



Ciencia e Ingeniería Neogranadina
ISSN: 0124-8170
revistaing@unimilitar.edu.co
Universidad Militar Nueva Granada
Colombia

Calvache, Juan Felipe; Pérez, Sergio Andrés; Moreno, Francisco Javier
MODELO BASADO EN FACTORES Y REGLAS PARA CUANTIFICAR LA AFINIDAD
ENTRE LOS INDIVIDUOS DE UN GRUPO HUMANO
Ciencia e Ingeniería Neogranadina, vol. 25, núm. 2, 2015, pp. 117-136
Universidad Militar Nueva Granada
Bogotá, Colombia

Disponible en: <http://www.redalyc.org/articulo.oa?id=91142868007>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

MODELO BASADO EN FACTORES Y REGLAS PARA CUANTIFICAR LA AFINIDAD ENTRE LOS INDIVIDUOS DE UN GRUPO HUMANO

A MODEL BASED ON FACTORS AND RULES TO QUANTIFY THE AFFINITY AMONG THE INDIVIDUALS OF A HUMAN GROUP

Juan Felipe Calvache¹, Sergio Andrés Pérez², Francisco Javier Moreno³

Fecha de recepción: 27 de Septiembre de 2014

Fecha de aprobación: 22 de Julio de 2015

Referencia: J.F. Calvache, S.A. Pérez, F.J. Moreno. (2015). Reglas para cuantificar la afinidad entre los individuos de un grupo humano. Ciencia e Ingeniería Neogranadina, 25 (2), pp. 117 - 136, DOI: <http://dx.doi.org/10.18359/rcin.1435>

RESUMEN

En este artículo se propone un modelo para cuantificar el grado de afinidad entre los individuos de un grupo humano. Para obtener el grado de afinidad se considera un conjunto de factores y de reglas (para cada factor) definidos por el analista. El modelo se puede aplicar prácticamente a cualquier grupo humano: estudiantes, trabajadores, miembros de una red social, etc. Para validar y mostrar la utilidad del modelo, se analizaron dos grupos de estudiantes de cursos universitarios. Los datos de los estudiantes, correspondientes a los factores definidos para los experimentos, se recopilaron mediante una encuesta que fue diseñada para tal efecto. Aunque se requieren experimentos más exhaustivos, los resultados evidenciaron posibles patrones; e.g., los grupos de estudiantes con mayor grado de afinidad fueron los de mayor calificación promedio grupal. También se observó que existen ciertos individuos que tienden a ser miembros de los grupos más afines y otros que tienden a ser miembros de los grupos menos afines.

Palabras clave: Afinidad, Comunidades, Grupos Humanos, Redes Sociales, Relaciones Sociales.

1. Ingeniero de Sistemas, Universidad Nacional de Colombia sede Medellín, Medellín, Colombia, jfcalvache@unal.edu.co

2. Ingeniero de Sistemas, Universidad Nacional de Colombia sede Medellín, Medellín, Colombia, seaperezpe@unal.edu.co

3. Ingeniero de Sistemas, Phd en Ingeniería, Profesor Asociado, Universidad Nacional de Colombia sede Medellín, Medellín, Colombia, fjmoreno@unal.edu.co

ABSTRACT

In this paper we propose a model to quantify the degree of affinity among the individuals of a human group. To obtain the degree of affinity, our model considers a set of factors and a set of rules (for each factor) which are defined by the analyst. Our model can be virtually applied to any human group: students, employees, members of a social network, etc. To validate and show the expediency of our model, we analyzed two groups of university courses. The students' data, corresponding to the factors identified for the experiments, were collected using a survey that was designed for this purpose. Although more extensive experiments are required, our results showed possible patterns, e.g., the groups of students with higher affinity were the groups with the highest average grades. It was also noted that there are certain individuals who tend to be members of the higher affinity groups and some who tend to be members of the lowest affinity groups.

Keywords: Affinity, Communities, Human Groups, Social Networks, Social Relationships.

INTRODUCCIÓN

En este artículo se propone un modelo para cuantificar el grado de afinidad entre los individuos de un grupo humano considerando sus preferencias personales y elementos en común de acuerdo con el contexto cultural, social y geográfico en el que se encuentren.

La noción de afinidad en grupos humanos ha tenido diversas aproximaciones. En [1] se propone un modelo para conformar equipos de trabajo. Los individuos se agrupan de acuerdo con un conjunto de criterios preestablecidos que incluyen el género, temas de interés, áreas de trabajo, tipos de proyectos en los que han participado, tipos de tareas que prefieren hacer, entre otros. Algunos de estos datos son extraídos de Facebook y LinkedIn, y otros son capturados por medio de un aplicativo diseñado por los autores. Sin embargo, el modelo no considera ciertos aspectos que influyen en las relaciones humanas, como tipos

de personalidad, ubicación geográfica, tendencias políticas y religiosas, entre otros.

Panigrahy [2] y Bapna [3] enfocan su trabajo a las redes sociales. Panigrahy propone una medida que considera el número de caminos disjuntos que existen entre un par de nodos para calcular la afinidad entre los usuarios de la red. Esta medida de afinidad representa la probabilidad de que un par dado de nodos permanezcan conectados cuando se elimina una parte de la red. Por su parte, Bapna propone un modelo basado en la confianza que tiene una persona en otra. Se proponen tres medidas: i) la intimidad, i.e., el número de amigos comunes entre dos personas, ii) la intensidad, i.e., el número de publicaciones en el muro (*wall*) intercambiados entre dos personas, y iii) el número de fotos etiquetadas (*photos tagged*), i.e., el número de fotos en las que se etiqueta el remitente en la foto del receptor o viceversa.

Por otro lado, en [4] se analizan las afinidades positivas y negativas en una red social. Cada

individuo (denominado actor) tiene un conjunto de preferencias y de afinidades positivas y negativas con respecto a los otros actores. Con base en estos elementos, se propone un modelo probabilístico de relaciones entre los actores y se concluye que la probabilidad de una relación es mayor si se percibe reciprocidad entre los actores.

En [5] se analizan las afinidades implícitas y explícitas en una red de individuos. Las afinidades explícitas surgen de las conexiones explícitas de los individuos en la red, y las implícitas surgen naturalmente a partir del comportamiento y actitudes de los individuos. Los autores plantean una medida de afinidad de los individuos basados en el número de valores (y no en factores y reglas ponderadas, como se propone aquí) compartidos en sus atributos.

En [6] se presenta un modelo para distribuir y organizar a los trabajadores de una empresa considerando los factores de compatibilidad de los individuos. Estos factores incluyen la disposición y facilidad para trabajar en equipo, la seguridad, la responsabilidad, la autoestima y la sinceridad. Eaglea [7] propone un modelo para deducir relaciones de amistad a partir de un sistema de telefonía (celulares).

Según Zhang y Dantu [8] las relaciones sociales cambian con el tiempo. Los autores proponen medir un índice de reciprocidad, el cual se obtiene a partir del registro de llamadas telefónicas entre los usuarios y se aplica el modelo Arima (*Autoregressive Integrated Moving Average*, un modelo estadístico usado en estadística y econometría) [9] para analizar las relaciones sociales durante un periodo. Por otro lado, en [10] se propone un sistema recomendador de películas basado en una red

social donde se analizan las relaciones entre los usuarios y se generan valores de afinidad grupales a partir de los perfiles de estos.

El modelo aquí propuesto soporta distintos factores que podrían influir en las relaciones de los individuos. Estos factores son definidos por el analista, i.e., no están predeterminados en el modelo a diferencia de los trabajos descritos. El modelo también permite tantas reglas (asociadas con cada factor) como el analista considere pertinentes para determinar el valor que tiene cada factor definido. Además, el modelo se puede aplicar a cualquier grupo humano, e.g., los estudiantes de un curso, los trabajadores de una empresa, los jugadores de un equipo, i.e., no solo a grupos enmarcados en el contexto de las redes sociales [11-12].

El artículo está organizado de la siguiente manera. En la primera sección se presenta el modelo para afinidad de dos individuos y su correspondiente extensión para grupos de más de dos individuos. En la segunda sección se presentan los experimentos y su análisis. En la tercera sección se presentan las conclusiones y se plantean trabajos futuros.

