de datos. Estos puntos permitieron dar una visión general del estado del arte alrededor de esta investigación.



## 5. Resultados

La búsqueda en las diferentes bases de datos definidas arrojó inicialmente un total de 274 artículos de los cuales 97 fueron excluidos por estar relacionados con otros temas, 18 por ser poco representativos, debido a que no contaban con el registro DOI, que es un identificador para las publicaciones electrónicas. Por otro lado, no se tuvo en cuenta 26 registros por ser consignados en fuentes no confiables o que son muy subjetivas, es decir, pertenecientes a revistas o bases de datos poco distinguidas y prestigiosas, obteniendo 133 artículos como referencia. Lo anterior, permitió obtener una revisión sistemática potencialmente relevante. Sin embargo, fue pertinente limitar aún más la búsqueda, para llegar a esto, se rechazó 118 estudios que no fueron orientados a organizaciones, a las relaciones colaborativas entre nodos, que no expusieron de forma contundente la aplicabilidad de alguno de los métodos relacionados con el análisis de redes sociales o la minería de datos para que sea información de gran utilidad, dejando un total de 15 artículos, a los que se adicionaron 5, que fueron seleccionados en el proceso de la revisión bibliográfica, logrando un total de 20 documentos los cuales constituyeron el insumo clave para la materialización de este proyecto.

Todos los estudios contemplados se centran en un ámbito organizacional, por lo tanto, las redes sociales y colaborativas entre los actores son de carácter empresarial. Así mismo, reconocen que para realizar una investigación se debe incluir a todos los empleados independientemente de su grado de jerarquía, es decir, todos los miembros que conforman la entidad deben ser involucrados e interconectados en un mismo sistema informativo. Puesto que de alguna forma cada uno puede influenciar o impactar en el desempeño. Por esta razón, es esencial identificar los aspectos, criterios y destrezas que convierten a un empleado en un nodo clave.



## 5.1 Propósito

Los estudios seleccionados para realizar esta revisión sistémica se enfocan en aplicar diferentes modelos en el ámbito empresarial y de este modo, construir la estructura organizacional e identificar la importancia de cada miembro perteneciente a la red. De las veinte investigaciones que conforman este trabajo (Albino & Garavelli, 1998; Chan & Chung, 2004; Crispim & Pinho de Sousa, 2009; de Toni & Nonino, 2010; Fire et al., 2013; Grassi et al., 2010; HATALA, 2006; Herrera, 2013; Hoppe & Reinelt, 2010; Kamali et al., 2020; León et al., 2017; N. Li et al., 2016; Luna, 2015; Nurek & Michalski, 2020; Olguin et al., 2009; Silva et al., 2009; Yu Sun et al., 2017; Yuan Sun et al., 2015; Yelupula & Ramaswamy, 2008; Zbieg et al., 2016), nueve de ellas emplean los modelos basados en el análisis de redes sociales corporativas (de Toni & Nonino, 2010; Fire et al., 2013; Grassi et al., 2010; HATALA, 2006; Hoppe & Reinelt, 2010; León et al., 2017; Luna, 2015; Yuan Sun et al., 2015; Zbieg et al., 2016) y once se enfocan en los métodos de la minería de datos (Albino & Garavelli, 1998; Chan & Chung, 2004; Crispim & Pinho de Sousa, 2009; Herrera, 2013; Kamali et al., 2020; N. Li et al., 2016; Nurek & Michalski, 2020; Olguin et al., 2009; Silva et al., 2009; Yu Sun et al., 2017; Yelupula & Ramaswamy, 2008).

Por otro lado, tres de los nueve artículos de SCNA involucran todas las medidas locales y globales (de Toni & Nonino, 2010; HATALA, 2006; Hoppe & Reinelt, 2010), cinco utilizan las medidas de centralidad tales como: grado, intermediación, cercanía, vector propio y Katz (Fire et al., 2013; Grassi et al., 2010; León et al., 2017; Luna, 2015; Zbieg et al., 2016), y uno emplea la centralidad de PageRank (Yuan Sun et al., 2015).

Con respecto a los métodos de la minería de datos, de los once estudios seleccionados, dos combinan diferentes técnicas de minería de datos (Herrera, 2013; Silva et al., 2009), dos realizan su análisis con base en los algoritmos genéticos (Chan & Chung, 2004; Kamali et al., 2020), otro utiliza la búsqueda tabú (Crispim & Pinho de Sousa, 2009), tres relacionan las redes neuronales artificiales como: SVM Y KNN (Albino & Garavelli, 1998; N. Li et al., 2016; Olguin et al., 2009), uno aplica el método de recocido simulado (Yu Sun et al., 2017), otro usa el K-means (Yelupula & Ramaswamy, 2008) y el último incluye los árboles de decisión y las redes neuronales como el SVM (Nurek & Michalski, 2020).

Ahora bien, las veinte investigaciones hacen hincapié en obtener resultados contundentes, verídicos y satisfactorios en las redes corporativas que se evalúan. Lo que buscan los autores es mejorar el ambiente laboral, aumentar la productividad, eficiencia, fomentar el trabajo colaborativo, participativo, modificar la estructura de cada empresa, realizar un mejor aprovechamiento del conocimiento y obtener de forma más rápida y eficiente flujos de información. Por otra parte, dichos proyectos tienen como fin destacar los nodos importantes de la red y que no cuentan con roles de equipo de Trabajo (RET) calificados, es decir,



pretenden desarticular y reestructurar el sistema jerárquico de la compañía para que cada miembro sea reconocido y valorado de acuerdo con su contribución, desempeño y actividades diarias que ejecuta.

## 5.2 Validación

Tres de los nueve estudios de SCNA (Fire et al., 2013; León et al., 2017; Yuan Sun et al., 2015) y ocho de las once investigaciones correspondientes a los métodos de minería de datos (Albino & Garavelli, 1998; Crispim & Pinho de Sousa, 2009; Kamali et al., 2020; N. Li et al., 2016; Nurek & Michalski, 2020; Silva et al., 2009; Yu Sun et al., 2017; Yelupula & Ramaswamy, 2008) realizaron validación. Un estudio de SCNA (León et al., 2017) para asegurar la confiabilidad y validez de los datos, utilizó varias medidas. Primero, entrevistó a los administradores del sistema para decidir si los datos habían sido manipulados o no, en segundo lugar, determinó si se habían registrado anomalías como truncamientos, pasos y cambios de fase no razonables. Dado que ambas técnicas ofrecieron un resultado positivo, los datos se consideraron fiables y válidos. Por otro lado, los autores de esta investigación realizaron una triangulación para validar las mediciones; Se aplicó los métodos clásicos del análisis de remuestreo estadístico (análisis de conglomerados, análisis factorial, regresión poisson, regresión no lineal y suavizada, series de tiempo), que se basan en las fórmulas matemáticas para describir la precisión de estadísticas de muestra. Así mismo, se utilizó la técnica de remuestreo Bootstrap, que significa “Salir adelante por uno mismo o levantarse mediante el propio esfuerzo”. Este método utiliza la distribución de estadísticas en sí para analizar la precisión de las estadísticas. Los métodos clásicos asumen que todas las observaciones son independientes, mientras que el segundo construye redes usando varios subconjuntos de nodos. Ambos métodos ofrecieron resultados estadísticamente significativos.

Uno de los artículos que implementó la medida de centralidad de PageRank del SCNA (Yuan Sun et al., 2015), para verificar el rendimiento del algoritmo propuesto, validó la precisión de los resultados, comparando la centralidad de PageRank con otros tres métodos, y analizó el efecto de rendimiento de cada algoritmo. Este resultado experimental, permitió determinar que el modelo PageRank es más preciso, eficiente, tiene una alta escalabilidad y presenta mayor rendimiento en comparación con las demás medidas de centralidad del SCNA. Así mismo, la investigación que se basa en el SCNA realizada por los autores Fire et al. (2013) ejecutó la recolección y clasificación de los datos manualmente, esta información fue suministrada por los empleados de las organizaciones, lo cual garantizó que la información fuera verídica y relevante. Por otro lado, se verificó los resultados de las medidas de centralidad y los de los algoritmos de aprendizaje automático a medida que avanzaban en la



detección de roles influyentes ocultos dentro de la empresa. Los datos fueron manipulados de forma anónima, reemplazando aleatoriamente los ID de Facebook de los empleados con una serie de números enteros contiguos.

Por otro lado, uno de los ocho estudios enfocados a las técnicas de minería de datos y que implementó un algoritmo genético denominado DINGA (Kamali et al., 2020) para validar la veracidad de los resultados y su precisión, consideró determinar el peso de cada criterio de forma completamente aleatoria. Para ello, ponderaron los criterios cien veces arbitrariamente. Luego, calcularon la precisión promedio cien veces utilizando una ponderación aleatoria. La precisión de la media en Enron fue del 60% respectivamente. La comparación de los resultados del algoritmo genético con la ponderación aleatoria indica una mejora de la precisión del 22% en los conjuntos de datos de la red de interacción Enron.

Así mismo, el artículo que utiliza el algoritmo de búsqueda tabú (Crispim & Pinho de Sousa, 2009) para validar sus datos calculó la correlación entre los criterios evaluados con el objetivo de verificar si los criterios elegidos deben cambiarse o arrojan resultados precisos (se considera que correlaciones menores a 0.30 indican poca o ninguna relación entre las variables). En este ejemplo los criterios seleccionados no presentan interdependencias significativas, puesto que todas las correlaciones fueron superiores a 0,30. Posteriormente, el estudio que se apoya en los métodos de árboles de decisión, redes neuronales y SVM (Nurek & Michalski, 2020) entrenó un modelo con el uso del algoritmo de búsqueda de cuadrícula con validación cruzada de 5 veces, por lo que se probaron todas las combinaciones posibles del rango de valores dados, y la mejor combinación con respecto a la puntuación más alta, devolvió el valor del promedio macro. Se verificó como la actividad de una persona puede influir en el resultado de la clasificación y, por lo tanto, se examinó si los meses en los cuales hubo un mayor intercambio de mensajes entre los empleados arrojan mejores resultados. En esta investigación se obtuvo que la clasificación de los miembros de la red social corporativa realizada por los métodos de ANN Y SVM tienen una precisión del 100%, mientras que los resultados obtenidos mediante el método de árboles de decisión cuentan con un índice de confiabilidad del 80%.

La investigación que incluye el algoritmo de recocido simulado perteneciente a la minería de datos (Yu Sun et al., 2017) para validar la efectividad del método, se probó en diferentes redes reales y redes artificiales. Además, se realizó una comparación del método seleccionado con otros dos métodos recientemente desarrollados que son: el método de clasificación basado en la entropía de mapeo (MME) y el método de clasificación basado en atributos múltiples (MMA). Según diferentes métodos, se obtuvo diferentes secuencias de importancia, por lo tanto, eliminar nodos importantes en redes complejas causó un gran daño a las redes. En otras palabras, si los nodos de una red compleja se eliminan uno por uno de acuerdo con su orden en una secuencia importante, la red compleja colapsará rápidamente.



En este estudio se obtuvo que la secuencia de nodos importantes que arrojó el método de recocido simulado es más precisa, eficaz y efectiva que los demás métodos.

Por otra parte, el proyecto que se centra en las redes neuronales artificiales (Albino & Garavelli, 1998) realizó la prueba NN, la cual consiste en la verificación del rendimiento de la red. Entre los principales resultados de NN, la generalización, es decir, la capacidad de proporcionar una respuesta adecuada en correspondencia de nueva información, y la tolerancia al ruido, es decir, la capacidad de reconocer información poco clara, se analizaron críticamente en esta fase. Por otro lado, se validó que la implementación de la red neuronal suele requerir otras fases auxiliares, como la elección del software, la integración de NN con otros sistemas de información y el mantenimiento de NN. Además, se comprobó la precisión del enfoque NN mediante una comparación con otros enfoques, como el AHP y la lógica difusa.

