

Análisis Estadísticos de Interacciones de Usuarios en Redes Sociales

Diego Medel^[0000-0001-7465-7062], Laura Aballay^[0000-0001-5462-7683],
Silvana Aciar^[0000-0002-6281-6924]

Instituto de Informática, Universidad Nacional de San Juan, San Juan, Argentina
{dmedel,lballay}@iinfo.unsj.edu.ar,saciar@unsj-cuim.edu.ar

Resumen En la actualidad, las redes sociales son utilizadas por millones de personas. Los usuarios se relacionan mediante publicaciones, reacciones, comentarios, entre otras formas. Resulta de gran utilidad un análisis con la información que proporcionan estas interacciones, estudiando el comportamiento de los usuarios para determinar la influencia que existe entre ellos. En este trabajo, se realizó un análisis estadístico sobre las interacciones entre usuarios de una red social orientada a la calificación de negocios. Se empleó un análisis de componentes principales para estudiar la correlación entre las acciones hechas por los usuarios y un análisis de conglomerados para agrupar a usuarios de acuerdo a sus interacciones.

Keywords: social relationships · social networks · multivariate analysis · recommender system.

1. Introducción

En la actualidad, las redes sociales son utilizadas por millones de personas. Su principal función es servir a la sociabilidad, y son cada vez más versátiles y completas en este sentido. La sociabilidad “virtual” entre personas puede darse por medio de comentarios, compartir publicaciones, imágenes, me gusta (likes), etc. Las redes sociales, como Instagram¹ permiten publicar fotos (y videos) para sus seguidores, Twitter² publicar texto con un número limitado de caracteres, o Tinder³ en contactar personas que se encuentren en un cierto radio de distancia para concretar citas, son algunos ejemplos. Las aplicaciones como Foursquare⁴, Facebook places⁵ y Yelp⁶, son redes sociales más populares entre los servicios basados en la ubicación, que permiten a los usuarios compartir fácilmente su ubicación geográfica y los contenidos relacionados a dicha ubicación en forma

¹ <https://www.instagram.com/>

² <https://twitter.com/>

³ <https://tinder.com/>

⁴ <https://es.foursquare.com/>

⁵ <https://www.facebook.com/places/>

⁶ <https://www.yelp.com/>

online. Otro tipo de redes sociales son las basadas en eventos sociales, como lo es Meetup ⁷ para reuniones grupales.

Debido al gran tráfico de información que generan las redes sociales, los investigadores exponen diversas técnicas en recuperación de información relevante. La entrega de contenido personalizado y adaptativo es un problema de investigación dentro del campo de la recuperación de información y sistemas de recomendación [1][2]. Los sistemas de recomendación sensibles al contexto ponen énfasis en el entorno externo, como puede ser el clima, ubicación actual y en el entorno interno del usuario, como es el estado emocional del mismo [3]. En [4] coinciden que el uso de la información contextual de un usuario puede mejorar la performance de un sistema de recomendación. Sin embargo, la mayor preocupación en diseñar un sistema de recomendación sensible al contexto es encontrar factores o variables del contexto que sean de valor para la predicción o recomendación.

En la vida cotidiana, la interacción entre personas siempre se encuentra presente, tales como, relaciones de amistad o de trabajo como así también relaciones con un mayor grado de cercanía que otras. Según [5] el contexto grupal es caracterizado por la presencia de varias personas, en dicho trabajo, refleja estudios recientes que demuestran que la satisfacción individual es influenciada por la satisfacción de otros, conocido como contagio emocional. Aunque remarca que la influencia es difícil de explicar, afirma que los factores tales como personalidad y tipo de relación impactan en el contagio emocional. Para entender el cambio en la satisfacción del usuario en un contexto de grupo es importante identificar los factores que caracterizan la influencia. A su vez, en [6] se asume la presencia de dos claves en la influencia: el contagio emocional y la conformidad o satisfacción. El contagio emocional es cómo influyen otras personas en el estado afectivo del usuario y cómo adaptarse en caso de la conformidad.