1. MODELO PROPUESTO

1.1 DEFINICIÓN DEL MODELO

El objetivo del modelo es cuantificar el grado de afinidad entre los individuos de un grupo humano de acuerdo con un conjunto de factores y reglas (asociadas con cada factor) definidas por el analista. Por ejemplo, el analista podría considerar factores como la música, la ubicación geográfica, el idioma, las tendencias políticas, religiosas y culturales

[13-14]. Un ejemplo de una regla para el factor música puede ser: "si el género musical de dos individuos es el mismo, entonces el valor de su afinidad es 1, 0 de lo contrario".

Se define una matriz simétrica de afinidad A para n individuos de un grupo humano, como se muestra en la figura 1.

$$\begin{matrix} & & 1 & 2 & \dots & n \\ 1 & \left[\begin{matrix} a_{11} & a_{12} & \dots & a_{1n} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{matrix} \right] \\ 2 \\ \dots \\ n \end{matrix}$$

Figura 1. Matriz de afinidad A .
Fuente: Elaboración propia.

Donde a_{ij} representa el grado de afinidad entre los individuos i y j . Si $i = j$, entonces $a_{ii} = 0$. Para calcular el valor de cada a_{ij} se aplica la ecuación (1).

$$a_{ij} = \sum_{k=1}^m W_k * (v_{ij})_k \quad (1)$$

Donde W_k , $0 < W_k \leq 1$, es el peso asignado por el analista al factor k y $(v_{ij})_k$ representa el valor de afinidad para los individuos i y j en el factor k . Por lo tanto, a_{ij} es el resultado de ponderar un conjunto de m factores definidos por el analista. Además, se debe cumplir la restricción de la ecuación (2).

$$\sum_{k=1}^m W_k = 1 \quad (2)$$

A su vez, $(v_{ij})_k$ es el resultado de ponderar un conjunto de z reglas y se calcula mediante la ecuación (3).

$$\sum_{i=1}^z p_i = 1 \quad (3)$$

Donde p_q , $0 < p_q \leq 1$, es el peso asignado por el analista a la regla q y $(r_{ij})_q$ es el valor de afinidad para los individuos i y j en la regla q . Además, se debe cumplir la restricción de la ecuación (4).

$$\sum_{i=1}^z p_i = 1 \quad (4)$$

Por ejemplo, un analista podría definir los valores para una regla q así:

$$(r_{ij})_q = \begin{cases} 0 & \text{Si no existe relación entre } i \text{ y } j \text{ con respecto a la regla } q. \\ 0.5 & \text{Si la relación entre } i \text{ y } j \text{ con respecto a la regla } q \text{ es débil.} \\ 1 & \text{Si la relación entre } i \text{ y } j \text{ con respecto a la regla } q \text{ es fuerte.} \end{cases}$$

1.1.1 Ejemplo

A continuación se calculará la matriz de afinidad A para cuatro individuos (numerados del 1 al 4). Se consideran tres factores ($m = 3$): 1. Música, 2. Ubicación geográfica y 3. Tendencia política, véase la tabla 1.

1.1.2 Reglas para el factor música

Se consideran dos reglas ($z = 2$) para este factor: 1. Género favorito y 2. Artista favorito:

- $(r_{ij})_{\text{Género_favorito}} = 1$ si el género favorito de la música es igual, 0 de lo contrario.
- $(r_{ij})_{\text{Artista_favorito}} = 1$ si el artista favorito es el mismo, 0.5 si los dos individuos tienen el mismo género musical y el artista favorito de cada individuo pertenece a ese género, 0.3 si los dos individuos tienen diferente género

Tabla 1. Reglas de afinidad.

| Factor | Regla | Individuo | | | |
|----------------------|------------------|-------------|---------------------|-------------|---------------------|
| | | 1 | 2 | 3 | 4 |
| Música | Género favorito | Reggaeton | Salsa | Reggaeton | Rock en español |
| | Artista favorito | J. Balvin | Gilberto Santa Rosa | Don Omar | Héroes del silencio |
| Ubicación geográfica | Ciudad | Medellín | Santuario | Medellín | Bogotá |
| | Zona | Rural | Urbana | Urbana | Rural |
| | Población | Media | Pequeña | Media | Grande |
| Tendencia política | Partido | Conservador | Liberal | Liberal | Conservador |
| | Ideología | Capitalismo | Socialismo | Capitalismo | Capitalismo |

Fuente: Elaboración propia.

musical y el artista favorito de cada individuo ha interpretado al menos una canción de esos dos géneros, 0 de lo contrario.

Los pesos son $p_{Género_favorito} = 0.7$ y $p_{Artista_favorito} = 0.3$. Con estos valores se obtienen los valores de la tabla 2. Por ejemplo, $(v_{12})_{Música} = 0.7*0+0.3*0.3=0.09$.

Tabla 2. Afinidad de acuerdo con el factor música para los cuatro individuos.

| | 1 | 2 | 3 | 4 |
|---|------|------|------|---|
| 1 | 0 | 0.09 | 0.85 | 0 |
| 2 | 0.09 | 0 | 0.09 | 0 |
| 3 | 0.85 | 0.09 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 |

Fuente: Elaboración propia.

1.1.3 Reglas para el factor ubicación geográfica

Se consideran tres reglas ($z = 3$) para este factor: 1. Ciudad, 2. Zona y 3. Población:

- $(r_{ij})_{Ciudad} = 1$ si las dos personas viven en la misma ciudad, 0.7 si viven a una distancia menor a 30 km la una de la otra, 0.4 si viven a una distancia menor a 70 km la una de otra, 0 de lo contrario.
- $(r_{ij})_{Zona} = 1$ si las dos personas viven en zona urbana, 0.5 si viven en zona rural, 0 de lo contrario.
- $(r_{ij})_{Población} = 1$ si las dos personas viven en la misma ciudad y la población es pequeña (menos de 100 mil habitantes), 0.5 si viven en la misma ciudad y la población es media (entre 100 mil y un millón de habitantes), 0.2 si viven en la misma ciudad y la población es grande (más de un millón de habitantes), 0 de lo contrario.

Los pesos son $P_{Ciudad} = 0.5$, $P_{Zona} = 0.3$ y $P_{Población} = 0.2$. Con estos valores se obtienen los valores de la tabla 3.

1.1.4 Reglas para el factor tendencia política

Se consideran dos reglas ($z = 2$) para este factor: 1. Partido y 2. Ideología:

Tabla 3. Afinidad de acuerdo con el factor ubicación geográfica para los cuatro individuos.

| | 1 | 2 | 3 | 4 |
|----------|----------|----------|----------|----------|
| 1 | 0 | 0.2 | 0.6 | 0.15 |
| 2 | 0.2 | 0 | 0.6 | 0.15 |
| 3 | 0.6 | 0.6 | 0 | 0 |
| 4 | 0.15 | 0.15 | 0 | 0 |

Fuente: Elaboración propia.

- $(r_{ij})_{\text{Partido}} = 1$ si el partido político es el mismo, 0 de lo contrario.
- $(r_{ij})_{\text{Ideología}} = 1$ si la ideología es la misma, 0 de lo contrario.

Los pesos son $P_{\text{Partido}} = 0.8$ y $P_{\text{Ideología}} = 0.2$. Con estos valores se obtienen los valores de la tabla 4.

Finalmente, para calcular la matriz de afinidad A se consideran los siguientes pesos para cada factor: $W_{\text{Música}} = 0.2$, $W_{\text{Ubicación_geográfica}} = 0.4$ y $W_{\text{Tendencia_política}} = 0.4$. Con estos valores se obtienen los valores de la Tabla 5. Por ejemplo, $a_{12} = 0.2 * 0.09 + 0.4 * 0.2 + 0.2 + 0.4 * 0 = 0.098$.

Tabla 5. Matriz de afinidad A para los cuatro individuos.

| | 1 | 2 | 3 | 4 |
|----------|----------|----------|----------|----------|
| 1 | 0 | 0.098 | 0.49 | 0.46 |
| 2 | 0.098 | 0 | 0.298 | 0.1 |
| 3 | 0.49 | 0.298 | 0 | 0.08 |
| 4 | 0.46 | 0.1 | 0.08 | 0 |

Fuente: Elaboración propia.