Por otro lado, el documento que implementa el método K-means de la minería de datos (Yelupula & Ramaswamy, 2008) como principal limitación es que no realizaron una validación de los datos de la relación que han extraído y analizado. A pesar de que hicieron la validación con la fuente que encontraron en Internet, para realizar una validación evidente de la fuente, necesitan comparar los datos y hallazgos con material de fuentes confiables, como informes y artículos de prensa sobre el caso Enron, cartas y entrevistas con ex empleados, información de otras personas con conocimiento directo de la empresa. Aunque predijeron la estructura organizativa de Enron con el atributo numérico, no pudieron obtener una precisión del 100%, por lo que deben analizar el contenido del correo electrónico para lograr una precisión total.

El estudio que implementa técnicas de minería de datos (Silva et al., 2009), mediante entrevistas y con el apoyo de un cuestionario a los profesores de una de las universidades, se analizaron las respuestas y las compararon con los resultados del enfoque. En este cuestionario se incluía preguntas que permitían identificar algunas características, tales como: si las áreas están interrelacionadas; cómo es la relación entre los investigadores; y consecuentemente, cómo es la relación entre las diferentes instituciones. Además, se realizaron preguntas a los investigadores sobre sus áreas de interés; si trabaja con profesores de otras áreas; si suele ser coautor con profesores de otras instituciones; y otros tipos de preguntas, para mapear la colaboración científica. Lo que arrojó como resultado que la mayoría de los profesores que respondieron el cuestionario dicen tener más publicaciones con profesores de la misma institución que con profesores externos. Cada profesor indicó los nombres del investigador con el que más se le relaciona. Entonces validaron las relaciones fuertes y débiles, así como los puntos críticos en la red social corporativa.



Por último, El artículo que se basa en el KNN (N. Li et al., 2016), el cual es un método de las redes neuronales para verificar la efectividad del algoritmo, en primer lugar, se comparó el algoritmo con otros siete algoritmos de clasificación KNN (Euclidiana, Euclidiana cuadrada, City Block, Canberra, Squared-chord, Squared-chi-square y MKNN), utilizando conjuntos de datos del repositorio de aprendizaje automático de UCI. Se formularon tres ecuaciones de distancia del algoritmo KNN propuesto, y se probaron, respectivamente, las cuales son: método KNN aplicado con: la Entropía-distancia Euclidiana (EEKNN), Entropía-distancia Manhattan (EMKNN) y Entropía-distancia Canberra (ECKNN). Los resultados muestran que EEKNN, EMKNN y ECKNN tienen la precisión de clasificación promedio más alta, correspondiendo a un 87.59%, 87.64% y 85.51%, respectivamente, con el criterio de distancia ponderada.

A continuación, en la tabla 6, se sintetiza toda la información acerca del proceso de validación de los métodos en cada uno de los artículos y los respectivos resultados alcanzados.



Tabla 6. Información sintetizada sobre la validación de los métodos

<b>Análisis de Redes Sociales Corporativas (SCNA)</b>				
<b>Autor</b>	<b>Año</b>	<b>Método</b>	<b>Proceso de validación</b>	<b>Resultados</b>
León, R. – D., Rodríguez-Rodríguez, R., Gómez-Gasquet, P., and Mula, J.	2017	SCNA	<ol style="list-style-type: none"> <li>1. Se entrevistó a los administradores del sistema para decidir si los datos habían sido manipulados y se determinó si se habían registrado anomalías.</li> <li>2. Se realizó una triangulación para validar las mediciones.</li> <li>3. Se aplicó los métodos clásicos del análisis de remuestreo estadístico y se utilizó la técnica de remuestreo Bootstrap</li> </ol>	Los métodos clásicos asumen que todas las observaciones son independientes, mientras que el segundo construye redes usando varios subconjuntos de nodos. Ambos métodos ofrecieron resultados estadísticamente significativos.
Sun, Yuan, Ma, Y., Zhang, F., Ma, Y., and Shen, W	2015	SCNA: medida de centralidad de PageRank	<ol style="list-style-type: none"> <li>1. Se validó la precisión de los resultados, comparando la centralidad de PageRank con otros tres métodos, y se analizó el efecto de rendimiento de cada algoritmo.</li> </ol>	Este resultado experimental, permitió determinar que el modelo PageRank es más preciso, eficiente, tiene una alta escalabilidad y presenta mayor rendimiento en comparación con las demás medidas de centralidad del SCNA.
Fire, M., Puzis, R., and Elovici, Y	2013	SCNA	<ol style="list-style-type: none"> <li>1. Se ejecutó la recolección y clasificación de los datos manualmente.</li> <li>2. Se verificó los resultados de las medidas de centralidad y los de los algoritmos de aprendizaje automático a medida que avanzaban en la detección de roles influyentes ocultos dentro de la empresa.</li> </ol>	Se corroboraron que los métodos del SCNA arrojaron resultados con un índice de confiabilidad mayor al 80%.



Minería de Datos (DM)				
Autor	Año	Método	Proceso de validación	Resultados
Kamali, H.; Rahmani, H. and Shah-Hosseini, H.	2020	Algoritmo genético	1. Se determinó el peso de cada criterio de forma completamente aleatoria. Para ello, se ponderaron los criterios cien veces arbitrariamente.	La precisión del algoritmo genético fue del 60% respectivamente. La comparación de los resultados del algoritmo genético con el método de ponderación aleatoria indica que los algoritmos genéticos son 22% más precisos que la ponderación aleatoria, según los datos de la red de interacción de la empresa Enron.
Crispim, J. and Pinho de Sousab, J.	2009	Búsqueda tabú	1. Se calculó la correlación entre los criterios evaluados con el objetivo de verificar si los criterios elegidos deben cambiarse o arrojan resultados precisos. 2. Se calculó la precisión promedio cien veces utilizando una ponderación aleatoria.	Según los datos, al realizar la validación del algoritmo de búsqueda tabú, se obtiene un nivel de confianza en los resultados del 90% $\pm$ 5. Lo cual indica, que los criterios seleccionados en la realización de la investigación no presentan interdependencias significativas, lo que significa que los resultados son precisos.
Nurek, M. and Michalski, R.	2020	Árboles de decisión, ANN Y SVM	1. Se entrenó un modelo con el uso del algoritmo de búsqueda de cuadrícula con validación cruzada de 5 veces, por lo que se probaron todas las combinaciones posibles del rango de valores dados, y la mejor combinación con respecto a la puntuación más alta, devolvió el valor del promedio macro. 2. Se verificó como la actividad de una persona puede influir en el resultado de la clasificación y, por lo tanto, se examinó si los meses en los cuales hubo un mayor intercambio de mensajes entre los empleados arrojan mejores resultados.	La clasificación de los miembros de la red social corporativa realizada por los métodos de ANN Y SVM tienen una precisión del 100%, mientras que los resultados obtenidos mediante el método de árboles de decisión cuentan con un índice de confiabilidad del 80%.



<b>Minería de Datos (DM)</b>				
<b>Autor</b>	<b>Año</b>	<b>Método</b>	<b>Proceso de validación</b>	<b>Resultados</b>
Sun, Y.; Yao, P.; Wan, L.; Shen, J. and Zhong, Y.	2017	Recocido simulado	<ol style="list-style-type: none"> <li>1. Se probó los métodos en diferentes redes reales y redes artificiales.</li> <li>2. Se realizó una comparación del método seleccionado con otros dos métodos recientemente desarrollados que son: el método de clasificación basado en la entropía de mapeo (MME) y el método de clasificación basado en atributos múltiples (MMA).</li> </ol>	Se obtuvo que la secuencia de nodos importantes que arrojó el método de recocido simulado es más precisa, eficaz y efectiva que los demás métodos.
Albino, V. and Garavelli, A.	1998	ANN	<ol style="list-style-type: none"> <li>1. Se realizó la prueba NN, la cual consiste en la verificación del rendimiento de la red.</li> <li>2. Se validó que la implementación de la red neuronal suele requerir otras fases auxiliares, como la elección del software, la integración de NN con otros sistemas de información y el mantenimiento de NN. Además, se comprobó la precisión del enfoque NN mediante una comparación con otros enfoques, como el AHP y la lógica difusa.</li> </ol>	Entre los principales resultados de NN, la generalización, es decir, la capacidad de proporcionar una respuesta adecuada en correspondencia de nueva información, y la tolerancia al ruido, es decir, la capacidad de reconocer información poco clara, se analizaron críticamente en esta fase y los resultados fueron verídicos y precisos.
Yelupula, K. and Ramaswamy, S.	2008	K-means	<ol style="list-style-type: none"> <li>1. Se validó los datos con la fuente que encontraron en Internet.</li> </ol>	Aunque predijeron la estructura organizativa de Enron con el atributo numérico, no pudieron obtener una precisión del 100%.
Silva, R., Menezes, V., Oliveira, J., de Souza, M. F., de Mello, C. E. R., Souza, J. M., and Zimbrão, G.	2009	DM	<ol style="list-style-type: none"> <li>1. Se analizaron las respuestas y las compararon con los resultados del enfoque, realizando entrevistas y cuestionarios a los profesores de la universidad.</li> <li>2. Se validó las relaciones fuertes y débiles, así como los puntos críticos en la red social corporativa.</li> </ol>	La mayoría de los profesores que respondieron el cuestionario dicen tener más publicaciones con profesores de la misma institución que con profesores externos. Cada profesor indicó los nombres del investigador con el que más se relaciona.



Minería de Datos (DM)				
Autor	Año	Método	Proceso de validación	Resultados
Li, N.; Kong, H.; Ma, Y.; Gong, G. And Huai1, W.	2016	K-NN	<p>1. Se comparó el algoritmo con otros siete algoritmos de clasificación KNN (Euclidiana, Euclidiana cuadrada, City Block, Canberra, Squared-chord, Squared-chi-square y MKNN).</p> <p>2. Se formularon tres ecuaciones de distancia del algoritmo KNN propuesto, y se probaron, respectivamente, las cuales son: método KNN aplicado con: la Entropía-distancia Euclidiana (EEKNN), Entropía-distancia Manhattan (EMKNN) y Entropía-distancia Canberra (ECKNN).</p>	<p>Los resultados muestran que EEKNN, EMKNN y ECKNN tienen la precisión de clasificación promedio más alta, correspondiendo a un 87.59%, 87.64% y 85.51%, respectivamente, con el criterio de distancia ponderada.</p>

Fuente: las autoras



### 5.3 Modelos basados en el análisis de redes sociales corporativas

En primer lugar, los autores de Toni & Nonino (2010) a través de las medidas locales y globales, identificaron a los miembros clave integrados en la estructura organizacional informal (redes informales) y encontraron la contribución o importancia de estos miembros clave en la red corporativa con respecto al desempeño. Por otro lado, las medidas de Centralidad de Grado, Centralidad de inmediatez (Intermediación), Centralidad de cercanía y centralidad del vector propio fueron analizadas en un caso de estudio de una empresa italiana de consultoría, tecnología e información, que opera por módulos o equipos. Este estudio se enfocó en investigar la red corporativa de un solo módulo, conformado por el gerente, dos consultores y dos analistas, donde se quiere determinar los miembros claves que contribuyen a la mejora y funcionamiento del equipo con el fin de diseñar un sistema integrado. Los resultados muestran que se deben incorporar dos analistas adicionales, debido a que el gerente es el miembro más central y el único encargado de intercambiar información, por lo que los demás miembros no tienen mayor importancia por su baja contribución al proyecto. La incorporación de estos nuevos integrantes permitirá encontrar un flujo directo de información entre todos los integrantes, haciendo que todos aporten y sean influyentes para el equipo, puesto que la centralidad del gerente disminuye (Grassi et al., 2010).