Yelp es una red social donde los usuarios, además de calificar un negocio al que asistieron, pueden valorar reseñas hechas por otros usuarios. En este trabajo se estudiaron las calificaciones o reseñas hechas por los usuarios de Yelp, como así también las calificaciones que reciben de otros usuarios por medio de un análisis multivariado. Se desea analizar si los usuarios son influenciados por otros que tienen mayor cantidad de reseñas realizadas o un alto promedio de calificaciones en las mismas.

Se llevó a cabo un Análisis de Componentes Principales (ACP), que es una técnica estadística utilizada para resumir la información, o reducir la dimensión (número de variables). Es decir, ante un banco de datos con muchas variables, el objetivo será reducirlas a un menor número perdiendo la menor cantidad de información posible. Los nuevos componentes principales o factores serán una combinación lineal de las variables originales, y además serán independientes entre sí. Este análisis permitió evidenciar la existencia de relaciones de influencia entre las acciones de los usuarios, como así también, en los cumplidos recibidos. Por otro lado, se llevó a cabo un análisis de conglomerados agrupando elementos en forma homogénea en función de las similitudes o similaridades entre ellos [7].

⁷ <http://www.meetup.com/about/>

Este análisis proporcionó agrupar los usuarios de acuerdo a su actividad en calificar reseñas así como en la interacción con la red social.

Este artículo se estructura de la siguiente manera: en la sección 2 se describen los trabajos relacionados a este análisis. En la sección 3, se describe el caso de estudio y se detalla el trabajo, a su vez, en la sección 4 y 5 se exponen los análisis de componentes principales y de conglomerados respectivamente. Por último, las conclusiones generales y trabajos futuros en la sección 6.

2. Trabajos Relacionados

En [8] los autores diseñan un sistema de recomendación, en el cual consideran las relaciones de confianza, y además, usuarios que viven en áreas cercanas que hayan calificados items de manera similar. Analizan la red social Epinions⁸, desde su dataset obtienen información de los usuarios inherente a la calificación de los negocios (un número del 1 a 5) y la red de confianza de cada usuario. De esta manera, su propuesta ha sido evaluada con las métricas de precisión y recall obteniendo resultados favorables. A diferencia de nuestro trabajo, nuestro análisis incluye todas las acciones del usuario en la red, además, no se tiene en cuenta ningún tipo de relación de amistad o de confianza.

En [9], proponen un mecanismo de recomendación social para mejorar el comercio electrónico de O2O (Online-to-offline), emplean un análisis de cercanía entre dos usuarios evaluando sus registro de llamadas, de mensajes e información de las interacciones en Facebook. Además, analizan la actividad del usuario teniendo en cuenta la cantidad de interacciones (llamadas, mensajes, publicaciones) en un período de tiempo t . A partir de un análisis de cercanía entre el usuario u y v ; y un análisis de actividad del usuario u , calculan la influencia del usuario u sobre el usuario v . El concepto de confianza o influencia es muy utilizada por investigadores para mejorar la precisión de los sistemas de recomendación, denotando una influencia de usuarios sobre otros en la sugerencia.

En el trabajo [10], proponen un sistema de recomendación basado en la confianza a partir de las interacciones entre usuarios en la red social Twitter. Por medio de recopilación de datos relacionados a tweets, hashtag, retweet y favoritos desarrollan métricas para evaluar la confianza entre usuarios. De manera similar en [11], se calcula el grado de confianza en base a intereses similares y a interacciones sociales entre usuarios (comentarios, mensajes, etc) con respecto al tiempo.

En [12] analizan la confianza que genera el organizador del evento en una red social basada en eventos (event-based social networks, EBSN), su análisis de influencia se basa en el promedio de calificaciones que reciben los eventos organizados y el contenido que aportan los mismos. A su vez, en [13] destacan que si un usuario participa a menudo en eventos organizados por un anfitrión, la probabilidad de que asista a otros eventos similares organizados por este debería ser alta. Estos trabajos realizan una observación entre un usuario (organizador

⁸ <http://www.epinions.com/>

del evento), con respecto a los demás, en contraste, en el presente trabajo se plantea un análisis general entre todos los usuarios evaluando las interacciones sociales que presenta la red social pero sin importar el rol que ocupe cada usuario.