Nótese que el modelo es flexible, ya que permite que el analista defina el número de

Tabla 4. Afinidad de acuerdo con el factor tendencia política para los cuatro individuos.

| | 1 | 2 | 3 | 4 |
|----------|----------|----------|----------|----------|
| 1 | 0 | 0 | 0.3 | 1 |
| 2 | 0 | 0 | 0.8 | 0 |
| 3 | 0.3 | 0.8 | 0 | 0.3 |
| 4 | 1 | 0 | 0.3 | 0 |

Fuente: Elaboración propia.

factores, el número de reglas de cada factor y los valores de los pesos de cada factor y de cada regla.

1.2 EXTENSIÓN DEL MODELO PARA EL CÁLCULO DE AFINIDAD PARA MÁS DE DOS INDIVIDUOS

Para calcular la afinidad, e.g., de un conjunto de cuatro individuos $G = \{1,2,3,4\}$, se considera el nivel de afinidad de cada pareja (compuesta por individuos distintos) que se puede formar a partir del conjunto G . Se suman las afinidades de todas las parejas y se divide entre el número total de estas (seis en este ejemplo).

Sea P el conjunto de todas las parejas (i, j) , $i \neq j$, de individuos que se pueden formar a partir del conjunto G . La cardinalidad de P es $\text{Card}(P) = \text{Combinaciones}(2, n)$, i.e., combinaciones de 2 en n , donde $n = \text{Card}(G)$. Para el ejemplo, $P = \{(1,2), (1,3), (1,4), (2,3), (2,4), (3,4)\}$, $\text{Card}(P) = \text{Combinaciones}(2, 4) = 6$.

Sea Afinidad una función con signatura: $P \rightarrow [0;1]$, i.e., la función recibe una pareja (i, j) y devuelve su valor de afinidad (un número

entre 0 y 1). El valor de afinidad de una pareja (i, j) se calcula de acuerdo con la ecuación (1). Entonces el valor de afinidad del conjunto G se calcula mediante la ecuación (5).

$$\forall p \in P, \sum \frac{Afinidad(p)}{Card(p)} \quad (5)$$

Con el fin de comparar el modelo propuesto con otras alternativas, se presenta a continuación una opción basada en *clustering* (agrupamiento). Considérese la matriz de afinidad A de la figura 1. Esta matriz se puede representar como un grafo no dirigido ponderado donde los vértices representan los individuos y a_{ij} (grado de afinidad) corresponde al peso asociado con la arista que conecta a los vértices i y j . Para $n = 4$ el grafo se muestra en la figura 2.

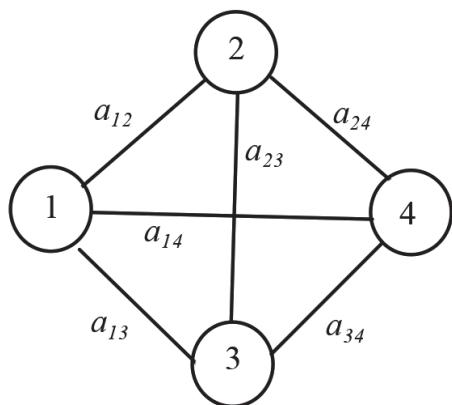


Figura 2. Grafo de afinidad para $n = 4$.

Fuente: Elaboración propia.

Con el fin de aplicar un método de *clustering* se presenta a continuación la forma general del algoritmo de *clustering* jerárquico [15].

Algoritmo de clustering jerárquico (forma general)

WHILE condición de fin de generación de *clusters* no se satisfaga **DO**

- Seleccionar los dos "mejores" *clusters* a juntar;
- Formar un nuevo *cluster* con los dos *clusters* seleccionados;

END WHILE;

Para cada problema particular se debe definir:
a) la condición de fin de generación de *clusters*, y b) el método para seleccionar los dos "mejores" *clusters* a juntar. La condición de fin puede ser, e.g., terminar cuando se haya obtenido un número específico de *clusters* o cuando el próximo *cluster* a formarse es inadecuado (e.g., la distancia promedio entre sus miembros supera un umbral establecido por el analista). En cuanto al método para seleccionar los dos mejores *clusters* a juntar existen muchas variantes, e.g., seleccionar la pareja de *clusters* que tienen la menor distancia entre sus centroides o la pareja de *clusters* que genera el *cluster* con el menor radio (el radio de un *cluster* es la máxima distancia entre todos sus miembros y el centroide del *cluster*); si se presentan empates se decide arbitrariamente.

Para nuestro caso de estudio, el objetivo es agrupar los individuos según su grado de afinidad. Es decir, mientras el grado de afinidad a_{ij} entre dos individuos i y j sea mayor (tiende a 1), se considera que los individuos son más cercanos. Por tanto, para el algoritmo de *clustering* se tomará $(1 - a_{ij})$ como la distancia entre dos individuos. Por otro lado, el algoritmo terminará (condición de fin) cuando se haya obtenido un número específico de *clusters*; ya que para nuestro primer experimento (ver segunda sección) se solicitó a los estudiantes

de los cursos examinados que conformaran grupos de entre dos y tres integrantes, lo cual determinará el número de *clusters* y con los cuales se comparará según los grupos (*clusters*) generados por el algoritmo de *clustering*.

El otro aspecto a definir es el método para seleccionar los dos “mejores” *clusters* a juntar. Inicialmente se tienen n *clusters* (cada individuo forma un *cluster*). Supongáse que los individuos i y j tienen la menor distancia ($1 - a_{ij}$) entre todos los *clusters*. Por tanto, en la siguiente iteración habrá $n-1$ *clusters* con i y j conformando un nuevo *cluster*. El problema es definir el centroide del *cluster* conformado por i y j con el fin de establecer la distancia de este *cluster* hacia los demás en la siguiente iteración. Una solución es seleccionar el

centroide como el miembro del *cluster* que minimiza la suma de las distancias a los otros puntos del *cluster* [15] (si el *cluster* solo tiene dos miembros la elección es arbitraria). Otras propuestas para seleccionar el centroide se exploran en dicha referencia: el problema en esencia es seleccionar un miembro del *cluster* como su miembro representativo (en general, denominado *clustroid*) y con base en él establecer las distancias a los demás *clusters* para seleccionar la próxima pareja de *clusters* que se deben juntar. En la figura 3 se bosqueja el proceso para un grafo de cuatro individuos.

En la primera iteración, figura 1 b), los individuos que tienen la menor distancia (0.2) son el 1 y el 2; por tanto, se genera un *cluster* con estos dos nodos. Como la distancia de 1 a 2 es la

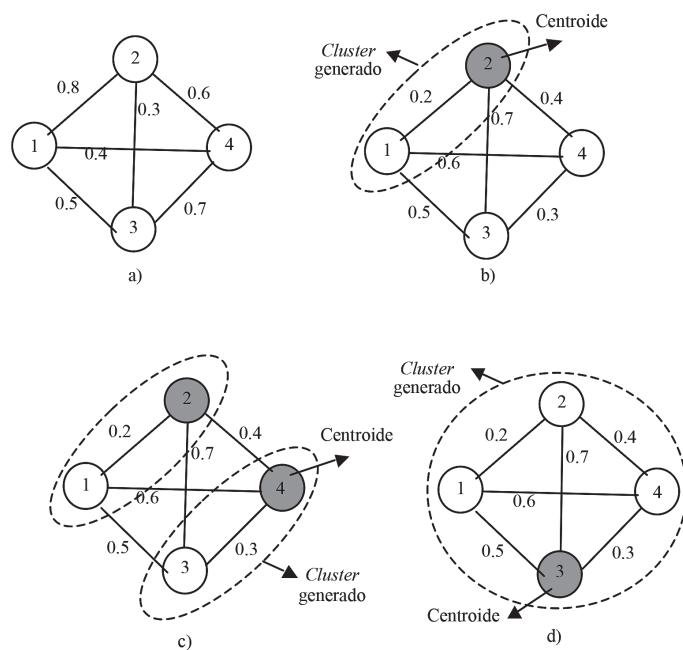


Figura 3. Aplicación del algoritmo de *clustering* por afinidad propuesto: a) grafo de afinidad, b) grafo de afinidad ($1 - a_{ij}$) y *cluster* generado en la primera iteración, c) *cluster* generado en la segunda iteración y d) *cluster* generado en la tercera iteración.