Así mismo, los autores Zbieg et al. (2016), a partir de las medidas de centralidad, realizaron un estudio que se enfocó en identificar empleados con un rol de agentes de cambio. Los datos incluyen redes de colaboración de tres empresas medianas. Se seleccionó agentes en función de cuatro medidas de centralidad de la red, jerarquía y antigüedad; los agentes fueron identificados al azar. Los agentes seleccionados con cada método se evaluaron contra la cobertura de la red, que es su área de impacto y realimentación. En primer lugar, analizaron si los agentes seleccionados con métodos basados en medidas de centralidad de la red (grado, intermediación, cercanía, vector propio) y medidas clásicas (jerarquía, antigüedad) optimizan mejor la cobertura del área de comunicación que los agentes seleccionados aleatoriamente. Los resultados arrojaron que los empleados con una antigüedad prolongada son más centrales dentro de una red, ya que han tenido más tiempo para establecer relaciones. A su vez, los empleados ubicados en una jerarquía alta coordinan el trabajo de otros empleados y deben tener contactos más amplios que especialistas que se ocupan de tareas particulares.

Otra aplicación enfocada en el análisis de redes sociales corporativas, la realizan los autores Yuan Sun et al. (2015) mediante la centralidad de PageRank del análisis de redes sociales, identificaron los nodos claves de la red logística de una empresa en Shanghai, China. En este estudio no solo tuvieron en cuenta la fuerza de la relación entre los nodos, sino que también analizaron el peso del nodo. En este documento, se define el grado de actividad del nodo basado en las características del comportamiento de los miembros de la red. Como resultado, las redes logísticas se pueden considerar como redes de doble ponderación tomando la fuerza de la relación como el peso del borde y la actividad del nodo como el peso del nodo. Los



pesos de los nodos se utilizaron como coeficiente de amortiguación y el peso de los bordes se utilizó para calcular la importancia de los nodos durante el proceso iterativo. Por último, seleccionaron un conjunto de datos de una empresa de logística y se llevaron a cabo experimentos integrales. Los resultados experimentales muestran que el algoritmo propuesto puede descubrir de forma eficaz y eficiente nodos clave en una red logística real.

Uno de los ejemplos aplicativos de los métodos del análisis de las redes sociales, es una investigación realizada por León et al. (2017), a partir de las medidas densidad de red, grado, centralidad de intermediación, cercanía, la centralidad del vector propio y la centralidad de Katz, la cual tuvo como objetivo identificar a los individuos que influyen en los procesos de intercambio de conocimiento desde una red social interna y pronosticar los futuros flujos de conocimiento que puedan atravesarla. Esto se aplica a una red corporativa de una compañía de seguros británica. Los principales resultados enfatizaron los miembros más importantes, sus relaciones, los flujos de conocimiento futuros y la conexión entre la heterogeneidad y estructura de la red y la intención de compartir conocimiento futuro de los empleados. Establecieron quién actuará como difusor de conocimiento, compartiendo lo que sabe con sus compañeros de trabajo, y quién actuará como repositorio de conocimiento, enfocándose en adquirir cada vez más conocimiento.

Por otra parte, los autores Hoppe & Reinelt (2010) realizaron una investigación por medio del modelo del análisis de redes sociales, la cual se focalizó en una organización sin ánimo de lucro denominada Sierra Health. Dicha empresa está conformada por líderes ejecutivos, por lo tanto, es considerada como una red de liderazgo organizacional. El objetivo de este estudio consistió en identificar miembros clave, es decir, se busca determinar dentro de la red de liderazgo “Líderes emergentes” que son los nodos más importantes. La investigación arrojó como resultado que uno de los miembros más influyentes y con mayor nivel en la actividad de asesoramiento es un empleado reciente. Los demás nodos informaron que con frecuencia este empleado estaba ocupado, sobrecargado de trabajo, inaccesible y que relativamente tenían poca interacción con él. Esta situación provocó que toda la responsabilidad del desempeño de las actividades recayera solamente sobre un empleado. El estudio permitió conocer que en la red social corporativa no hay un trabajo equitativo ni colaborativo, no todos los miembros tienen el mismo grado de participación, lo cual a su vez causa que los trabajadores que más se destacan en sus funciones estén aislados y sobrecargados de trabajo, y que los demás no se esmeren en sus labores y aporten en menor medida. Los datos destacaron un área específica para mejorar en la red de liderazgo existente y subrayaron la necesidad del programa de Líderes Emergentes. Se reposicionó al líder sobrecargado y se agregaron recursos de liderazgo adicionales para ayudar a los empleados a dejar de depender de él. Cuando se entrevistó al director ejecutivo de Commonwealth después de la evaluación y reestructuración, informó que, a pesar de que la evaluación procedió como él esperaba, los beneficios más importantes para él no fueron anticipados y le



dieron una nueva perspectiva sobre cómo encajaban las piezas de Commonwealth. Reorientó su enfoque de desarrollo de liderazgo de los individuos al sistema de liderazgo colectivo.

En estas mismas circunstancias, la empresa constructora Hugo René Luna, realizó un estudio con base en el análisis de las redes sociales con el objetivo de identificar la importancia de los miembros de la empresa constructora H.R.L. Para este trabajo específicamente, las medidas que son examinadas son las de centralidad, un indicador de cómo un nodo está conectado con otros nodos. Con la información obtenida del análisis de redes sociales organizacionales, se realizó una encuesta que se distribuyó a todos los empleados de la constructora Hugo René Luna en la ciudad de Guayaquil. La encuesta consistió en un conjunto de 5 relaciones que han demostrado ser útiles en la revelación de las características de la red social informal, es decir, cómo se realiza la colaboración en la empresa. Se obtuvo respuesta de las 93 personas que conforman la organización, con lo cual se recolectó la información que permitió analizar la red social informal de toda la empresa. Con los resultados arrojados de las encuestas se procedió a analizar la información, mediante la herramienta de análisis de redes sociales (UciNet), y así detectar los actores y puntos críticos de la empresa. En la caracterización de los roles claves, se pudo observar que a pesar de que las redes analizadas son densas, es decir, hay muchas relaciones entre los miembros, existen puntos críticos que han sido identificados en el análisis. La alta dependencia, la vulnerabilidad de las relaciones de salida en la red social corporativa y la presencia de personas aisladas son algunos problemas que se han identificado, lo que determina las intervenciones para mejorar la eficiencia del desarrollo del trabajo diario (Luna, 2015).

De esta forma, las medidas locales y globales permiten determinar los actores más importantes que forman parte de una red social. Estos actores están representados por nodos, que se encuentran interconectados y comparten algún tipo de información, lo que permite analizar por medio de diversos algoritmos los nodos más relevantes, que tienen una mayor participación y colaboración en una comunidad (Can & Alatas, 2019). Así, estas medidas de centralidad y prestigio han sido durante mucho tiempo el método tradicional en las redes sociales para clasificar nodos en diferentes campos de acción como corporaciones, investigación científica y proyectos a gran escala.

Lo anterior, se puede evidenciar en una investigación donde analizaron a seis reconocidas empresas de alta tecnología de diversos tamaños, desde pequeñas empresas con varios cientos de empleados hasta empresas a gran escala con cientos de miles de empleados, utilizando diferentes medidas de centralidad para detectar los roles influyentes ocultos dentro de cada organización. Además, se hizo uso de algoritmos de aprendizaje automático para clasificar los roles de gestión en la red corporativa. A través de rastreadores web, recopilaban una cantidad considerable de datos de fuentes públicas sobre las organizaciones estudiadas e información que sus empleados han expuesto en Facebook, LinkedIn y otras fuentes



disponibles públicamente. Mediante una serie de pasos, se pudo llevar a cabo esta investigación. Primero, adquirieron la topología de la red social informal de la organización a partir de información disponible públicamente, seguidamente recopilaron la información sobre la estructura de la empresa expuesta por los empleados de la empresa en Facebook y por último, se empleó un algoritmo de última generación para agrupar la red social de la organización en comunidades inconexas, lo cual permitió comparar los líderes y comunidades revelados con la información obtenida de LinkedIn; esto permitió derivar los roles de muchas comunidades dentro de una organización, proporcionando información importante sobre la estructura de la organización y los patrones de comunicación (Fire et al., 2013).

Por último, el autor HATALA (2006) introdujo el análisis de redes sociales como una metodología única para estudiar las relaciones sociales y su importancia en el campo del desarrollo de recursos humanos de una organización. El SCNA contribuye significativamente al campo, al medir las relaciones que existen entre los individuos y el impacto que esas relaciones tendrán en la producción de capital humano. Esta investigación se hizo mediante el estudio de un practicante del área de Recursos Humano para incrementar la efectividad organizacional mediante el uso de métodos de aprendizaje, mejora del desempeño y desarrollo de liderazgo. El comportamiento individual es un reflejo del entorno y las respuestas conductuales específicas no se pueden pronosticar con precisión sin el conocimiento del contexto en el que funciona el individuo o el grupo. Por lo tanto, es importante comprender las relaciones interpersonales que ocurren en una organización y el impacto que los factores contextuales tienen en la respuesta del individuo al entorno laboral. Además, el SCNA ayudó a desarrollar aún más el campo de los recursos humanos al permitir al investigador analizar la interacción entre las personas y su entorno. El análisis de redes sociales puede agregar rigor empírico a áreas tan diversas como el cambio organizacional, el diseño instruccional y la entrega de capacitación. La utilidad práctica del SCNA puede ayudar a los profesionales de Recursos Humanos a medir la eficacia de la intervención y su impacto a lo largo del tiempo.

#### **5.4 Métodos basados en la minería de datos**

Con respecto a los métodos de la minería de datos, los autores Chan & Chung (2004) desarrollaron mediante un algoritmo genético un procedimiento de optimización genética multicriterio, diseñado específicamente para resolver problemas de optimización en la gestión de la cadena de suministro. El algoritmo propuesto se analiza con un problema de distribución de pedidos en una red de cadena de suministro impulsada por la demanda. Lo



cual está directamente relacionado con el desempeño e importancia de los integrantes que forman parte del reparto de pedidos.

Del mismo modo, Kamali et al. (2020) en su trabajo propusieron un sistema, utilizando Algoritmos Genéticos (DINGA), el objetivo de este estudio es descubrir los nodos importantes en la red organizacional de la empresa Enron, a través del conjunto de datos de correos electrónicos. El correo electrónico representa las interacciones entre los empleados de la empresa. En este conjunto de datos, los empleados se consideran nodos y se envían correos electrónicos entre ellos como bordes. Para implementar el método propuesto en la fase de entrenamiento del algoritmo genético, se consideraron que 33 miembros de la empresa son los nodos importantes con cargos de presidente, vicepresidente, director de operaciones, director general y ejecutivo relativo del gobierno. En este sistema propuesto, los nodos importantes de las redes sociales se descubrieron empleando una combinación de ocho criterios informativos y su ponderación inteligente. Como resultado se obtuvo que las centralidades de la conectividad de promedio local “LAC” (describe qué tan cerca están los vecinos de un nodo) y de la red son más importantes que los otros criterios. Además, las centralidades de vector propio y proximidad son menos importantes que los otros criterios para descubrir los nodos importantes en la red de Enron.

Por otro lado, el estudio denominado “Selección de socios en empresas virtuales: un enfoque de apoyo a la decisión de criterios múltiples” es un ejemplo aplicativo del método de búsqueda tabú, el cual consiste en una investigación que busca determinar los miembros más importantes de una red compuesta de múltiples empresas donde se pueden realizar 12 actividades diferentes que requieren 10 recursos diferentes, y está formada por 100 candidatos (empresas) caracterizados por 20 criterios (criterios de 12 nodos y ocho criterios de borde). En este caso cada nodo es una organización y se requiere seleccionar a las empresas que cuenten con un mayor desempeño, liderazgo y competencias gerenciales para construir una red de socios, es decir, se busca acotar los miembros de la red inicial con el objetivo de obtener una red final conformada por nodos claves. Los componentes principales de este estudio son: la función objetivo, la solución inicial, la estructura de vecindad y la lista tabú. Los resultados computacionales preliminares demuestran claramente el potencial del enfoque para la aplicación práctica (Crispim & Pinho de Sousa, 2009).