3. Caso de Estudio

Yelp permite los usuarios hacer reseñas (en inglés, reviews), que son comentarios o sugerencias acerca de un negocio en particular. Asimismo, las reseñas pueden ser calificadas con estrellas, de 1 a 5, y el promedio de calificaciones totales se almacena en la variable `average-stars`. Además un usuario puede calificar las reseñas de otros, siendo a criterio del usuario-lector calificar la reseña de otro como útil (en inglés, useful), divertida (en inglés, funny) o genial (en inglés, cool) si así lo desea. Por otro lado, los usuarios también pueden calificarse entre sí teniendo en cuenta sus actividades en la red social, lo que en Yelp se llaman cumplidos.

Para este artículo, se trabajó con la tabla usuarios (users), en la Tabla 3 se describen las variables que se tuvieron en cuenta en el análisis. Yelp dataset⁹ es un subconjunto de negocios (del inglés businesses), reseñas (del inglés reviews), y usuarios registrados. Esta base de datos se puede descargar en archivos JavaScript Object Notation (JSON) y Structured Query Language (SQL). Para llevar

| Variable | Descripción |
|--------------------------------|---|
| <code>review_count</code> | Número de reseñas que ha escrito |
| <code>useful</code> (útil) | Número de votos útiles enviados por el usuario |
| <code>funny</code> (divertido) | Número de votos divertidos enviados por el usuario |
| <code>cool</code> (genial) | Número de votos geniales enviados por el usuario |
| <code>fans</code> | Número de fans que tiene el usuario |
| <code>average_stars</code> | Puntuación media de todas las evaluaciones |
| <code>compliment_cool</code> | número de cumplidos “cool” recibidos por el usuario |
| <code>compliment_funny</code> | número de cumplidos “funny” recibidos por el usuario |
| <code>compliment_writer</code> | número de cumplidos “writer” recibidos por el usuario |
| <code>compliment_hot</code> | número de cumplidos “hot” recibidos por el usuario |

Cuadro 1. Variables usadas. Tabla Users (Usuarios)

a cabo el estudio del comportamiento de los usuarios de esta red, se analizó la relación entre la cantidad de reseñas (`review_count`) y el promedio de calificaciones (`average_stars`) con los votos divertido (`funny`), útil (`useful`) y genial (`cool`) emitidos por el usuario. Por otro lado, también se analizó la relación entre la cantidad

⁹ <https://www.yelp.com/dataset>

de reseñas (`review_count`) y el promedio de calificaciones (`review_count`) hechas por un usuario con los cumplidos recibidos por otros usuarios. Para ello se llevó a cabo en primer lugar un análisis preliminar, un proceso de análisis exploratorio univariado de las variables a utilizar, análisis de componentes principales, y además se realizó un análisis de conglomerados.

3.1. Análisis Preliminar

Se revisan las medidas resumen más usadas en un análisis descriptivo univariado, ya que es recomendable que todo análisis comience con un análisis descriptivo (gráficos y medidas resumen) de cada una de las variables antes de continuar con el análisis multivariado. En este análisis, se obtuvieron valores de Kurtosis extremadamente alta, lo que indica la existencia de puntos atípicos. Por tal motivo, se decidió realizar la transformación del logaritmo (\ln) de todas las variables (Figura 1) antes de los siguientes análisis estadísticos [7].