Fuente: Elaboración propia.

misma que de 2 a 1, se elige arbitrariamente cualquiera de los dos nodos como el centroide del *cluster* (acá se eligió el nodo 2). En la segunda iteración, figura 1 c), los individuos que tienen la menor distancia (0.3) son el 3 y el 4; por tanto, se genera un *cluster* con estos dos nodos y se elige arbitrariamente como centroide el nodo 3. Si se continúa el algoritmo, en la tercera iteración, figura 1 d), la distancia entre los dos *clusters* es 0.4 (distancia entre los centroides de los dos *clusters*); se genera un *cluster* con todos los nodos del grafo. Nótese que el centroide de este *cluster* no es arbitrario: como el *cluster* tiene más de dos miembros, entonces se encuentra que el miembro que minimiza la suma de las distancias a los otros miembros del *cluster* es el 3 ($0.5 + 0.7 + 0.3 = 1.5$).

Nótese que al aplicar un algoritmo de *clustering* con base en el grado de afinidad, tal y cómo se ha propuesto, puede ocurrir lo siguiente. Supongáse que

- en la primera iteración los individuos i y j tienen la menor distancia ($1 - a_{ij}$) entre todos los *clusters*. En la figura 3, los nodos 1 y 2.
- en la segunda iteración los individuos p y q tienen la menor distancia ($1 - a_{pq}$) entre todos los *clusters*. En la figura 3, los nodos 3 y 4.
- en la tercera iteración los *clusters* conformados por (i, j) y (p, q) tienen la menor distancia entre todos los *clusters*. Es decir, se forma un *cluster* (i, j, p, q) . En la figura 3, los nodos 1, 2, 3 y 4.

De esta forma, el *cluster* resultante está conformado por dos *subclusters* cada uno con alta afinidad; sin embargo, esto no garantiza que la afinidad del *cluster* resultante sea alta. Por ejemplo, i y j podrían ser altamente afines porque viven en zona urbana, mientras que p y q podrían ser altamente afines porque

viven en zona rural. Por otro lado, estos cuatro individuos podrían ser afines en otras características y el alto nivel de afinidad de las parejas puede contribuir a un nivel de afinidad entre los cuatro individuos, i.e., afinidad entre los más afines así haya aspectos en los que puede haber posiciones opuestas entre los *subclusters*, lo que incluso puede llegar a ser beneficioso para el desempeño del grupo [16]. Aunque esto también se manifiesta en nuestro método, como se evidencia en la ecuación (5): allí se calcula la afinidad de un grupo con base en el promedio de las afinidades de todas las parejas que lo conforman.

Otra diferencia con nuestro método es que el algoritmo de *clustering* no garantiza la generación de *clusters* donde cada uno tiene un mínimo y un máximo número de miembros. De esta forma, si se requieren, e.g., 18 grupos cada uno con un mínimo de dos y un máximo de tres miembros, el algoritmo de *clustering* generará 18 *clusters* pero algunos podrían tener, e.g., uno, dos, tres, cuatro o más miembros.

2. EXPERIMENTOS

Para los experimentos se consideraron dos grupos de estudiantes correspondientes a dos cursos universitarios de ingeniería de sistemas de los últimos semestres. El primero de 60 estudiantes y el segundo de 50 estudiantes. Se consideraron 4 (m) factores para evaluar su afinidad: *i*) el aspecto técnico, *ii*) el aspecto académico, *iii*) las cualidades grupales, y *iv*) las cualidades individuales. Para los cuatro factores se usó el mismo peso (0.25). Para la obtención de los datos de los estudiantes se aplicó la encuesta que se muestra en la tabla 6; allí también se muestran las reglas que se consideraron (E1 y E2 representan estudiantes).

Tabla 6. Reglas para ejemplo propuesto.

| Pregunta | Opciones de respuesta | Reglas |
|---|---|--|
| En la creación de software, ¿cuál es su etapa preferida o en la que considera que tiene fortalezas? | 1. Análisis y definición 2. Diseño 3. Construcción 4. Pruebas | <ul style="list-style-type: none"> ▪ Si $E_1 = E_2 \rightarrow r_{ij} = 1$. (Es decir, si la elección de E_1 es la misma que E_2). ▪ Si $(E_1: a \wedge E_2: b) \vee (E_1: a \wedge E_2: b) \rightarrow r_{ij} = 1$. ▪ Si $(E_1: b \wedge E_2: c) \vee (E_1: c \wedge E_2: b) \rightarrow r_{ij} = 0.7$. ▪ Si $(E_1: c \wedge E_2: d) \vee (E_1: d \wedge E_2: c) \rightarrow r_{ij} = 0.5$. ▪ Si $(E_1: b \wedge E_2: d) \vee (E_1: b \wedge E_2: d) \rightarrow r_{ij} = 0.3$. ▪ Si $(E_1: d \wedge E_2: a) \vee (E_1: a \wedge E_2: d) \rightarrow r_{ij} = 0.1$. ▪ En otro caso, $r_{ij} = 0$. |
| ¿Qué tipo de software prefiere usar para sus desarrollos? | 1. Libre 2. Licenciado | <ul style="list-style-type: none"> ▪ Si $(E_1: a \wedge E_2: a) \rightarrow r_{ij} = 1$. ▪ Si $(E_1: b \wedge E_2: b) \rightarrow r_{ij} = 0.5$. ▪ En otro caso, $r_{ij} = 0$. |
| ¿Está de acuerdo que en el curso se exija la herramienta a usar? | 1. Sí 2. No 3. Le es indiferente | <ul style="list-style-type: none"> ▪ Si $(E_1: a \wedge E_2: a) \vee (E_1: b \wedge E_2: b) \rightarrow r_{ij} = 0.6$. ▪ Si $(E_1: c \wedge E_2: c) \rightarrow r_{ij} = 1$. ▪ Si $(E_1: a \wedge E_2: c) \vee (E_1: b \wedge E_2: c) \vee (E_1: c \wedge E_2: d) \vee (E_1: d \wedge E_2: c) \rightarrow r_{ij} = 0.5$. ▪ En otro caso, $r_{ij} = 0$. |
| ¿Qué metodología de desarrollo piensa que es más efectiva? | 1. La metodología ágil que logra un desarrollo en menos tiempo. 2. Las metodologías monumentales porque se identifica claramente el problema y se documenta de mejor manera. 3. Piensa que se debe analizar cada caso y elegir la metodología más adecuada. 4. Prefiero no usar ninguna metodología y seguir mis propios estándares. | <ul style="list-style-type: none"> ▪ Si $E_1 = E_2 \rightarrow r_{ij} = 1$. ▪ Si $(E_1: a \wedge E_2: b) \vee (E_1: b \wedge E_2: a) \rightarrow r_{ij} = 0.4$. ▪ Si $(E_1: c \wedge E_2: a) \vee (E_1: a \wedge E_2: c) \vee (E_1: c \wedge E_2: b) \vee (E_1: b \wedge E_2: c) \vee (E_1: c \wedge E_2: d) \vee (E_1: d \wedge E_2: c) \rightarrow r_{ij} = 0.6$. ▪ En otro caso, $r_{ij} = 0$. |
| ¿Qué desempeño académico considera que tiene? | 1. Medio 2. Alto 3. Muy alto 4. Aceptable | <ul style="list-style-type: none"> ▪ Si $E_1 = E_2 \rightarrow r_{ij} = 1$. ▪ Si $(E_1: c \wedge E_2: d) \vee (E_1: d \wedge E_2: c) \rightarrow r_{ij} = 0.3$. ▪ Si $(E_1: c \wedge E_2: a) \vee (E_1: a \wedge E_2: c) \rightarrow r_{ij} = 0.5$. ▪ Si $(E_1: c \wedge E_2: b) \vee (E_1: b \wedge E_2: c) \rightarrow r_{ij} = 0.7$. ▪ En otro caso, $r_{ij} = 0$. |