Los autores Yu Sun et al. (2017) en su investigación estudiaron el problema de clasificación de nodos de una red de aerolíneas estadounidenses (NAA) y propusieron el método de clasificación del recocido simulado. El método propuesto describe el concepto de secuencia de importancia de los nodos y define una medida de razonabilidad para ello, comparando una secuencia importante y su medida de razonabilidad con un estado de red compleja y la energía del estado, respectivamente. Este método construye la secuencia importante más razonable con la ayuda de recocido simulado. Los resultados del experimento muestran que este método



de clasificación no sólo es eficaz, sino que también se puede aplicar a diferentes tipos de redes complejas.

En otro contexto, los métodos de la minería de datos fueron aplicados en las empresas constructoras, para calificar subcontratistas. Debido a su intrínseca ambigüedad y difícil formalización, está es una tarea particularmente compleja, generalmente realizada por la experiencia y la intuición de la gerencia. En el artículo desarrollado por Albino & Garavelli (1998), se propone una red neuronal para apoyar la gestión en la calificación de subcontratistas, es decir, se requiere identificar a los subcontratistas más importantes, preparados y capacitados para efectuar las labores. Para mostrar la implementación práctica de una red neuronal a la calificación de subcontratistas, se analiza un caso de aplicación, con el apoyo del director de compras de un contratista general que opera principalmente en una compañía que realiza grandes proyectos de obra civil. Para evaluar las empresas especializadas que se postulan como potenciales subcontratistas (competidores) y determinar cuáles de ellos son los nodos claves de la red corporativa que vincula a todos los subcontratistas, se consideraron los siguientes parámetros: reducción de precio, reducción de tiempo, características técnicas / cualitativas de la oferta, confiabilidad contractual del competidor, capacidad de gestión del competidor.

Con el mismo propósito Olguin et al. (2009) realizó un estudio para medir automáticamente el comportamiento organizacional, en este estudio se utilizaron diversos algoritmos, entre ellos se encuentra el SVM. El objetivo fue comprender cómo los patrones de comportamiento dan forma a las personas y las organizaciones. Mediante el uso de sensores corporales en grandes grupos de personas durante largos períodos de tiempo en entornos naturalistas, han podido identificar, medir y cuantificar interacciones sociales, comportamiento grupal y dinámicas organizacionales. A su vez, los autores (N. Li et al., 2016), realizaron un estudio para predecir el desempeño e importancia humana en una organización, que puede ser significativo para garantizar la competitividad de la empresa. Este algoritmo mostró una alta precisión de predicción cuando la distancia ponderada entre el individuo predicado y las muestras es pequeña. Cuando se implemente más este método para la fabricación, con un gran tamaño de grupo de muestra y diversidad de atributos humanos mejorará la efectividad, la capacidad de predicción y la aplicabilidad.

Mediante el uso de algoritmos y metodologías de minería de datos, se analizó el flujo de correos electrónicos dentro de una organización. Con el fin de determinar si se puede predecir la red corporativa de Enron, para llevar a cabo este estudio se utilizó el algoritmo k-means para clasificar a los empleados en diferentes escalones organizacionales, identificando y evaluando diferentes características. Los resultados muestran que se puede obtener información significativa sobre la estructura de una organización incluso si se descuida el cuerpo (contenido) de los correos electrónicos. Se extraen suficientes datos relevantes sobre



la red social de correos electrónicos, utilizando un análisis simple del flujo de estos, y se obtiene una imagen general de la estructura organizativa y las relaciones que se tienen entre los empleados (Yelupula & Ramaswamy, 2008).

Por consiguiente, en el ámbito corporativo la formación de una jerarquía dentro de una organización es una forma natural de asignar las funciones, delegar responsabilidades y optimizar el flujo de información. Por otro lado, los empleados suelen enviar docenas de correos electrónicos cada día y, al hacerlo, forman naturalmente una red social informal donde los nodos son los individuos y los bordes son las acciones que los vinculan. Los autores Nurek & Michalski (2020) realizaron un estudio en una empresa de fabricación ubicada en Polonia donde el conjunto de datos analizados contiene un intercambio de mensajes de nueve meses entre empleados administrativos y consta de dos archivos: el primero contiene la jerarquía de la empresa y el segundo almacena el historial de comunicaciones. El concepto clave de este trabajo es obtener una clasificación aproximada de los miembros de la red social organizacional. Para esto se emplearon métodos de aprendizaje automático como los Árboles de Decisión, Redes Neuronales y Máquinas de Vector de Soporte (SVM), así como un algoritmo de clasificación colectiva, que también se propone en este trabajo. El enfoque utilizado permitió comparar el rendimiento de los métodos tradicionales de clasificación del análisis de redes sociales con los métodos propuestos en el estudio. Los resultados demuestran que la red social construida con los métodos propuestos expone una estructura organizacional más acertada que la del análisis de redes sociales.

En otra instancia es indispensable mencionar que, por medio de la minería de datos, la utilidad y sus técnicas se estudia la gestión empresarial, así como la importancia de la gestión de información. Un punto significativo para la posible mejora en la gestión de los procesos de una organización lo constituye la auditoría mediante la cual se valida la información y sus procesos de negocio (Jaramillo et al., 2018). Se realiza un análisis desde las técnicas de minería de datos de un log de eventos estandarizando su formato y considerando el uso de datos históricos y actuales como elemento para la minería de datos registrando en forma de trazas las acciones que se van realizando cuando se ejecutan instancias o casos del proceso de negocio y estas trazas contienen información de los casos, lo que posibilita que se usen las técnicas de minería de datos en las auditorías. Con la información que soportan los procesos de negocio de la empresa el auditor puede verificar si el proceso ejecutado está o no en correspondencia con lo establecido, pudiendo así detectar y esclarecer las violaciones en la ejecución de las tareas. Para la aplicación de las técnicas de minería de datos se extrajo una muestra de la trazabilidad a partir de la ejecución del proceso Gestionar (Herrera, 2013).

Este estudio arrojó el número de actividades correspondientes al proceso de gestión de los roles, aunque contenía sólo 31 casos, 804 eventos, 52 clases de eventos y 3 tipos de eventos, permitió determinar algunas anomalías que se presentaban en dicho proceso. Para poder



realizar un análisis más detallado se empleó la técnica denominada Advanced Dotted Chart Analysis, la cual ha sido implementada en ProM, que es una herramienta genérica, de código abierto que agrupa las principales técnicas desarrolladas en la minería de datos e instrumento especializado para procesos de auditoría. Con este tipo de técnica se hace un diagnóstico visual del proceso analizado. Como resultado de este análisis se determinó que la tarea que fallaba con mayor frecuencia era la de Condicionar Operación, ejecutándose en 31 ocasiones y falló el 38,7% de las veces. En total se produjeron en el proceso analizado, 60 eventos de tipo fallo, por lo cual los fallos asociados a la tarea Condicionar Operación (consiste en la verificación de las condiciones necesarias para que se seleccione entre agregar, editar, eliminar o terminar) representan el 20 % de los fallos ocurridos. Este es un valor significativo considerando que el proceso contenía 21 tareas, determinando que uno de los administradores del sistema no dominaba totalmente los datos que se debían especificar antes de optar por algunas de las tareas agregar, editar, eliminar o cerraba la ventana del navegador dejando incompleto el proceso en curso. Cabe destacar que la minería de datos se ha convertido en una herramienta útil para el proceso de auditoría de la empresa y para la mejora de la gestión empresarial de manera general. Estas técnicas permiten detectar fraudes, malas prácticas, ineficiencias y riesgos en el proceso ejecutado. También es posible aplicarlas como soporte a la operatividad de la empresa (Herrera, 2013).

Por otro lado, en la tabla 7, se encuentran todos los aspectos que se tuvieron en cuenta en cada uno de los artículos a la hora de incluirlos en este estudio. Primeramente, se revisa los datos básicos con los que cuenta toda investigación que son: autores, fecha de publicación, el título y el respectivo resumen. Luego, se detalla el tipo de modelo y submodelo que aplica el estudio para identificar el grado de complejidad de este y lograr una clasificación apropiada de las investigaciones encontradas. Por consiguiente, se verifica si los artículos tienen enfoque corporativo y cuál es el objetivo al que los autores quieren llegar con su trabajo. Así mismo, se le hace una revisión a la estructura del modelo utilizado y se menciona el tipo de organización en la cual se desarrolla el estudio.

Así mismo, en la tabla 7 se definen los aspectos más comunes que se encontraron en las 20 investigaciones, tales como: variables clave, métodos y procedimientos implementados por los autores para recolectar los datos, modelo matemático construido, software utilizado para el análisis de los resultados, funcionamiento de la red social corporativa construida, validación de los modelos aplicados para determinar el grado de confiabilidad y precisión de los datos, resultados generados, aporte de la investigación para mejorar la comunicación, participación e información entre los miembros de la empresa y las limitaciones y futuros estudios.

Por último, cabe aclarar que los artículos no necesariamente cumplen con todos los criterios, sino que como se mencionó anteriormente, los criterios definidos corresponden a los más



---

comunes que se encontraron al momento de inspeccionar y revisar de forma detallada cada uno de ellos. Por otra parte, la tabla 7, solo relaciona a los tres primeros artículos del SCNA (de Toni & Nonino, 2010; Fire et al., 2013; Grassi et al., 2010; Zbieg et al., 2016) para que el lector pueda identificar los aspectos seleccionados y evaluados, que contribuyeron a cumplir los objetivos de este trabajo. Por cuestión de espacio, la tabla completa, que involucra a las 20 investigaciones, se puede observar en el anexo de Excel “Criterios para determinar relacionamiento efectivo” (Solano & Murillo, 2021), el cual está adjunto a este documento.



Tabla 7. Aspectos y criterios relevantes para determinar la importancia de un miembro en una red social corporativa en inglés.