| Variable | n | Media | D.E. | Mín | Máx | Q1 | Q3 | Kurtosis |
|------------------------------------|-----|-------|------|------|-------|------|------|----------|
| <code>LN_review_count</code> | 400 | 4,25 | 1,67 | 0,00 | 8,31 | 3,04 | 5,42 | -0,21 |
| <code>LN_useful</code> | 327 | 3,36 | 2,33 | 0,00 | 10,86 | 1,61 | 4,34 | 0,33 |
| <code>LN_funny</code> | 294 | 3,01 | 2,32 | 0,00 | 9,71 | 1,10 | 4,62 | -0,25 |
| <code>LN_cool</code> | 281 | 2,93 | 2,46 | 0,00 | 10,67 | 0,69 | 4,53 | 0,09 |
| <code>LN_fans</code> | 317 | 2,25 | 1,61 | 0,00 | 7,51 | 1,10 | 3,14 | 0,15 |
| <code>LN_average_stars</code> | 400 | 1,34 | 0,15 | 0,00 | 1,61 | 1,29 | 1,41 | 19,85 |
| <code>LN_compliment_hot</code> | 254 | 2,34 | 2,03 | 0,00 | 8,49 | 0,69 | 3,43 | 0,25 |
| <code>LN_compliment_nice</code> | 259 | 1,48 | 1,26 | 0,00 | 5,64 | 0,69 | 2,20 | 0,62 |
| <code>LN_compliment_profile</code> | 182 | 1,16 | 1,38 | 0,00 | 5,59 | 0,00 | 1,79 | 1,03 |
| <code>LN_compliment_cute</code> | 184 | 1,53 | 1,67 | 0,00 | 6,72 | 0,00 | 2,30 | 0,52 |
| <code>LN_compliment_list</code> | 115 | 1,49 | 1,39 | 0,00 | 4,99 | 0,00 | 2,40 | -0,38 |
| <code>LN_compliment_nice</code> | 263 | 2,03 | 1,71 | 0,00 | 7,16 | 0,69 | 2,94 | 0,38 |
| <code>LN_compliment_plain</code> | 292 | 2,27 | 1,84 | 0,00 | 8,66 | 0,69 | 3,09 | 0,79 |
| <code>LN_compliment_cool</code> | 288 | 2,53 | 1,91 | 0,00 | 8,98 | 1,10 | 3,58 | 0,17 |
| <code>LN_compliment_funny</code> | 299 | 2,53 | 1,91 | 0,00 | 8,98 | 1,10 | 3,58 | 0,17 |
| <code>LN_compliment_writer</code> | 267 | 1,89 | 1,60 | 0,00 | 7,40 | 0,69 | 2,71 | 0,76 |
| <code>LN_compliment_photos</code> | 179 | 1,46 | 1,56 | 0,00 | 7,78 | 0,00 | 2,30 | 1,48 |

Figura 1. Tabla de Medidas luego de Transformar a \ln .

4. Análisis de Componentes Principales

Para realizar un Análisis de Componentes Principales (ACP) de forma exploratoria o descriptiva, no es necesario que los datos sigan una distribución normal. Lo único a controlar es que las variables sean cuantitativas o de una naturaleza que otorgue sentido al calcular varianzas y covarianzas y que las correlaciones subyacentes sean lineales, por esto último, es que algunas veces las variables originales son transformadas a escala logaritmo antes de realizar el análisis. Un análisis de componentes principales tiene sentido si existen altas correlaciones entre las variables, ya que esto es indicativo de que existe información redundante y, por tanto, pocos factores explicarán gran parte de la variabilidad total [7]. Antes de aplicar el ACP debe comprobarse si es necesario, es decir, si la

correlación entre las variables analizadas es lo suficientemente grande como para justificar la factorización de la matriz de coeficientes de correlación.

4.1. Acciones de los Usuarios sobre las Calificaciones o Reseñas (reviews)

El índice de Kaiser-Meyer-Olkin o medida de adecuación muestral KMO, compara los valores de las correlaciones entre las variables y sus correlaciones parciales. Si el índice KMO está próximo a 1, el ACP se puede hacer. Si el índice es bajo (próximo a 0), el ACP no será relevante. Algunos autores han definido una escala para interpretar el índice KMO de un conjunto de datos [14].

La elección de los factores se realiza de tal forma que el primero recoja la mayor proporción posible de la variabilidad original; el segundo factor debe recoger la máxima variabilidad posible no recogida por el primero, y así sucesivamente. Del total de factores se elegirán aquéllos que recojan el porcentaje de variabilidad que se considere suficiente. A éstos se les denominará componentes principales.

En los gráficos biplot, los ángulos agudos indican correlaciones positivas, ángulos obtusos corresponden a correlaciones negativas y ángulos rectos indican que no hay correlación entre las variables. En estos gráficos no se busca la proximidad entre observaciones y variables, sino sus direcciones es lo que importa.