| | | |
|--|---|---|
| <p>¿Tengo unos métodos de estudio claramente identificados?</p> | <p>1. Sí 2. No</p> | <ul style="list-style-type: none"> ▪ Si ($E1: a \wedge E2: a \rightarrow r_{ij} = 0.7$). ▪ Si ($E1: a \wedge E2: b \rightarrow r_{ij} = 0.2$). ▪ Si ($E1: a \wedge E2: b \vee (E1: b \wedge E2: a) \rightarrow r_{ij} = 0.1$). ▪ En otro caso, $r_{ij} = 0$. |
| <p>¿Para elegir el grupo considera el desempeño de sus compañeros?</p> | <p>1. Es muy importante. 2. Podría no ser tan importante si él/ella tiene interés en el tema. 3. No es importante.</p> | <ul style="list-style-type: none"> ▪ Si $E1 = E2 \rightarrow r_{ij} = 1$. ▪ Si ($E1: a \wedge E2: c \vee (E1: c \wedge E2: a) \rightarrow r_{ij} = 0.2$). ▪ Si ($E1: a \wedge E2: b \vee (E1: b \wedge E2: a) \rightarrow r_{ij} = 0.5$). ▪ Si ($E1: b \wedge E2: c \vee (E1: c \wedge E2: b) \rightarrow r_{ij} = 0.7$). ▪ En otro caso, $r_{ij} = 0$. |
| <p>¿Ha trabajado en grupos o proyectos de investigación?</p> | <p>1. Sí 2. No</p> | <ul style="list-style-type: none"> ▪ Si ($E1: a \wedge E2: a \rightarrow r_{ij} = 1$). ▪ Si ($E1: a \wedge E2: b \rightarrow r_{ij} = 0.5$). ▪ Si ($E1: a \wedge E2: b \vee (E1: b \wedge E2: a) \rightarrow r_{ij} = 0.4$). ▪ En otro caso, $r_{ij} = 0$. |
| <p>¿Qué calidad considera puede aportarle más al grupo de trabajo?</p> | <p>1. Liderazgo 2. Experiencia 3. Capacidad técnica 4. Inteligencia</p> | <ul style="list-style-type: none"> ▪ Si $E1 = E2 \rightarrow r_{ij} = 1$. ▪ Si ($E1: a \wedge E2: b \vee (E1: b \wedge E2: a) \rightarrow r_{ij} = 1$). ▪ Si ($E1: c \wedge E2: a \vee (E1: a \wedge E2: c) \rightarrow r_{ij} = 0.6$). ▪ Si ($E1: b \wedge E2: d \vee (E1: d \wedge E2: b) \rightarrow r_{ij} = 0.6$). ▪ Si ($E1: d \wedge E2: a \vee (E1: a \wedge E2: d) \rightarrow r_{ij} = 0.3$). ▪ En otro caso, $r_{ij} = 0$. |
| <p>A la hora de tomar una decisión en su equipo de trabajo usted:</p> | <p>1. Prefiere siempre seguir instrucciones. 2. Le gusta dar ideas pero que otro tome la decisión. 3. Le gusta dar ideas y participar en la toma de la decisión. 4. Toma la decisión considerando los aportes de los compañeros. 5. Toma la decisión sin considerar a los compañeros.</p> | <ul style="list-style-type: none"> ▪ Si $E1 = E2 \rightarrow r_{ij} = 1$. ▪ Si ($E1: a \wedge E2: d \vee (E1: d \wedge E2: a) \rightarrow r_{ij} = 1$). ▪ Si ($E1: a \wedge E2: e \vee (E1: e \wedge E2: a) \rightarrow r_{ij} = 0.6$). ▪ Si ($E1: c \wedge E2: d \vee (E1: d \wedge E2: c) \rightarrow r_{ij} = 0.5$). ▪ Si ($E1: d \wedge E2: b \vee (E1: b \wedge E2: d) \rightarrow r_{ij} = 0.3$). ▪ En otro caso, $r_{ij} = 0$. |
| <p>Bajo ambientes de presión cuál es su actitud:</p> | <p>1. Calmado y piensa en acciones que lo lleven a resolver la dificultad. 2. Se desespera y no logra pensar en una solución. 3. Lo deja pasar y no presta atención, espera las consecuencias.</p> | <ul style="list-style-type: none"> ▪ Si $E1 = E2 \rightarrow r_{ij} = 1$. ▪ Si ($E1: a \wedge E2: c \vee (E1: c \wedge E2: a) \rightarrow r_{ij} = 0.6$). ▪ Si ($E1: b \wedge E2: c \vee (E1: c \wedge E2: b) \rightarrow r_{ij} = 0.4$). ▪ Si ($E1: a \wedge E2: b \vee (E1: b \wedge E2: a) \rightarrow r_{ij} = 0.2$). ▪ En otro caso, $r_{ij} = 0$. |



| | | |
|--|---|--|
| Si usted se equivoca siente que: | 1. Esto es inevitable por su condición de ser humano. 2. Revisa el caso y saca lecciones de lo que le sucedió. 3. Cree que esto le sucede pocas veces y por eso no le presta mucha atención. 4. Corrige los errores y no se preocupa en analizarlos. | <ul style="list-style-type: none"> ▪ Si $E1 = E2 \rightarrow r_{ij} = 1$. ▪ Si $(E1: b \wedge E2: d) \vee (E1: d \wedge E2: b) \rightarrow r_{ij} = 0.5$. ▪ Si $(E1: a \wedge E2: c) \vee (E1: c \wedge E2: a) \rightarrow r_{ij} = 0.4$. ▪ Si $(E1: a \wedge E2: d) \vee (E1: d \wedge E2: a) \rightarrow r_{ij} = 0.2$. ▪ En otro caso, $r_{ij} = 0$. |
| Cuál es su género musical favorito: | 1. Pop 2. Rock 3. Tropical 4. Otro | <ul style="list-style-type: none"> ▪ Si $E1 = E2 \rightarrow r_{ij} = 1$. ▪ Si $(E1: a \wedge E2: b) \vee (E1: b \wedge E2: a) \rightarrow r_{ij} = 0.6$. ▪ Si $(E1: b \wedge E2: c) \vee (E1: c \wedge E2: b) \rightarrow r_{ij} = 0.5$. ▪ Si $(E1: b \wedge E2: d) \vee (E1: d \wedge E2: b) \rightarrow r_{ij} = 0.2$. ▪ En otro caso, $r_{ij} = 0$. |
| Cuál de estas respuestas lo identifican más en sus creencias religiosas: | 1. Soy una persona espiritual pero no estoy de acuerdo con la religión. 2. Respeto las posiciones religiosas de las personas. 3. Aunque respeto las diferentes posiciones religiosas, creo que mis creencias deberían ser adoptadas por todas las personas que me rodean. | <ul style="list-style-type: none"> ▪ Si $E1 = E2 \rightarrow r_{ij} = 1$. ▪ Si $(E1: a \wedge E2: b) \vee (E1: b \wedge E2: a) \rightarrow r_{ij} = 0.5$. ▪ Si $(E1: b \wedge E2: c) \vee (E1: c \wedge E2: b) \rightarrow r_{ij} = 0.3$. ▪ Si $(E1: a \wedge E2: c) \vee (E1: c \wedge E2: a) \rightarrow r_{ij} = 0.2$. ▪ En otro caso, $r_{ij} = 0$. |
| ¿Considera que se enoja con facilidad? | 1. Sí. 2. Me enojo frecuentemente. 3. Depende la situación. 4. No. 5. Es difícil que pierda la calma. | <ul style="list-style-type: none"> ▪ Si $E1 = E2 \rightarrow r_{ij} = 1$. ▪ Si $(E1: a \wedge E2: e) \vee (E1: e \wedge E2: a) \rightarrow r_{ij} = 0.6$. ▪ Si $(E1: e \wedge E2: c) \vee (E1: c \wedge E2: e) \rightarrow r_{ij} = 0.35$. ▪ Si $(E1: a \wedge E2: c) \vee (E1: c \wedge E2: a) \rightarrow r_{ij} = 0.35$. ▪ Si $(E1: a \wedge E2: d) \vee (E1: d \wedge E2: a) \rightarrow r_{ij} = 0.35$. ▪ En otro caso, $r_{ij} = 0$. |
| ¿Cómo reacciona ante la crítica? | 1. La tomo positivamente, aunque me molesta. 2. La tomo positivamente y pienso que sí debo mejorar ese aspecto. 3. Creo que no me aporta nada. | <ul style="list-style-type: none"> ▪ Si $E1 = E2 \rightarrow r_{ij} = 1$. ▪ Si $E1: b = E2: b \rightarrow r_{ij} = 0.95$. ▪ Si $(E1: a \wedge E2: b) \vee (E1: b \wedge E2: a) \rightarrow r_{ij} = 0.75$. ▪ Si $(E1: b \wedge E2: c) \vee (E1: c \wedge E2: b) \rightarrow r_{ij} = 0.5$. ▪ En otro caso, $r_{ij} = 0$. |

Fuente: Elaboración propia.