ARTÍCULO	1	2	3
AUTORES	Toni, A. and Nonino, F.	Grassi, R., Stefani, S., and Torriero, A.	Zbieg, A., Batorski, D., and Żak, B.
AÑO DE PUBLICACIÓN	2010	2010	2016
TÍTULO	Roles clave en la organización informal: una perspectiva de análisis de redes	Centralidad en las redes organizacionales	¿Cómo seleccionar agentes de cambio en las organizaciones? Una comparación de los enfoques clásico y de red
MÉTODO MODELADO	Análisis de Redes Sociales	Análisis de Redes Sociales	Análisis de Redes Sociales
CLASE DE MODELO	Combinada	Medidas Locales	Medidas Locales
ENFOQUE ORGANIZACIONAL	SI	SI	SI
OBJETIVO	Identificar un nuevo rol informal clave	Configurar de forma óptima un equipo de proyecto. Considerando el caso de estudio de una empresa italiana de consultoría ITC, llamémosla ABC. ABC es un actor importante en consultoría tecnológica e informativa. Se debe diseñar un equipo de proyecto para implementar un sistema ERP (planificación de recursos empresariales).	Este estudio utiliza tres redes de colaboración organizacional y tiene como objetivo explorar y verificar este enfoque centrándose en las medidas de centralidad, jerarquía y antigüedad, y seleccionar agentes al azar.
BASE DE LA ESTRUCTURA DEL MODELO	Construcción del modelo de grupo	Literatura	Construcción del modelo de grupo
NOMBRE DE LA ORGANIZACIÓN	Grupo Euris	Empresa italiana de consultoría ITC	No especificada
OBJETO SOCIAL	Mercado de tecnología de la información	Consultoría	No especificado



ARTÍCULO	1	2	3
<b>VARIABLES CLAVE</b>	La investigación se centra en el análisis en profundidad de la estructura informal de un estudio de caso de una empresa basada en el conocimiento que opera en la industria de sistemas de información aplicando la metodología del SCN. Partiendo del modelo de cinco redes informales propuesto por Cross. Estructurar un marco general de análisis permitirá tanto un análisis de la estructura organizativa informal (redes informales) como la identificación y caracterización de los roles informales clave dentro de las empresas y entre las empresas del Grupo Euris, a saber, líderes de opinión, conectores centrales, cuellos de botella, expertos, consultores o personas útiles.	En este artículo, intentan dar un paso más en la comprensión del papel de los individuos en las redes organizacionales a través de un enfoque topológico gráfico basado en medidas de centralidad.	El estudio contiene varios pasos de análisis de datos. En primer lugar, se crean tres redes compuestas por lazos de colaboración para representar las estructuras de las empresas estudiadas. A continuación, se calculan las centralidades básicas de la red para todos los nodos de la red. Estas medidas de red asignadas a cada empleado, junto con la información sobre la jerarquía y antigüedad de los empleados, sirvieron para los cálculos de las clasificaciones de los agentes de cambio.
<b>MÉTODO DE RECOLECCIÓN DE DATOS</b>	Entrevistas	Métodos de observación	Encuestas
<b>ECUACIONES PROPORCIONADAS</b>	SI	SI	NO
<b>En caso afirmativo, ¿se proporcionó una justificación para la (s) ecuación (es) / forma (s) funcional (es) utilizada?</b>	Sí, para algunos parámetros	Sí, para todos los parámetros	Sí, para algunos parámetros
<b>SOFTWARE</b>	SI	NO	SI



ARTÍCULO	1	2	3
<b>En caso afirmativo, defina el software utilizado</b>	Excel, UCINET 6, NetDraw		ONA
<b>En caso afirmativo, describa la contribución del software a la investigación</b>	En la recopilación de datos, utilizan uno de los softwares más comunes para el análisis de redes: UCINET 6w. Además, usan NetDraww para visualizar las redes.		Simplifica y estandariza el proceso de recopilación de datos y proporciona visualización y análisis de la red de administradores y empleados al instante después de completar las encuestas.
<b>CONSTRUCCIÓN DE LA RED</b>	SI	SI	NO
<b>En caso afirmativo, indique el número total de redes construidas</b>	6	4	
<b>En caso afirmativo, describa las redes construidas</b>	La red de información muestra el análisis de los flujos de información con referencia a la red de información del Grupo Euros, los datos sobre la centralización de la red y el grado de centralidad medio. La red de conocimiento muestra cómo se distribuye y utiliza el conocimiento, con el fin de identificar a los actores más competentes y experimentados según sus colegas, es decir, los expertos. El análisis identifica a las personas potencialmente capacitadas para ocupar puestos importantes en la empresa y los denominados especialistas periféricos.	Se visualiza la complejidad de los diferentes escenarios. Además, el papel central del director general ha cambiado y su papel central ha ido disminuyendo progresivamente. Esto permitió introducir un nuevo puesto de director general dividiendo las responsabilidades entre dos directores de proyecto.	
<b>¿SE REALIZÓ VALIDACIÓN?</b>	NO	NO	NO



ARTÍCULO	1	2	3
<b>En caso afirmativo, seleccione el proceso utilizado para validar el comportamiento del modelo</b>			
<b>RESULTADOS GENERADOS</b>	<p>Uno de los principales objetivos de la investigación fue identificar un nuevo rol informal clave, al que llamamos pilus prior, que sintetiza la resolución de problemas, la experiencia y las características de accesibilidad. El objetivo se ha logrado utilizando un enfoque original, el análisis conjunto de tres redes informales. El caso sugiere que los pilus priors son los líderes emergentes en la empresa y superan a sus colegas asumiendo el rol protagónico de “gerente informal” reconocido por su grupo, coordinando informalmente las actividades y transfiriendo su experiencia a colegas no calificados. Además, el valor de pilus prior podría utilizarse como indicador para una evaluación “de abajo hacia arriba” de los empleados.</p>	<p>Los resultados muestran que se deben incorporar dos analistas adicionales, ya que el gerente es el miembro más central y el único encargado de intercambiar información, por lo que los demás miembros no son de mayor importancia por su baja contribución al proyecto.</p>	<p>Los resultados muestran que los empleados a largo plazo son más centrales dentro de una red, ya que han tenido más tiempo para construir relaciones. A su vez, los empleados ubicados en una jerarquía superior coordinan el trabajo de otros empleados y deben tener contactos más amplios que los especialistas que se ocupan de tareas particulares.</p>
<b>REESTRUCTURACIÓN DEL SISTEMA ORGANIZATIVO</b>	SI	SI	SI



ARTÍCULO	1	2	3
<p><b>En caso afirmativo, indique la propuesta de reestructuración</b></p>	<p>En la investigación, se centran en un análisis en profundidad de la estructura informal de un grupo de empresas utilizando la metodología y el marco del SNA. El Grupo Euris ha sido seleccionado por su pertenencia a la industria intensiva en conocimiento; de hecho, en estos contextos, las estructuras y las relaciones informales juegan un papel fundamental en el logro de objetivos estratégicos y operativos. El documento identifica y caracteriza los roles informales clave (a saber, líderes de opinión, conectores centrales, cuellos de botella, expertos, consultores o personas serviciales) y uno nuevo, llamado pilus prior (primer lanzador) que sintetiza la resolución de problemas, la experiencia y las características de accesibilidad. El caso sugiere que los pilus priors son los líderes emergentes informales en la empresa que superan a sus colegas.</p>	<p>Incorporaron nuevos integrantes, que permitirán encontrar un flujo directo de información entre todos los integrantes, haciendo que todos contribuyan y sean influyentes para el equipo, ya que la centralidad del gerente disminuye, y ABC lo soluciona introduciendo nuevos actores y reordenando tareas y responsabilidades.</p>	<p>A partir de las medidas de centralidad, realizaron un estudio que busca identificar a los empleados con el rol de agentes de cambio. Los datos incluyen redes de colaboración de tres empresas medianas. Seleccionaron agentes basándose en cuatro medidas de centralidad, jerarquía y antigüedad de la red, e identificaron agentes al azar. Los agentes seleccionados con cada método fueron evaluados contra la cobertura de la red, que es su área de impacto y retroalimentación. En primer lugar, analizaron si los agentes seleccionados con métodos basados en medidas de centralidad de la red (grado, intermediación, cercanía, vector propio) y medidas clásicas (jerarquía, antigüedad) optimizan mejor la cobertura del área de comunicación que los agentes seleccionados aleatoriamente.</p>



ARTÍCULO	1	2	3
<p><b>LIMITACIONES DEL ESTUDIO Y RECOMENDACIONES PARA FUTURAS INVESTIGACIONES</b></p>	<p>La investigación futura se moverá por dos caminos: el primero se refiere a la ampliación de la muestra que permitirá una mejor generalización de los resultados, mientras que el segundo se refiere al desarrollo de un modelo para evaluar cuantitativamente el desempeño de las personas reconocidas como líderes informales.</p>	<p>No especificado</p>	<p>La limitación de la investigación radica en el hecho de que el método fue verificado para un número reducido de redes organizacionales. Además, todas las redes estudiadas eran representaciones de empresas medianas; todavía no sabemos qué tan alta cobertura de red se puede obtener para redes organizativas más grandes. Esto abre el primer campo para futuras investigaciones. En segundo lugar, el documento muestra la importancia de la cobertura de la red en las iniciativas de gestión del cambio basadas en supuestos específicos. El cambio se estudia como un proceso de un solo paso. Este sencillo modelo de difusión se puede combinar con el proceso de dos o más pasos. Además, la cobertura de la red se calcula basándose en vínculos dirigidos a agentes seleccionados (en grado). Si bien los autores están convencidos teórica y prácticamente de este enfoque, se deberían explorar otros tipos de cálculos de cobertura en investigaciones futuras. Finalmente, incluso si se trabaja con datos reales, la investigación actual probó los métodos de selección de agentes de cambio y obtuvo una cobertura de red bastante teórica. En futuras investigaciones, sería sumamente valioso verificar los conceptos aplicándolos al proceso de cambio organizacional real.</p>
<p><b>REFERENCIAS BIBLIOGRÁFICAS</b></p>	<p>SI</p>	<p>SI</p>	<p>SI</p>

Fuente: las autoras



En la tabla 7 se puede visualizar todos los aspectos y criterios relevantes usados en el campo de la investigación teórica para determinar el relacionamiento efectivo y predecir la importancia de un miembro en una red social corporativa. Según los 20 artículos que conforman este estudio, la posición que tiene el nodo en la red social corporativa en la mayoría de los casos indica el alcance que tiene el trabajador en la organización, es decir, los que pertenecen al personal de alta dirección, tienden a ser empleados indispensables en la red por su gran capacidad de conocimientos, liderazgo y experiencia laboral, y los empleados que se encuentran en cargos básicos no son tan relevantes para la toma de decisiones y direccionamiento de la empresa. Sin embargo, en una proporción menor de las 20 investigaciones, se demuestra que muchas veces los miembros de la organización con menor nivel jerárquico o que aparentemente son pocos calificados para participar en el desarrollo estratégico y aportar ideas constructivas, sean los que más se destaquen en las actividades diarias y generen mayor valor e impacto interno (Carrión, 2010).

Ahora bien, otros autores destacan que la posición que tiene el nodo en la red es insignificante a la hora de definir la importancia de un miembro en la red social corporativa, según ellos, no importa si el cargo que ocupa el trabajador en la empresa es directivo, administrativo, básico u operativo, lo que realmente destaca a una persona son las competencias informativas, persuasivas, participativas, expresivas, actitudinales, delegatorias, y generar potencial en los demás (Koch, 1993). Los criterios claves que hacen que un miembro sea importante son: la experticia, que significa dominio, maestría en el hacer y uso instrumental de ese hacer. La maestría implica concentración absoluta, entrega total de la persona y eficiencia. Por otro lado, debe adquirir un alto grado de sinergia completa de sus habilidades técnicas, para llevar a feliz término el desarrollo de su cargo.



## 6. Discusión

Este trabajo es de gran trascendencia en el campo de la literatura, puesto que su contribución se enmarca en realizar una revisión sistémica de modelos basados en el análisis de redes y los métodos de la minería de datos para determinar la importancia de un miembro en una red social corporativa. No obstante, existen diversos estudios de investigación enfocados en aplicar los modelos de SCNA Y DM a sistemas organizacionales, orientados a identificar nodos clave en dicha red. Sin embargo, se evidencia la carencia de un estudio que consolide la aplicabilidad y relevancia que han tenido estos métodos en las redes corporativas y como han contribuido al mejoramiento de la comunicación, transferencia de información, colaboración, participación, ambiente laboral y productividad de las empresas. Este trabajo, pretende llenar este vacío, siendo guía o soporte para futuros investigadores que requieran información y ejemplos aplicativos de los modelos, con el objetivo de analizar el grado de éxito, alcance, precisión y eficiencia de los mismo.

En esta investigación se detecta que los algoritmos utilizados actualmente para determinar la importancia de un miembro en una red social corporativa son los que están asociados a los métodos de las redes sociales y de la minería de datos. Dichos algoritmos han sido aplicados a la solución de problemas o anomalías en el ámbito laboral como: la inadecuada distribución de responsabilidades y tareas, el bajo nivel de participación de algunos miembros, la poca interacción, comunicación y transferencia de información que tienen las diferentes áreas que conforman la empresa y desconocimiento de las habilidades, capacidades y competencias de algunos empleados, generando que las actividades que se realizan no sean valoradas ni reconocidas, lo cual repercute en la desmotivación laboral de estos actores. A partir de lo anterior, los directivos han podido plantear estrategias para mejorar el funcionamiento organizacional, promover el grado de contribución y cooperación de los empleados, para que este no se centre en nodos específicos de la red, sino que todos los miembros evaluados generen un mismo impacto e influencia.