Se calculó el índice KMO con un valor cercano a 1, lo cual, es aceptable realizar el análisis. Además, observando la Figura 2 de autovalores se puede observar que con dos componentes se explica el 91 % de la totalidad de los datos. De esta manera, son suficientes para explicar el comportamiento de los datos. Luego, en la Figura 2 es posible diferenciar los dos componentes, en la primera columna, el componente principal 1 (CP1), es de forma, debido a que tiene valores con diferente signo y presenta un 70 % de la variabilidad total de los datos. Sin embargo, el único valor con signo negativo corresponde a la variable average-stars. Todas las variables restantes explican el CP1 con coeficientes similares. Este componente se podría denominar “acciones de los usuarios”. Con respecto al CP2, explica el 20 % de los datos, y también se observan valores con diferentes signos, pero esta vez la variable average-stars se lleva toda la explicación, con un valor de 0.99. El resto de las variables no determinan influencia en este componente. A este componente se le puede denominar tal como la variable que lo explica “promedio de calificaciones del usuario”. Por un lado, la variable average-stars forma casi un ángulo de 90° con el resto de las variables (Figura 3). Esto se interpreta que dicha variable no tiene correlación con el resto. Por lo tanto, se puede decir que se observa que el promedio de calificaciones utilizado por los usuarios no tiene influencia en los votos que usan ni en la cantidad de fans que los siguen. A su vez, las variables que corresponden a los votos funny, cool y useful y fans están estrechamente correlacionadas entre sí, ya que sus autovectores se muestran muy cercanos, formando ángulos agudos. Es decir que, se observa que los usuarios votan de forma equilibrada las reseñas (reviews) que visitan. Además, la variable review_count (cantidad de reseñas hechas) se observa muy cercana a todo el conjunto de variables pero se nota opuesta a la

Autovalores

| Lambda | Valor | Proporción | Prop Acum |
|--------|-------|------------|-----------|
| 1 | 3,94 | 0,66 | 0,66 |
| 2 | 1,02 | 0,17 | 0,83 |
| 3 | 0,67 | 0,11 | 0,94 |
| 4 | 0,22 | 0,04 | 0,97 |
| 5 | 0,09 | 0,02 | 0,99 |
| 6 | 0,06 | 0,01 | 1,00 |

Autovectores

| Variables | e1 | e2 |
|------------------|-------|----------|
| LN_average_stars | -0,04 | 0,99 |
| LN_review_count | 0,37 | -0,07 |
| LN_fans | 0,44 | 0,14 |
| LN_useful | 0,47 | -0,04 |
| LN_funny | 0,47 | -6,1E-04 |
| LN_cool | 0,47 | 0,05 |

Figura 2. Tabla autovalores y autovectores

variable average-stars. Es decir, que la cantidad de reseñas del usuario no tienen relación con las calificaciones colocadas por ellos.

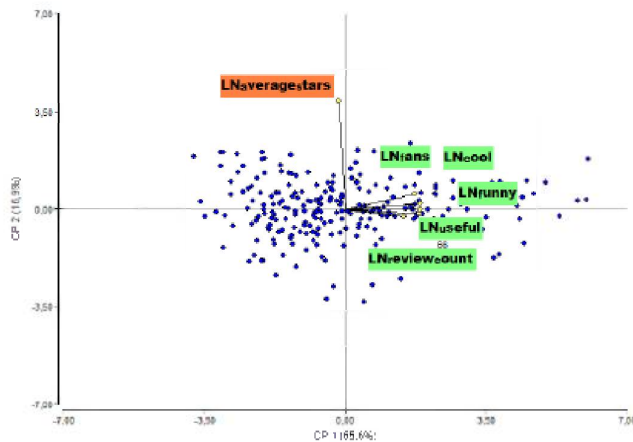


Figura 3. Gráfico biplot

4.2. Cumplidos Recibidos de un Usuario

Es posible que los usuarios reciban cumplidos (similar a “me gusta”) acerca de sus reseñas o calificaciones por parte de otros usuarios. Estos cumplidos pueden ser de diferentes tipos. Se trata de analizar si estos cumplidos tienen relación

con el número de reseñas (variable `review_count`) o con el promedio de estrellas (variable `average-stars`) del usuario. Se calculó el índice KMO, se obtuvo un número cercano a 1, lo cual es aceptable realizar el análisis.