En un primer experimento se consideraron los dos cursos. En ambos cursos, al principio del semestre académico se solicitó a los

estudiantes que conformaran grupos de entre dos y tres integrantes. Cada estudiante solo pertenece a un grupo, y cada uno elige

libremente a sus compañeros de grupo (ningún estudiante está en ambos cursos, de hecho el primer curso es requisito del segundo). El total de grupos fue de 22 en el primer curso y 19 en el segundo. Al final del semestre se comparó la afinidad de cada grupo (calculada mediante la ecuación 5) con la calificación grupal promedio obtenida en el curso (calificación obtenida mediante evaluaciones grupales).

Con el fin de comparar la propuesta con la afinidad percibida por los integrantes de cada grupo, al final del semestre también se solicitó a cada integrante de cada grupo que calificase, según su experiencia vivida durante el curso, el grado de afinidad (entre 0

y 1) del grupo en cada uno de los 4 factores. Los resultados se promediaron y se obtuvo un nivel de afinidad percibido. En las tablas 7 y 8 se presentan los resultados obtenidos para 10 grupos de cada curso. Se presentan los 5 grupos con calificación más alta y los 5 con calificación más baja. Por otro lado, en la tabla 9 se presentan los resultados del algoritmo de *clustering* (los seis *clusters* con afinidad promedio más alta: la afinidad de cada *cluster* se calculó promediando las afinidades entre sus miembros). Con el fin de facilitar las comparaciones, la condición de fin de generación de *clusters* fue el número de grupos formado por los estudiantes en cada curso (22 en el primero y 19 en el segundo).

Tabla 7. Primer Curso: calificación por grupo vs. afinidad por grupo.

| Grupo | Integrantes | Calificación grupal final obtenida en el curso | Afinidad (Ecuación (5)) | Afinidad percibida por los miembros del grupo |
|--------------------------------------|-------------|--|-------------------------|---|
| Grupos con mayor calificación grupal | | | | |
| 15 | (37,38,39) | 4.753 | 0.776 | 0.8 |
| 3 | (7,8,9) | 4.706 | 0.562 | 0.8 |
| 8 | (19,20,21) | 4.619 | 0.513 | 0.75 |
| 6 | (15,16) | 4.510 | 0.502 | 0.75 |
| 18 | (46,47,48) | 4.373 | 0.547 | 0.75 |
| Grupos con menor calificación grupal | | | | |
| 16 | (40,41,42) | 3.426 | 0.438 | 0.5 |
| 7 | (17,18) | 3.559 | 0.416 | 0.5 |
| 11 | (26,27,28) | 3.63 | 0.493 | 0.5 |
| 2 | (4,5,6) | 3.63 | 0.48 | 0.6 |
| 1 | (1,2) | 3.6 | 0.495 | 0.5 |

Fuente: Elaboración propia.

Tabla 8. Segundo Curso: calificación por grupo vs. afinidad por grupo.

| Grupo | Integrantes | Calificación grupal final obtenida en el curso | Afinidad (Ecuación (5)) | Afinidad percibida por los miembros del grupo |
|--------------------------------------|-------------|--|-------------------------|---|
| Grupos con mayor calificación grupal | | | | |
| 8 | (20,21) | 4.5 | 0.606 | 0.75 |
| 10 | (24,25) | 4.5 | 0.595 | 0.6 |
| 9 | (22,23) | 4.3 | 0.513 | 0.75 |
| 4 | (9,10) | 4.3 | 0.498 | 0.5 |
| 5 | (11,12,13) | 4.1 | 0.518 | 0.5 |
| Grupos con menor calificación grupal | | | | |
| 18 | (35,36) | 3.3 | 0.575 | 0.6 |
| 19 | (45,46,47) | 3.2 | 0.481 | 0.5 |
| 14 | (48,49,50) | 3.2 | 0.498 | 0.4 |
| 12 | (29,30,31) | 3.0 | 0.498 | 0.5 |
| 16 | (39,40,41) | 2.9 | 0.481 | 0.5 |

Fuente: Elaboración propia.

Tabla 9. *Clusters* generados en las seis primeras iteraciones del algoritmo de *clustering* en cada curso.

| Iteración | Clusters primer curso | Clusters segundo curso |
|-----------|-----------------------|------------------------|
| 1 | (37,38,9) | (3,15,27,29) |
| 2 | (7,8,36) | (2,9,16,20) |
| 3 | (20,16,46) | (10,12,41) |
| 4 | (17,41,45) | (24,40,48) |
| 5 | (2,12,48) | (7,11,22) |
| 6 | (8,33,40) | (36,50) |

Fuente: Elaboración propia.

Aunque los resultados no son concluyentes, sí es al menos sugerente que los grupos con

mayor afinidad (según la ecuación 5) son los que obtuvieron en general las mayores calificaciones grupales promedio. De hecho, en el primer curso los cinco grupos (y en el segundo curso los dos grupos) con mayor calificación grupal fueron precisamente los grupos con mayor afinidad. Esto sugiere que la afinidad de los individuos puede contribuir a un mejor desempeño grupal. En cuanto a la afinidad percibida, también es significativo que los grupos que se autocalificaron como los más afines, tendieron a obtener las calificaciones más altas mientras que los que se autocalificaron como los menos afines, efectivamente tendieron a obtener las calificaciones más bajas. Por otro lado, la afinidad percibida tiende a ser más alta que la

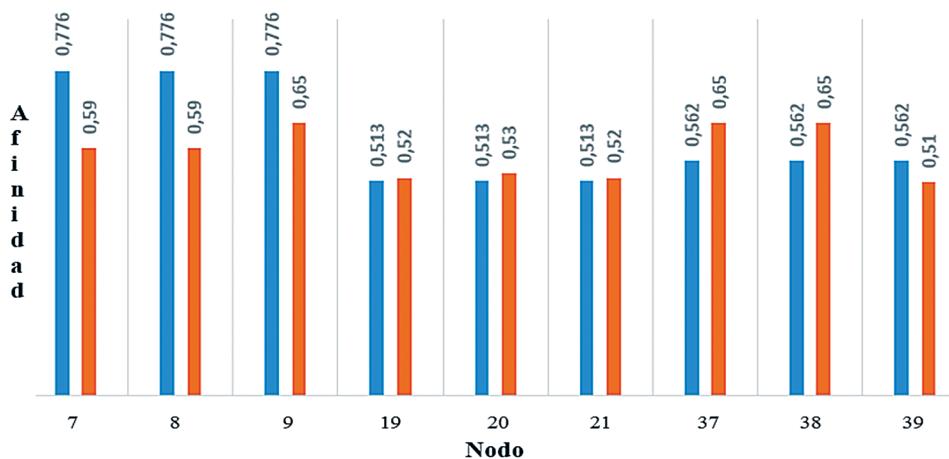
afinidad calculada por el método propuesto. Esto se puede deber a que los individuos de cada grupo se conocían de otros cursos, i.e., ya habían establecido lazos de amistad lo que contribuyó a que tuviesen una apreciación alta con respecto a su afinidad grupal.