Dentro de las redes sociales se encuentra las medidas locales (grado de centralidad, centralidad de intermediación, de cercanía, del vector propio, de Katz y centralidad de PageRank), y las medidas globales (diámetro y radio, distancia promedio, calificación promedio, densidad y modularidad). Los métodos más usados y que más han contribuido en



el campo de la investigación con respecto a la identificación de nodos claves son los asociados a las medidas locales, puesto que éstas permiten definir y medir la interacción de un miembro y el relacionamiento que tiene dentro de una organización. Con respecto al DM los métodos más usados han sido los algoritmos genéticos, búsqueda tabú, recocido simulado, redes neuronales artificiales (SVM Y KNN), arboles de decisión y K-means. Los algoritmos Apriori, EM, Adaboost y Naive Baye están en un proceso de evolución y desarrollo, puesto que son muy escasas las investigaciones que se encuentran relacionados a la aplicabilidad de estos en el ámbito organización para detectar miembros importantes en una RSC. Son algoritmos que seguramente se seguirán estudiando y validando.

Para llevar a cabo esta investigación, en primer lugar, se plantea una búsqueda en diferentes bases de datos, la cual abarca todos los criterios y restricciones estipuladas. Una vez definida la estrategia de búsqueda, se determina un mecanismo de extracción de datos para la selección de aquella literatura que satisfaga los objetivos principales del proyecto. Para esto, se tiene en cuenta en cada artículo el tipo de modelo utilizado, la clase de modelo, si la investigación se enmarca en un contexto organizacional, las variables utilizadas en dicho estudio, el método de recolección de datos, modelo matemático formulado, software utilizado, funcionamiento de la red corporativa, la validación de los métodos, y las limitaciones. Lo anterior, tiene como fin determinar el grado de confiabilidad de los resultados y lograr que los parámetros definidos en la extracción de datos, permitan que la información obtenida contribuya significativamente en la construcción de este trabajo.

A partir de lo mencionado, se obtienen 20 documentos que cumplen con los criterios planteados. De las veinte investigaciones que conforman este trabajo, nueve de ellas emplean los modelos basados en el análisis de redes sociales corporativas y once se enfocan en los métodos de la minería de datos. Por otro lado, tres de los nueve artículos de SCNA involucran todas las medidas locales y globales, cinco utilizan las medidas de centralidad tales como: grado, intermediación, cercanía, vector propio y Katz, y uno emplea la centralidad de PageRank. Con respecto a los métodos de la minería de datos, de los once estudios seleccionados, dos combinan diferentes técnicas de minería, dos realizan su análisis con base en los algoritmos genéticos, otro utiliza la búsqueda tabú, tres relacionan las redes neuronales artificiales como: SVM Y KNN, uno aplica el método de recocido simulado (Yu Sun et al., 2017), otro usa el K-means y el último incluye los árboles de decisión y las redes neuronales como el SVM.

De acuerdo con la revisión bibliográfica, once de los veinte artículos que hacen parte de este trabajo, demuestran que los métodos utilizados son precisos y generan resultados confiables. Tres de los nueve estudios de SCNA y ocho de las once investigaciones correspondientes a los métodos de minería de datos realizaron validación. Los autores León et al. (2017)



identifican que los métodos del SCNA ofrecen resultados estadísticamente significativos. Yuan Sun et al. (2015) determinan que el modelo PageRank es más preciso, eficiente, tiene una alta escalabilidad y presenta mayor rendimiento en comparación con las demás medidas de centralidad del SCNA y Fire et al. (2013) corroboran que los métodos del SCNA arrojan resultados con un índice de confiabilidad mayor al 80%. Por otro lado, el estudio que implementa los algoritmos genéticos demuestra que los algoritmos genéticos tienen una exactitud en sus resultados del 60% respectivamente. La comparación de los resultados del algoritmo genético con el método de ponderación aleatoria indica que los algoritmos genéticos son 22% más precisos que la ponderación aleatoria, según los datos de la red de interacción de la empresa Enron (Kamali et al., 2020).

Asimismo, al realizar la validación del algoritmo de búsqueda tabú, se obtiene un nivel de confianza en los resultados del  $90\% \pm 5$ . Lo cual indica, que los criterios seleccionados en la realización de la investigación no presentan interdependencias significativas, es decir, los resultados son precisos (Crispim & Pinho de Sousa, 2009). Los autores Nurek & Michalski (2020) en su respectivo estudio de clasificar los miembros de la red social corporativa, implementando los métodos de ANN Y SVM, obtienen que estos algoritmos tienen una precisión del 100%, mientras que los resultados arrojados mediante el método de árboles de decisión cuentan con un índice de confiabilidad del 80%. Por otro lado, se evidencia que la secuencia de nodos importantes que arroja el método de recocido simulado es más precisa, eficaz y efectiva que los demás métodos con los cuales se realiza la comparación (el método de clasificación basado en la entropía de mapeo (MME) y el método de clasificación basado en atributos múltiples (MMA)) (Yu Sun et al., 2017).

Por consiguiente Albino & Garavelli (1998) en su artículo de redes neuronales artificiales, realizan la prueba NN, la cual consiste en la verificación del rendimiento de la red e identifican que entre los principales resultados de la prueba NN, la generalización, es decir, la capacidad de proporcionar una respuesta adecuada en correspondencia de nueva información, y la tolerancia al ruido, es decir, la capacidad de reconocer información poco clara, arrojan resultados verídicos y precisos. Los autores Yelupula & Ramaswamy (2008), aplicando el método K-mean, predijeron la estructura organizativa de Enron con el atributo numérico, sin embargo, no pudieron obtener una precisión del 100%. Con respecto a la implementación de la minería de datos, los investigadores analizan las respuestas y las comparan con los resultados del enfoque, además validan las relaciones fuertes y débiles, así como los puntos críticos en la red social corporativa. Los datos reflejan que los miembros tienen una relación más fuerte y afectiva con profesores de la misma institución que con profesores externos (Silva et al., 2009). Por último, Li et al. (2016) formulan tres ecuaciones de distancia del algoritmo KNN propuesto, y las validan respectivamente, las cuales son: método KNN aplicado con: la Entropía-distancia Euclidiana (EEKNN), Entropía-distancia



Manhattan (EMKNN) y Entropía-distancia Canberra (ECKNN). Los resultados muestran que EEKNN, EMKNN y ECKNN tienen la precisión de clasificación promedio más alta, correspondiendo a un 87.59%, 87.64% y 85.51%.

De este modo, mediante este estudio se determina que el análisis de redes sociales (Wasserman & Faust, 1994) realizado dentro de una organización puede ser útil para hacer que el proceso de identificación de miembros importantes sea más preciso, medible y manejable. El SCNA proporciona la imagen de la estructura organizacional como un conjunto de conexiones entre empleados (por ejemplo, colaboración o flujo de información) que es menos formal, mucho más actual y detallada a la estructura representada por un organigrama (Borgatti & Cross, 2003; Tichy et al., 1979). La literatura demuestra que los conocimientos del análisis de redes y las visualizaciones entran en juego en cada etapa del proceso de selección de nodos claves (Zbieg et al., 2016) y son útiles para encontrar personas capacitadas, eficientes, comprometidas y con gran impacto e influencia para la ejecución de las actividades diarias (Battilana & Casciaro, 2012; Tichy et al., 1979). Sin embargo, los métodos tradicionales de descubrimiento de nodos claves en las redes sociales tienen limitaciones para lidiar con big data (grandes cantidades de datos e información), enfrentando desafíos como los límites de memoria y la gran carga de información que se transfiere por medio de los enlaces que interconectan los nodos.

El SCNA permite a los investigadores examinar la estructura social de un entorno para ayudar a explicar por qué existen ciertos fenómenos dentro de un grupo determinado. Por ejemplo, puede examinar qué tan prominente es un actor dentro de un grupo a través de medidas de centralidad. Este ejemplo puede llevar a la construcción de teorías en áreas como por qué ciertas personas tienen más probabilidades de ser promovidas que otras y por qué algunos departamentos exhiben niveles más altos de colaboración. Por otro lado, la identificación de miembros importantes depende en gran medida de la estructura de la red, específicamente, de los patrones de mezcla (Newman, 2003). Las redes con mezcla selectiva son aquellas en las que los nodos que tienen muchas relaciones están vinculados con otros que también tienen muchas relaciones, mientras que los nodos que tienen pocas relaciones suelen tener relaciones entre sí. Por lo tanto, la centralidad de grado sería un buen método para seleccionar nodos claves en redes con mezcla desasortativa; es decir aquellas en las que los nodos con un gran número de relaciones están conectados con los que tienen un bajo nivel de relacionamiento y viceversa. Sin embargo, las redes sociales suelen mostrar una mezcla clasificada, mientras que la mezcla desasortativa se encuentra en las redes biológicas y físicas.

Por otra parte, la minería de datos abre nuevas oportunidades de negocios, al contribuir a la toma de decisiones tácticas y estratégicas. Esto se debe, a que los modelos de Minería de



Datos pueden predecir valores, buscar relaciones, dependencias y generar resúmenes de datos (Jaramillo et al., 2018). La detección de patrones y descubrimiento se refiere a una colección de enfoques cuyo objetivo es identificar estructuras inusuales o relaciones en los datos. Otra importante aplicación de la minería de datos para generar relaciones de confianza es el análisis de sentimientos. Mediante el análisis de sentimientos es posible comprender la actitud que tiene la persona que escribe sobre una temática o la polaridad contextual (positiva, negativa o neutra) que expresa (Jaramillo et al., 2018). Con estos métodos las organizaciones pueden analizar continuamente grandes volúmenes de datos, automatizar las decisiones rutinarias y las más importantes sin tener que dilatarlas a la espera de que alguien decida qué debe hacerse. De este modo, si se implementan las técnicas de minería de datos en un sistema empresarial, se puede recopilar, analizar y decidir qué hacer a partir de los datos de forma independiente para optimizar la toma de decisiones y mejorar los procesos diarios de una organización.

Las dificultades asociadas al almacenamiento y gestión de grandes volúmenes de datos son prolíficas y se extienden a todos los campos que recopilan, almacenan y analizan datos. Los big data (conjunto de datos grandes y complejos) se caracterizan por cuatro dificultades significantes: volumen, variedad, veracidad y velocidad. El objetivo de la minería de datos es servir de solución para obtener el verdadero valor de los datos. El volumen se refiere a la dificultad de almacenar y procesar la cantidad ingente de datos recopilados por las organizaciones. La variedad se refiere a los numerosos tipos de datos recopilados y almacenados. Las herramientas de minería de datos deben poder procesar de forma simultánea toda una serie de formatos de datos y la velocidad detalla la mayor celeridad con la que se crean, recopilan y almacenan datos nuevos (Silva et al., 2009).

Por otro lado, es importante mencionar que éste trabajo permitió identificar los aspectos y criterios relevantes usados en el campo de la investigación teórica para determinar el relacionamiento efectivo y predecir la importancia de un miembro en una red social corporativa. Según los 20 artículos que conforman este estudio, la posición que tiene el nodo en la red social corporativa en la mayoría de los casos indica el alcance que tiene el trabajador en la organización, es decir, los que pertenecen al personal de alta dirección, tienden a ser empleados indispensables en la red por su gran capacidad de conocimientos, liderazgo y experiencia laboral, y los empleados que se encuentran en cargos básicos no son tan relevantes para la toma de decisiones y direccionamiento de la empresa. Sin embargo, en una proporción menor de las 20 investigaciones, se demuestra que muchas veces los miembros de la organización con menor nivel jerárquico o que aparentemente son pocos calificados para participar en el desarrollo estratégico y aportar ideas constructivas, sean los que más se destaquen en las actividades diarias y generen mayor valor e impacto interno (Carrión, 2010).