Observando Figura 4 de autovalores se puede observar que con dos componentes se explica el 87% de la totalidad de los datos. De manera que son suficientes para explicar el comportamiento de los datos. Luego, en la tabla de autovectores (Figura 4) es posible diferenciar los dos componentes, en la primera columna, el CP1, es de forma, debido a que tiene valores con diferente signo y presenta un 79% de la variabilidad total de los datos. Sin embargo, el único valor con signo negativo corresponde a la variable `average-stars`. Todas las variables restantes explican el CP1 con coeficientes similares. Este componente se podría denominar “cumplidos recibidos”. Con respecto al CP2, explica el 0.8% de los datos, y también se observan valores con diferentes signos, pero esta vez la variable `average-stars` se lleva toda la explicación, con un valor de 0.87. El resto de las variables no determinan influencia en este componente. A este componente se le puede denominar tal como la variable que lo explica “promedio de calificaciones del usuario”. En el gráfico biplot (Fig 5) las variables `average-stars` y

| Autovalores | | | |
|-------------|-------|------------|-----------|
| Lambda | Valor | Proporción | Prop Acum |
| 1 | 10,27 | 0,79 | 0,79 |
| 2 | 1,08 | 0,08 | 0,87 |
| 3 | 0,70 | 0,05 | 0,93 |
| 4 | 0,25 | 0,02 | 0,95 |
| 5 | 0,19 | 0,01 | 0,96 |
| 6 | 0,12 | 0,01 | 0,97 |
| 7 | 0,11 | 0,01 | 0,98 |
| 8 | 0,09 | 0,01 | 0,99 |
| 9 | 0,07 | 0,01 | 0,99 |
| 10 | 0,05 | 4,2E-03 | 1,00 |
| 11 | 0,03 | 2,1E-03 | 1,00 |
| 12 | 0,02 | 1,7E-03 | 1,00 |
| 13 | 0,00 | 0,00 | 1,00 |

| Autovectores | | |
|------------------------------------|-------|-------|
| Variables | e1 | e2 |
| <code>LN_compliment_hot</code> | 0,30 | 0,10 |
| <code>LN_compliment_more</code> | 0,30 | 0,01 |
| <code>LN_compliment_profile</code> | 0,30 | 0,05 |
| <code>LN_compliment_cute</code> | 0,29 | 0,12 |
| <code>LN_compliment_list</code> | 0,28 | 0,02 |
| <code>LN_compliment_nice</code> | 0,30 | -0,06 |
| <code>LN_compliment_plain</code> | 0,30 | -0,01 |
| <code>LN_compliment_cool</code> | 0,31 | 0,04 |
| <code>LN_compliment_funny</code> | 0,31 | 0,04 |
| <code>LN_compliment_writer</code> | 0,29 | -0,02 |
| <code>LN_compliment_photos</code> | 0,29 | 0,03 |
| <code>LN_average_stars</code> | -0,03 | 0,87 |
| <code>LN_review_count</code> | 0,16 | -0,44 |

Figura 4. Tabla autovalores y autovectores

`review-count` forman casi un ángulo de 90° con el resto de las variables. Esto se interpreta que estas variables no tienen correlación con el resto. Por lo tanto, se

puede decir que tanto el promedio de calificaciones como la cantidad de reseñas de un usuario no tiene influencia en los cumplidos que reciben de otros usuarios. Por otro lado, las variables que corresponden a los cumplidos (“me gusta como escribes”, “me gusta tu perfil”, “me gusta tu foto”, etc) están estrechamente correlacionadas entre sí, ya que sus autovectores se muestran muy cercanos. Es decir que, se observa que no hay diferencias significativas entre los cumplidos que le llegan al usuario. Además, la variable review-count se observa ortogonal al conjunto de variables de cumplidos, lo que denota una baja correlación con ellas. Y completamente opuesta al promedio de estrellas (variable average-stars) que coloca el usuario. Es decir, que la cantidad de reseñas del usuario no tienen relación con las calificaciones colocadas por ellos ni con los cumplidos recibidos.