En cuanto a los resultados generados por el algoritmo de *clustering* e.g., para el primer curso en su *cluster* con afinidad promedio más alta (37, 38, 9) se observa que todos sus miembros se encuentran entre los cinco primeros grupos de mayor afinidad obtenidos con el método propuesto (ecuación 5). Esto sugiere al menos que los dos métodos identifican ciertos individuos que contribuyen en los grupos más afines. En la figura 4 se compara la afinidad (calculada con la ecuación 5) de los miembros de los tres grupos con mayor calificación grupal vs. la afinidad de estos miembros según la afinidad obtenida mediante el *clustering*. Por otro lado, en el segundo curso, en su *cluster* con mayor afinidad promedio (3, 15, 27, 29) se observa

que *ninguno de sus miembros se encuentra entre los cinco primeros grupos de mayor afinidad obtenidos con el método propuesto*. Esto es sugerente y pone de manifiesto la diferencia de los dos métodos. Esto se explica porque la pareja (3, 27) es la de mayor afinidad en el segundo curso (véase la tabla 11), por ello el algoritmo de *clustering* la incluye en un *cluster* en la primera iteración y en posteriores iteraciones la une con los miembros 3 y 29. En conclusión, los métodos generan resultados distintos y no se garantiza que los individuos incluidos en los grupos más afines de uno también figuren como miembros de los grupos más afines del otro método.

En un segundo experimento se calculó la afinidad para todos los grupos de tres estudiantes distintos que se pueden formar a partir del grupo de 50 estudiantes del segundo curso. En la tabla 10 se presentan los diez grupos con mayor afinidad. Nótese que los estudiantes con códigos 15 y 16 aparecen en ocho de estos diez grupos. Esto sugiere que

Figura 4. Afinidad (según el grupo al que pertenecen) de algunos estudiantes del curso 1: calculada con la Ecuación (5) (en azul) y obtenida mediante el *clustering* (en naranja).



Fuente: Elaboración propia.

son individuos con un alto nivel de afinidad (de hecho, el grupo con mayor afinidad incluye a estos dos estudiantes). Sin embargo, su afinidad como pareja no está entre las diez parejas con mayor afinidad (véase tabla 11). Por otro lado, la pareja conformada por los estudiantes 3 y 27 que posee la mayor afinidad (véase tabla 11), no aparece en ninguno de los diez grupos más afines de tres estudiantes de la tabla 10.

Tabla 10. Grupos de tres estudiantes con mayor afinidad.

| Posición | Grupo | Afinidad |
|----------|------------|----------|
| 1 | (15,16,29) | 0.700 |
| 2 | (7,15,16) | 0.696 |
| 3 | (11,15,20) | 0.688 |
| 4 | (15,20,29) | 0.687 |
| 5 | (11,15,16) | 0.685 |
| 6 | (7,16,29) | 0.683 |
| 7 | (15,16,47) | 0.678 |
| 8 | (15,16,20) | 0.676 |
| 9 | (9,15,16) | 0.675 |
| 10 | (7,16,25) | 0.666 |

Fuente: Elaboración propia.

También se calculó la afinidad para todos los grupos de cuatro estudiantes distintos que se pueden formar a partir del grupo de 50 estudiantes. En la tabla 12 se presentan los diez grupos con mayor afinidad. Similarmente al caso anterior (grupos de tres estudiantes), se observó que los estudiantes 15 y 16 aparecen en la mayoría de estos grupos (de hecho, el estudiante 15 aparece en todos los grupos y el 16 en ocho de los diez grupos).

Tabla 11. Parejas con la mayor afinidad y con la menor afinidad.

| Pareja | Afinidad |
|---------|----------|
| (3, 27) | 0.6935 |
| (3, 15) | 0.6395 |
| (3, 12) | 0.626 |
| (2, 16) | 0.6165 |
| (3, 29) | 0.6165 |
| (3, 19) | 0.6155 |
| (3, 25) | 0.6155 |
| (3, 10) | 0.6105 |
| (2, 20) | 0.6005 |
| (2, 9) | 0.596 |
| (1,21) | 0.2085 |
| (21,39) | 0.2265 |
| (23,42) | 0.2285 |
| (23,44) | 0.2285 |
| (1,43) | 0.2285 |
| (21,37) | 0.2295 |
| (31,42) | 0.2360 |
| (1,44) | 0.2385 |
| (1,45) | 0.2385 |
| (1,14) | 0.2395 |

Fuente: Elaboración propia.

Tabla 12. Grupos de cuatro estudiantes con mayor afinidad.

| Posición | Grupo | Afinidad |
|----------|---------------|----------|
| 1 | (7,15,16,29) | 0.678 |
| 2 | (15,16,20,29) | 0.673 |
| 3 | (11,15,16,20) | 0.668 |
| 4 | (7,15,16,20) | 0.666 |
| 5 | (7,11,15,16) | 0.664 |
| 6 | (7,15,16,47) | 0.662 |
| 7 | (7,15,20,29) | 0.661 |
| 8 | (15,16,29,47) | 0.660 |
| 9 | (11,15,20,29) | 0.659 |
| 10 | (9,15,16,29) | 0.658 |

Fuente: Elaboración propia.

Para los grupos de cinco estudiantes, ver la tabla 13, los resultados mostraron el mismo comportamiento (de nuevo los estudiantes 15

y 16 aparecen en la mayoría de los grupos). Nótese, además, que los valores de afinidad de la tabla 13 son menores que los de las tablas 11 y 12; esto es razonable, ya que a medida que los grupos tienen más integrantes, *existirán más diferencias entre ellos*.

Tabla 13. Grupos de cinco estudiantes con mayor afinidad.

| Posición | Grupo | Afinidad |
|----------|-----------------|----------|
| 1 | (7,9,15,16,29) | 0.6503 |
| 2 | (7,9,11,15,16) | 0.6399 |
| 3 | (7,9,15,20,29) | 0.6361 |
| 4 | (7,9,15,16,20) | 0.6358 |
| 5 | (7,9,15,16,40) | 0.6335 |
| 6 | (3,7,15,16,29) | 0.6332 |
| 7 | (7,8,15,16,20) | 0.6319 |
| 8 | (7,9,15,16,25) | 0.6316 |
| 9 | (7,9,11,15,20) | 0.6302 |
| 10 | (6,11,15,16,20) | 0.6301 |

Fuente: Elaboración propia.

Tabla 14. Grupos de tres estudiantes con menor afinidad.

| Posición | Grupo | Afinidad |
|----------|-----------|----------|
| 1 | (1,12,43) | 0.347 |
| 2 | (1,12,44) | 0.369 |
| 3 | (1,12,45) | 0.372 |
| 4 | (1,12,46) | 0.382 |
| 5 | (1,12,47) | 0.405 |
| 6 | (1,12,48) | 0.405 |
| 7 | (1,12,49) | 0.411 |
| 8 | (1,12,50) | 0.428 |
| 9 | (1,13,14) | 0.449 |
| 10 | (1,13,15) | 0.491 |

Fuente: Elaboración propia.

Con respecto a los diez grupos con menor afinidad, se puede observar un comportamiento análogo al de los grupos más afines, e.g., el individuo 1 aparece en todos los grupos de 3, 4 y 5 estudiantes, ver tablas 14, 15 y 16.

Tabla 15. Grupos de cuatro estudiantes con menor afinidad.

| Posición | Grupo | Afinidad |
|----------|-----------------|----------|
| 1 | (1,21,23,31,42) | 0.2806 |
| 2 | (1,21,23,42,44) | 0.2829 |
| 3 | (1,21,23,34,42) | 0.2878 |
| 4 | (1,6,21,23,42) | 0.2912 |
| 5 | (1,23,31,42,44) | 0.2925 |
| 6 | (1,21,23,37,42) | 0.2948 |
| 7 | (1,23,42,43,44) | 0.2952 |
| 8 | (1,23,31,34,42) | 0.2957 |
| 9 | (1,19,23,31,42) | 0.2959 |
| 10 | (1,23,34,42,44) | 0.2964 |

Fuente: Elaboración propia.