Ahora bien, otros autores destacan que la posición que tiene el nodo en la red es insignificante a la hora de definir la importancia de un miembro en la red social corporativa, según ellos, no importa si el cargo que ocupa el trabajador en la empresa es directivo, administrativo,



básico u operativo, lo que realmente destaca a una persona son las competencias informativas, persuasivas, participativas, expresivas, actitudinales, delegatorias, y generar potencial en los demás (Koch, 1993). Los criterios claves que hacen que un miembro sea importante son: la experticia, que significa dominio, maestría en el hacer y uso instrumental de ese hacer. La maestría implica concentración absoluta, entrega total de la persona y eficiencia. Por otro lado, debe adquirir un alto grado de sinergia completa de sus habilidades técnicas, para llevar a feliz término el desarrollo de su cargo.

Las características que determinan la importancia de un nodo en una red corporativa son: el liderazgo y la inteligencia emocional. No es que el coeficiente intelectual y las destrezas técnicas sean irrelevantes. Son importantes, pero como “aptitudes de umbral”; es decir, son los requisitos básicos para puestos ejecutivos. Pero estudios recientes, muestran claramente que la inteligencia emocional es una condición crucial que debe de tener un empleado que se considere clave (TORRES, 2011). Sin ella, una persona puede tener la mejor preparación del mundo, una mente incisiva y analítica, y un infinito surtido de ideas inteligentes, pero aun así no logrará influenciar, controlar, planificar, guiar, dirigir, ayudar, trazar metas, formular valores, crear y transformar la cultura (TORRES, 2011).

Según García et al. (2018) los criterios y destrezas generales que convierten a un individuo en alguien productivo, significativo, participativo e indispensable son las habilidades técnicas, las cuales integran los conocimientos acerca de los métodos, procesos, procedimientos y técnicas para la conducción de las actividades de la unidad de trabajo del líder. Así mismo, son relevantes las habilidades interpersonales, que corresponden a la conducta y procesos interpersonales, habilidad para comprender los sentimientos, actitudes y motivos de otros a partir de lo que dicen o hacen, es la habilidad para comunicarse de manera clara y persuasiva, para establecer relaciones cooperativas (tacto, diplomacia, encanto, empatía, sensibilidad social, fluidez en el lenguaje, etc.). Por último, se deben de tener en cuenta las habilidades conceptuales, que se refieren a la habilidad analítica general, pensamiento lógico, eficiencia en la formación de conceptos y conceptualización de relaciones ambiguas y complejas, creatividad en la generación de ideas y solución de problemas, habilidad para analizar los hechos, percibir las tendencias, anticipar los cambios y reconocer las oportunidades y problemas potenciales.

Por último, la revisión sistémica de modelos basados en el análisis de redes y los métodos de la minería de datos para determinar la importancia de un miembro en una red social corporativa cuenta con fortalezas y limitaciones. Dentro de las fortalezas que tiene esta revisión bibliográfica, merece mencionarse que esta permitió consignar de forma detallada cada uno de los estudios encontrados en el campo de la investigación que han aplicado o hecho uso de estos métodos para el fin buscado. Además, toda la información fue extraída de bases de datos confiables, teniendo en cuenta los criterios y parámetros de búsqueda, para



obtener una calidad en el proceso de extracción de datos. Por otro lado, en la revisión sistémica, se logró encontrar ejemplos claros y representativos del aporte de los modelos en el ámbito empresarial. Así mismo, es importante mencionar que este trabajo de grado aportó en la elaboración de dos artículos. El primero es denominado “Identificación de líderes potenciales mediante el análisis de redes sociales”, realizado por los autores Leung Alex P, Wang Z et al, documento que se encuentra en proceso de publicación y participan las autoras. El segundo artículo se deriva en su totalidad de este trabajo de grado, realizado por las mismas autoras; este artículo se encuentra en proceso de construcción.

Por otra parte, una gran limitación de la construcción de este documento es que no realizó una búsqueda más amplia que incluyera diferentes bases de datos, por lo que pueden existir estudios omitidos y algunos otros relevantes que no fueron considerados, o que no incluyeron los términos de búsqueda en el título, resumen o palabras clave. Del mismo modo, existieron documentos resultantes que se basaban en la implementación de estos métodos, pero el enfoque no estaba relacionado con sistemas organizacionales.



## 7. Bibliografía

- Albino, V., & Garavelli, A. C. (1998). A neural network application to subcontractor rating in construction firms. *International Journal of Project Management*, 16(1), 9–14. [https://doi.org/10.1016/S0263-7863\(97\)00007-0](https://doi.org/10.1016/S0263-7863(97)00007-0)
- Anguita, D., & Boni, A. (2002). Improved neural network for SVM learning. *IEEE Transactions on Neural Networks*, 13(5), 1243–1244. <https://doi.org/10.1109/TNN.2002.1031958>
- Arif, M., Alam, K. A., & Hussain, M. (2015). Application of Data Mining Using Artificial Neural Network: Survey. *International Journal of Database Theory and Application*, 8(1), 245–270. <https://doi.org/10.14257/ijdta.2015.8.1.25>
- Bailyn, L., Sproull, L., & Kiesler, S. (1992). Connections: New Ways of Working in the Networked Organization. *Administrative Science Quarterly*, 37(3), 491. <https://doi.org/10.2307/2393454>
- Bar, F., Kane, N., & Simard, C. (2000). Digital networks and Organizational Change: The Evolutionary Deployment of Corporate Information Infrastructure. *International Sunbelt Social Network Conference*, 28.
- Basogain, X. (2015). *REDES NEURONALES ARTIFICIALES Y SUS APLICACIONES*. [https://ocw.ehu.eus/file.php/102/redes\\_neuro/contenidos/pdf/libro-del-curso.pdf](https://ocw.ehu.eus/file.php/102/redes_neuro/contenidos/pdf/libro-del-curso.pdf)
- Battilana, J., & Casciaro, T. (2012). Change Agents, Networks, and Institutions: A Contingency Theory of Organizational Change. *Academy of Management Journal*, 55(2), 381–398. <https://doi.org/10.5465/amj.2009.0891>



- Borgatti, S. P., & Cross, R. (2003). A Relational View of Information Seeking and Learning in Social Networks. *Management Science*, 49(4), 432–445. <https://doi.org/10.1287/mnsc.49.4.432.14428>
- Can, U., & Alatas, B. (2019). A new direction in social network analysis: Online social network analysis problems and applications. *Physica A: Statistical Mechanics and Its Applications*, 535, 122372. <https://doi.org/10.1016/j.physa.2019.122372>
- Carley, K. M., & Svoboda, D. M. (1996). Modeling Organizational Adaptation as a Simulated Annealing Process. *Sociological Methods & Research*, 25(1), 138–168. <https://doi.org/10.1177/0049124196025001005>
- Carrión, M. J. (2010). *Análisis de la importancia y efectos de la cultura organizacional en el comportamiento de los miembros de Porta* [PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR]. <http://repositorio.puce.edu.ec/bitstream/handle/22000/8162/T-PUCE-3225.pdf?sequence=1&isAllowed=y>
- Chan, F. T., & Chung, S. (2004). A multi-criterion genetic algorithm for order distribution in a demand driven supply chain. *International Journal of Computer Integrated Manufacturing*, 17(4), 339–351. <https://doi.org/10.1080/09511920310001617022>
- Chen, Chiang, & Storey. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165. <https://doi.org/10.2307/41703503>
- Cook, K. S., & Emerson, R. M. (1978). Power, Equity and Commitment in Exchange Networks. *American Sociological Review*, 43(5), 721. <https://doi.org/10.2307/2094546>



- Crispim, J. A., & Pinho de Sousa, J. (2009). Partner selection in virtual enterprises: A multi-criteria decision support approach. *International Journal of Production Research*, 47(17), 4791–4812. <https://doi.org/10.1080/00207540902847348>
- de Toni, A. F., & Nonino, F. (2010). The key roles in the informal organization: A network analysis perspective. *The Learning Organization*, 17(1), 86–103. <https://doi.org/10.1108/09696471011008260>
- Dehmer, M. (Ed.). (2011). *Structural Analysis of Complex Networks*. Birkhäuser Boston. <https://doi.org/10.1007/978-0-8176-4789-6>
- Fire, M., Puzis, R., & Elovici, Y. (2013). Organization Mining Using Online Social Networks. *ArXiv:1303.3741 [Physics]*. <http://arxiv.org/abs/1303.3741>
- Fraley, C., & Raftery, A. E. (2002). Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association*, 97(458), 611–631. <https://doi.org/10.1198/016214502760047131>
- Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1), 119–139. <https://doi.org/10.1006/jcss.1997.1504>
- García, D., Lozano, J., & Pacheco, A. (2018). *LAS EMPRESAS DEL FUTURO UNA TENDENCIA PARA CARACTERIZAR EL LIDERAZGO [UNIVERSIDAD COOPERATIVA DE COLOMBIA]*. [https://repository.ucc.edu.co/bitstream/20.500.12494/7076/1/2018\\_%20empresas\\_futuro\\_tendencia.pdf](https://repository.ucc.edu.co/bitstream/20.500.12494/7076/1/2018_%20empresas_futuro_tendencia.pdf)
- Gaur, P. (2012). Neural Networks in Data Mining. *International Journal of Electronics and Computer Science Engineering*, 1449–1453. <https://doi.org/2277-1956>



- Ghali, N., Panda, M., Hassanien, A. E., Abraham, A., & Snasel, V. (2012). Social Networks Analysis: Tools, Measures and Visualization. In A. Abraham (Ed.), *Computational Social Networks* (pp. 3–23). Springer London. [https://doi.org/10.1007/978-1-4471-4054-2\\_1](https://doi.org/10.1007/978-1-4471-4054-2_1)
- Glover, F. (1989). Tabu Search—Part I. *ORSA Journal on Computing*, 1(3), 190–206. <https://doi.org/10.1287/ijoc.1.3.190>
- Glover, F., & Laguna, M. (1998). Tabu Search. In D.-Z. Du & P. M. Pardalos (Eds.), *Handbook of Combinatorial Optimization* (pp. 2093–2229). Springer US. [https://doi.org/10.1007/978-1-4613-0303-9\\_33](https://doi.org/10.1007/978-1-4613-0303-9_33)
- Grassi, R., Stefani, S., & Torriero, A. (2010). Centrality in organizational networks. *International Journal of Intelligent Systems*, 25(3), 253–265. <https://doi.org/10.1002/int.20400>
- Hand, D. J., & Adams, N. M. (2015). Data Mining. In N. Balakrishnan, T. Colton, B. Everitt, W. Piegorisch, F. Ruggeri, & J. L. Teugels (Eds.), *Wiley StatsRef: Statistics Reference Online* (pp. 1–7). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118445112.stat06466.pub2>
- HATALA, J.-P. (2006). Social Network Analysis in Human Resource Development: A New Methodology. *Sage Publications*, 5(1), 45–71. <https://doi.org/10.1177/1534484305284318>
- Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2010). *Metodología de la investigación* (5a ed). McGraw-Hill.
- Herrera, R. Y. (2013). Minería de proceso como herramienta para la auditoría. *Ciencias de La Información*, 44(2), 25–32.