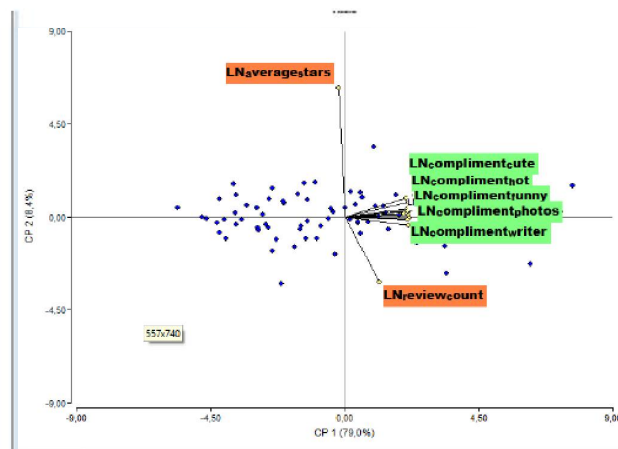


Figura 5. Gráfico biplot

5. Análisis de Conglomerados

En esta sección se pretende estudiar el agrupamiento de los usuarios de acuerdo a sus comportamientos en la red social. Se realizaron dos análisis utilizando el método de k-means. En el primer análisis de agrupamiento se estudió como es el comportamiento de los usuarios calificando reseñas de otros usuarios. En el segundo estudio se analizó como los usuarios reciben cumplidos de distintos usuarios lectores.

En el primer análisis se estudió el agrupamiento con las variables review_count, average_stars, useful, funny, cool y fans. En primera instancia, se analizó los valores de silueta, teniendo en cuenta la interpretación Rousseeuw [15]. Se obtuvo

hasta 5 conglomerados. Se eligió el grupo de 2 conglomerados debido a que el valor de silueta general es el que mejor se adapta a la interpretación de Rousseeuw. En la matriz de correlación del primer conglomerado se observó una correlación fuerte entre *useful* – *funny* (0.68), *useful* – *cool* (0.70), *cool*-*funny* (0.68). Es decir que estas variables que detallan las calificaciones que hacen los usuarios sobre las reseñas realizadas de terceros están relacionadas. No obstante, hay una correlación del 0.52 entre la cantidad de reseñas del usuario con sus fans. La variable promedio de calificaciones de reseñas tiene muy poca correlación con las demás variables. En la matriz de correlación del segundo conglomerado se observa una correlación fuerte entre *useful* – *funny* (0.90), *useful* – *cool* (0.94), *cool*-*funny* (0.95) . Es decir que estas variables que detallan las calificaciones que hacen los usuarios sobre las reseñas realizadas de terceros están relacionadas. No obstante, con una correlación del 0.65 la cantidad de reseñas del usuario con sus fans. La variable promedio de calificaciones de reseñas tiene muy poca correlación con las demás variables.

En ambos conglomerados el análisis es similar, las correlaciones con mayores valores son similares entre variables como así también la correlación baja en la variable *average_stars*. Sin embargo, en el vector de media de cada conglomerado (Fig 6) se puede ver una diferencia de los valores medios de cada vector con respecto al otro. Por ejemplo, la variable *funny*, en el primer conglomerado, el valor medio es 2.34 y , en el segundo conglomerado el valor es de 5.99, denotando una diferencia en término de media en calificaciones realizadas por el usuario. Vale aclarar que el valor de la variable *average_stars* es el mismo en ambos vectores, indicando que es totalmente indiferente del análisis de agrupamiento que se realizó. La interpretación en este análisis de conglomerados utilizando

```

1
Vector medio total
LN review count LN useful LN funny LN cool LN fans LN average stars
4,68      2,90      2,34      1,99      1,94      1,35

2
Vector medio total
LN review count LN useful LN funny LN cool LN fans LN average stars
6,08      6,42      5,99      6,05      4,14      1,35

```

Figura 6. Vector de media. Conglomerado 1 y 2

el método empleado con k-means es la división de los usuarios en dos grupos de acuerdo a sus calificaciones en las reseñas de otros usuarios, sin tener en cuenta tanto la cantidad de reseñas realizadas como el promedio general de sus calificaciones. En un grupo se encuentran los usuarios que en general han tenido una poca participación en calificar reseñas y en el otro grupo los que tiene una mayor actividad en calificar reseñas de otros usuarios.