Tabla 16. Grupos de cinco estudiantes con menor afinidad.

| Posición | Grupo | Afinidad |
|----------|-----------------|----------|
| 1 | (1,21,23,31,42) | 0.2806 |
| 2 | (1,21,23,42,44) | 0.2829 |
| 3 | (1,21,23,34,42) | 0.2878 |
| 4 | (1,6,21,23,42) | 0.2912 |
| 5 | (1,23,31,42,44) | 0.2925 |
| 6 | (1,21,23,37,42) | 0.2948 |
| 7 | (1,23,42,43,44) | 0.2952 |
| 8 | (1,23,31,34,42) | 0.2957 |
| 9 | (1,19,23,31,42) | 0.2959 |
| 10 | (1,23,34,42,44) | 0.2964 |

Fuente: Elaboración propia.

Estos mismos experimentos se hicieron con el grupo de estudiantes del primer curso y se obtuvieron resultados similares.

3. CONCLUSIONES Y TRABAJO FUTURO

En este artículo se propuso un modelo matemático para cuantificar el grado de afinidad entre los individuos de un grupo humano. Inicialmente, el modelo cuantifica el grado de afinidad entre dos individuos y luego se extiende para calcularlo para más de dos individuos. Para obtener el grado de afinidad, el modelo considera factores y reglas para cada factor. Una de las ventajas del modelo es que el analista puede especificar el número de factores y asociar con cada factor el número de reglas que desee. Además, puede especificar los pesos para cada factor y para cada regla de acuerdo con sus intereses. Esto hace que el modelo sea flexible y que se pueda aplicar prácticamente a cualquier grupo humano: estudiantes, trabajadores, miembros de una red social, etc.

Los experimentos se hicieron con dos cursos de estudiantes universitarios, uno de 60 estudiantes y el otro de 50. Los resultados indicaron que los grupos con mayor grado de afinidad fueron los de mayor calificación promedio grupal. En otro de los experimentos se analizó el grado de afinidad de todos los grupos que se pueden formar con 3, 4 y 5 estudiantes a partir del curso de 50 estudiantes. Se observó que existen ciertos individuos que regularmente conforman los grupos más afines. Los resultados también mostraron que a medida que los grupos tienen más individuos, el grado de afinidad tiende a bajar con respecto a los grupos con menos individuos. Esto es razonable, ya que

mientras más grandes sean los grupos más diferencias se pueden encontrar entre los individuos.

Como trabajos futuros se tienen planeados los siguientes: i) automatizar la recolección de datos para los factores y reglas. En este trabajo se usó una encuesta (ver tabla 6); sin embargo, esta labor es tediosa para los usuarios y los analistas. La idea es desarrollar un aplicativo que recopile automáticamente los datos de los individuos correspondientes a los factores definidos por el analista. El aplicativo debería permitir que el analista especifique los factores, reglas y las fuentes de datos (e.g., una red social, una base de datos), ii) analizar cuáles son los rasgos que caracterizan a los individuos que suelen contribuir en mayor/menor medida con el grado de afinidad de los grupos y ver si es posible promover/aminorar estos rasgos en otros individuos, iii) determinar, con base en su grado de afinidad, cuál es el número apropiado de individuos para desarrollar una determinada tarea, iv) analizar aspectos de discriminación en grupos [17] y cómo maximizar la influencia de un grupo en una comunidad [18], v) automatizar la asignación de los valores para los r_{ij} , e.g., una asignación automática podría ser $r_{ij} = 1$ si los individuos comparten el mismo valor de una regla dada y $r_{ij} = 0$, de lo contrario. Sin embargo, estos valores se podrían refinar dependiendo de cada regla, de sus valores y de los objetivos de los analistas, y vi) extender medidas de similitud entre dos vértices, tales como las basadas en caminos aleatorios (*random walks*) [19] para obtener la afinidad de grupos y compararlo con el método aquí propuesto. Un trabajo análogo se puede hacer con grafos de afinidad y con otros algoritmos de *clustering*. Un punto de partida es el trabajo de Liu [20].

AGRADECIMIENTOS

Los autores desean expresar sus agradecimientos a los profesores Jesús Antonio Hernández y Jaime Guzmán Luna, de la Universidad Nacional, sede Medellín, por sus valiosos comentarios sobre el modelo propuesto. También al profesor Fernando Sánchez Lasheras, de la Universidad de Oviedo, por sus aportes en la parte de *clustering*.

BIBLIOGRAFÍA

- [1] Molina, F., Loyola, P., Velásquez, J. D. (2010). Generación de equipos de trabajo mediante análisis de redes sociales e identificación de atributos personales. En: Revista Ingeniería de Sistemas, Vol. 14 (1), pp. 103-122.
- [2] Panigrahy, R., Najork, M., Xie, Y. (2012). How User Behavior is Related to Social Affinity. En: 5th ACM International Conference on Web Search and Data Mining (WSDM).
- [3] Bapna, R., Gupta, A., Rice, S., Sundararajan, A. (2011). Trust, Reciprocity and the Strength of Social Ties: An Online Social Network based Field Experiment. En: Conference on Information Systems and Technology (CIST).
- [4] Chung-Chien, H., Medhin, N. (2006). Positive and Negative Affinities Model For Social Networks. En: Department of Mathematics, 3rd International Conference on Neural, Parallel and Scientific Computations.
- [5] Smith, M., Giraud-Carrier, C., Purser, N. (2009). Implicit affinity networks and social capital. En: Journal Information Technology and Management, Vol. (2-3), pp. 123-134.
- [6] Gil, J. (2007). La idónea agrupación de trabajadores según grados de compatibilidad psicológica. En: Asociación Española de Dirección y Economía de la Empresa (Aedem).
- [7] Eaglea, N., Pentlandb, A., Lazerc, D. (2009). Inferring friendship network structure by using mobile phone data MIT Media Laboratory. En: Proceedings of the National Academy of Sciences of the United States of America PNAS, Vol. 106 (36), pp. 15274-15278.
- [8] Zhang, H., Dantu, R. (2010). Predicting Social Ties in Mobile Phone Networks. En: IEEE International Conference on Intelligence and Security Informatics ISI.
- [9] Bisgaard, S., Kulahci, M. (2011). Time Series Analysis and Forecasting by Example. New Jersey. Wiley.
- [10] Kim, M., Park, S. O. (2013). Group affinity based social trust model for an intelligent movie recommender system. En: Multimedia Tools and Applications, Vol. 64, pp. 505-516.
- [11] Wasserman, S., Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge. Cambridge University Press.
- [12] Kossinets, G., Kleinberg, J. M., Watts, D. J. (2001). The structure of information pathways in a social communication network. En: 14th ACM SIGKDD

- International Conference on Knowledge Discovery and Data Mining.
- [13] Aiken, L. (2003). Test psicológicos y evaluación. México. Prentice Hall.
- [14] Morgeson, M., Reider, F., Campion, M. (2005). Selecting individuals in team settings: The importance of social skills, personality characteristics, and teamwork knowledge. En: Personnel Psychology, Vol. 58 (3), pp. 583-611.
- [15] Leskovec, J., Rajaraman, A., Ullman, J. D. (2014). Mining of Massive datasets. Cambridge University Press.
- [16] Gompers, P., Mukharlyamov, V. (2015). The Cost of Friendship. En: Journal of Financial Economics, aceptado para publicación.
- [17] Fershtman, C., Gneezy, U. (2001). Discrimination in a Segmented Society: An Experimental Approach. En: Quarterly Journal of Economics, Vol. 115 (3), pp. 351-377.
- [18] Kempe, D., Kleinberg, J. M., Tardos, E. (2003). Maximizing the spread of influence through a social network. En: 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
- [19] Rao, D., Yarowsky, D., Callison-Burch, C.;(2008). Affinity measures based on the graph Laplacian. En: Proceedings of the 3rd TextGraphs Workshop on Graph-based Algorithms for NLP, pp. 41-48.
- [20] Liu, H., Yang, X., Latecki, L. J., Yan, S. (2012). Dense neighborhoods on affinity graph. International Journal of Computer Vision, Vol. 98 (1), pp. 65-82.