- Hoppe, B., & Reinelt, C. (2010). Social network analysis and the evaluation of leadership networks. *The Leadership Quarterly*, 21(4), 600–619. <https://doi.org/10.1016/j.leaqua.2010.06.004>
- Hu, W., Hu, W., & Maybank, S. (2008). AdaBoost-Based Algorithm for Network Intrusion Detection. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS*, 38(2), 577–583.
- Jaramillo, S., Espitia, E., & Cardona, S. (2018). Minería sobre streams de datos: Una nueva estrategia para afianzar las relaciones de confianza en el mundo empresarial. *Revista ESPACIOS*, 39(13), 6.
- Jiang, Y., & Zhou, Z.-H. (2004). Editing Training Data for kNN Classifiers with Neural Network Ensemble. In F.-L. Yin, J. Wang, & C. Guo (Eds.), *Advances in Neural Networks – ISNN 2004* (Vol. 3173, pp. 356–361). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-540-28647-9\\_60](https://doi.org/10.1007/978-3-540-28647-9_60)
- Kamali, H., Rahmani, H., & Shah-Hosseini, H. (2020). DINGA: A Genetic-algorithm-based Method for Finding Important Nodes in Social Networks. *Journal of AI and Data Mining*, 12.
- Kantarci, B., & Labatut, V. (2013). Classification of Complex Networks Based on Topological Properties. *2013 International Conference on Cloud and Green Computing*, 297–304. <https://doi.org/10.1109/CGC.2013.54>
- Keerthi, S. S., & Gilbert, E. G. (2002). Convergence of a Generalized SMO Algorithm for SVM Classifier Design. *Machine Learning*, 46(1/3), 351–360. <https://doi.org/10.1023/A:1012431217818>



- Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science*, 220(4598), 671–680. <https://doi.org/10.1126/science.220.4598.671>
- Koch, S. (1993). Características motivacionales y actitudinales de supervisores. *Revista Interamericana de Psicología Ocupacional*, 12(2), 91–105.
- León, R. – D., Rodríguez-Rodríguez, R., Gómez-Gasquet, P., & Mula, J. (2017). Social network analysis: A tool for evaluating and predicting future knowledge flows from an insurance organization. *Technological Forecasting and Social Change*, 114, 103–118. <https://doi.org/10.1016/j.techfore.2016.07.032>
- Li, N., Kong, H., Ma, Y., Gong, G., & Huai, W. (2016). Human performance modeling for manufacturing based on an improved KNN algorithm. *The International Journal of Advanced Manufacturing Technology*, 84(1–4), 473–483. <https://doi.org/10.1007/s00170-016-8418-6>
- Li, Y., & Wu, H. (2012). A Clustering Method Based on K-Means Algorithm. *Physics Procedia*, 25, 1104–1109. <https://doi.org/10.1016/j.phpro.2012.03.206>
- Lü, L., & Zhou, T. (2011). Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and Its Applications*, 390(6), 1150–1170. <https://doi.org/10.1016/j.physa.2010.11.027>
- Luna, A. (2015). *IDENTIFICACIÓN DE LOS LÍDERES DE OPINIÓN MEDIANTE EL ANÁLISIS DE REDES SOCIALES EN UNA EMPRESA CONSTRUCTORA*. [UNIVERSITAT POLITÈCNICA DE VALÈNCIA]. <https://riunet.upv.es/bitstream/handle/10251/57344/PamelaLunaTFM.pdf?sequence=1&isAllowed=y>



- Masangcap, M. L. G., Sison, A. M., & Medina, R. P. (2018). Application of enhanced expectation maximization (EnEM) algorithm for image segmentation. *Proceedings of the 2018 International Conference on Data Science and Information Technology - DSIT '18*, 74–78. <https://doi.org/10.1145/3239283.3239310>
- Michalski, R., Kazienko, P., & Krol, D. (2012). Predicting Social Network Measures Using Machine Learning Approach. *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 1056–1059. <https://doi.org/10.1109/ASONAM.2012.183>
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.
- Mukkamala, S., Janoski, G., & Sung, A. (2002). Intrusion detection using neural networks and support vector machines. *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290)*. <https://doi.org/10.1109/IJCNN.2002.1007774>
- Murthy, S., & Salzberg, S. (1995). Decision tree induction: How effective is the greedy heuristic? *KDD'95: Proceedings of the First International Conference on Knowledge Discovery and Data Mining*, 222–227.
- Nafie Ali, F. M., & Mohamed Hamed, A. A. (2018). Usage Apriori and clustering algorithms in WEKA tools to mining dataset of traffic accidents. *Journal of Information and Telecommunication*, 2(3), 231–245. <https://doi.org/10.1080/24751839.2018.1448205>
- Newman, M. E. J. (2003). The Structure and Function of Complex Networks. *SIAM Review*, 45(2), 167–256. <https://doi.org/10.1137/S003614450342480>



- Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2), 026113. <https://doi.org/10.1103/PhysRevE.69.026113>
- Nurek, M., & Michalski, R. (2020). Combining Machine Learning and Social Network Analysis to Reveal the Organizational Structures. *Applied Sciences*, 10(5), 1699. <https://doi.org/10.3390/app10051699>
- Olguin, D. O., Waber, B. N., Taemie Kim, Mohan, A., Ara, K., & Pentland, A. (2009). Sensible Organizations: Technology and Methodology for Automatically Measuring Organizational Behavior. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(1), 43–55. <https://doi.org/10.1109/TSMCB.2008.2006638>
- Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of Information Science*, 28(6), 441–453. <https://doi.org/grassi>
- Parvin, H., Alizadeh, H., & Minaei-Bidgoli, B. (2008). MKNN: Modified K-Nearest Neighbor. *Proceedings of the World Congress on Engineering and Computer Science*. [http://www.iaeng.org/publication/WCECS2008/WCECS2008\\_pp831-834.pdf](http://www.iaeng.org/publication/WCECS2008/WCECS2008_pp831-834.pdf)
- Pham, D. T., & Karaboga, D. (2000). *Intelligent Optimisation Techniques*. Springer London. <https://doi.org/10.1007/978-1-4471-0721-7>
- Riquelme, J. C., Ruiz, R., & Gilbert, K. (2006). *Minería de Datos: Conceptos y Tendencias*. 10(29), 11–18.
- Scott, J. (1988). Social Network Analysis. *Sociology*, 22(1), 109–127. <https://doi.org/10.1177/0038038588022001007>



- Shetty, J., & Adibi, J. (2005). Discovering important nodes through graph entropy the case of Enron email database. *Proceedings of the 3rd International Workshop on Link Discovery - LinkKDD '05*, 74–81. <https://doi.org/10.1145/1134271.1134282>
- Siddiqi, A., Targio, I., Yaqoob, I., Marjani, M., Shamshirband, S., Gani, A., & Nasaruddin, F. (2016). A survey of big data management: Taxonomy and state-of-the-art. *Journal of Network and Computer Applications*, 71, 151–166. <https://doi.org/10.1016/j.jnca.2016.04.008>
- Silva, R., Menezes, V., Oliveira, J., de Souza, M. F., de Mello, C. E. R., Souza, J. M., & Zimbrao, G. (2009). Mining and analyzing organizational social networks for collaborative design. *2009 13th International Conference on Computer Supported Cooperative Work in Design*, 504–509. <https://doi.org/10.1109/CSCWD.2009.4968109>
- Sivanandam, S. N., & Deepa, S. N. (2008). Genetic Algorithms. In *Introduction to Genetic Algorithms* (pp. 15–37). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-540-73190-0\\_2](https://doi.org/10.1007/978-3-540-73190-0_2)
- Snow, C. C., & Fjeldstad, Ø. D. (2015). Network Paradigm: Applications in Organizational Science. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 546–550). Elsevier. <https://doi.org/10.1016/B978-0-08-097086-8.73107-2>
- Solano, P. A., & Murillo, H. G (2021). Criterios para determinar relacionamiento efectivo. Anexo de Excel
- Sparrowe, R. T., Liden, R. C., Wayne, S. J., & Kraimer, M. L. (2001). Social Networks and the Performance of Individuals and Groups. *Academy of Management Journal*, 44(2), 316–325. <https://doi.org/10.5465/3069458>



- Sun, Yu, Yao, P.-Y., Wan, L.-J., Shen, J., & Zhong, Y. (2017). Ranking important nodes in complex networks by simulated annealing. *Chinese Physics B*, 26(2), 020201. <https://doi.org/10.1088/1674-1056/26/2/020201>
- Sun, Yuan, Ma, Y., Zhang, F., Ma, Y., & Shen, W. (2015). Key Nodes Discovery in Large-Scale Logistics Network Based on MapReduce. *2015 IEEE International Conference on Systems, Man, and Cybernetics*, 1309–1314. <https://doi.org/10.1109/SMC.2015.233>
- Tam, K. Y., & Kiang, M. Y. (1992). Managerial Applications of Neural Networks: The Case of Bank Failure Predictions. *Management Science*, 38(7), 926–947. <https://doi.org/10.1287/mnsc.38.7.926>
- Teixeira, R. M., Andreassi, T., Köseoglu, M. A., & Okumus, F. (2019). How do hospitality entrepreneurs use their social networks to access resources? Evidence from the lifecycle of small hospitality enterprises. *International Journal of Hospitality Management*, 79, 158–167. <https://doi.org/10.1016/j.ijhm.2019.01.006>
- Tichy, N. M., Tushman, M. L., & Fombrun, C. (1979). Social Network Analysis For Organizations. *Academy of Management Review*, 4(4), 507–519. <https://doi.org/10.5465/amr.1979.4498309>
- TORRES, G. P. (2011). *LIDERAZGO Y DIRECCIÓN: DOS CONCEPTOS DISTINTOS CON RESULTADOS DIFERENTES*. XIX (1), 213–228.
- Vapnik, V. (2010). *The Nature of Statistical Learning Theory* (2., nd ed. Softcover version of original hardcover edition 2000). Springer New York.
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511815478>



- Weissman, D. B., Jackson, S. H., Heicher, D. A., & Rockoff, M. A. (1990). Benzyl Alcohol Administration in Neonates: *Anesthesia & Analgesia*, 70(6), 673-674. <https://doi.org/10.1213/00000539-199006000-00025>
- Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z.-H., Steinbach, M., Hand, D. J., & Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1), 1-37. <https://doi.org/10.1007/s10115-007-0114-2>
- Xiong, L., Chitti, S., & Liu, L. (2007). Mining multiple private databases using a kNN classifier. *Proceedings of the 2007 ACM Symposium on Applied Computing - SAC '07*, 435. <https://doi.org/10.1145/1244002.1244102>
- Xu, S. (2018). Bayesian Naïve Bayes classifiers to text classification. *Journal of Information Science*, 44(1), 48-59. <https://doi.org/10.1177/0165551516677946>
- Yamadori, L., Morikawa, T., Kobayashi, S., & Ohmori, M. (1990). Gauchev's Disease Type I: Report of a Case with Prominent Deposition of Ceroid in Splenic Endothelial Cells and Intestinal Smooth Muscle Fibers. *The Japanese Society of Pathology*, 40(6), 425-430.
- Yaqoob, I., Hashem, I. A. T., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N. B., & Vasilakos, A. V. (2016). Big data: From beginning to future. *International Journal of Information Management*, 36(6), 1231-1247. <https://doi.org/10.1016/j.ijinfomgt.2016.07.009>
- Yelupula, K., & Ramaswamy, S. (2008). Social network analysis for email classification. *Proceedings of the 46th Annual Southeast Regional Conference on XX - ACM-SE 46*, 469. <https://doi.org/10.1145/1593105.1593229>



- 
- Zbieg, A., Batorski, D., & Żak, B. (2016). How to Select Change Agents in Organizations? A Comparison of the Classical and Network Approaches. *Problemy Zarzadzania*, 14(4 (64)), 120–143. <https://doi.org/10.7172/1644-9584.64.7>
- Zhang, C., & Fang, Z. (2012). An improved K-Means clustering algorithm. *2008 Congress on Image and Signal Processing*, 193–199.
- Zhang, G. P. (2000). Neural networks for classification: A survey. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, 30(4), 451–462. <https://doi.org/10.1109/5326.897072>