En el segundo análisis se estudió el agrupamiento de las variables *review_count*, *compliment_profile*, *compliment_plain*, *compliment_photos*, *average_stars*, *com-*

pliment_cute, compliment_cool, compliment_hot, compliment_list, complement_funny, compliment_more, compliment_note, compliment_writer. Se obtuvieron hasta grupos de 5 conglomerados, pero se eligió el grupo de 2 conglomerados debido a que el valor de silueta general es el que mejor se adapta a la interpretación de Rousseeuw (valores de 0.51 y 0.70). Tanto el grupo 2 como el 3 son óptimos para el análisis. En la matriz de correlación del conglomerado 1 se obtuvieron correlaciones fuertes entre las variables compliment, en algunos casos cercano a 1. Hay una correlación muy baja entre la variable review_count y average_stars con respecto a las demás variables compliment y entre ambas sucede lo mismo. En el conglomerado 2 el análisis es similar al anterior. Se observan correlaciones muy fuertes entre las variables compliment y correlaciones bajas entre review_count y average_stars. A continuación, se analizan los valores de cada vector de media de cada conglomerado y se puede observar una diferencia significativa entre ambos vectores. Se observan diferencias en cada valor medio de cada variable

```

Vector medio total
LN review count LN average stars LN compliment hot LN compliment more LN compliment profile LN compliment cute
-----
6,62          1,35          6,51          3,99          3,68          4,56

Vector medio total
LN review count LN average stars LN compliment hot LN compliment more LN compliment profile LN compliment cute
-----
3,73          1,30          3,57          2,10          1,30          1,73

```

Figura 7. Vector de media. Conglomerado 1 y 2

compliment de los dos conglomerados (Fig 7). Por ejemplo, el valor de compliment_profile es de 3.68 en un vector y en el otro 1.30 indicando una diferencia en sus valores medio. Sin embargo, no sucede lo mismo con average_stars donde el valor medio es prácticamente igual y con review_count una diferencia menor en ambos vectores.

La interpretación en este análisis de conglomerados utilizando el método empleado con k-means es la división de los usuarios en dos grupos de acuerdo a la cantidad de cumplidos recibidos por usuarios, sin tener en cuenta tanto la cantidad de reseñas realizadas como el promedio general de sus calificaciones. Se puede deducir que el grupo que tiene usuarios con un número alto de cumplidos tienen una mayor interacción con la red social con respecto al otro grupo.

6. Conclusiones

La realización de este trabajo permitió estudiar mediante un análisis estadístico el comportamiento de los usuarios de una red social basada en la calificación de negocios. Por medio de un análisis de componentes principales se determinó que el promedio de calificaciones y las acciones (votos funny, cool, useful) realizadas por un usuario no evidencian tener una relación de influencia, lo mismo sucede con la cantidad de fans del usuario. Son variables que se comportan de manera

independiente en la red social. De manera similar, la cantidad de reseñas del usuario no tiene relación con las calificaciones realizadas. Las acciones realizadas por el usuario son interpretadas como que el usuario emite votos de manera equilibrada a las distintas reseñas.

A su vez, se puede decir que tanto el promedio de calificaciones como la cantidad de reseñas de un usuario no tienen influencia en los cumplidos que reciben de otros usuarios. Estas conclusiones nos permiten afirmar que es información redundante o poco relevante para mejorar la precisión en el proceso de recomendación.

Se visualizan dos grupos bien definidos en términos de comportamiento. Usuarios que tienen una mayor actividad calificando reseñas que otros, como así también, usuarios que tienen una mayor interacción en la red social que otros.

Es importante remarcar que los resultados obtenidos permiten analizar el comportamiento de los usuarios en la red social de acuerdo a variables que los describen. Las conclusiones obtenidas demuestran que un usuario que participa activamente en la red calificando negocios no influye en la cantidad de cumplidos que reciben, como así tampoco en la cantidad de fans. No obstante, en un análisis a futuro se analizará el comportamiento de usuarios que mantienen una relación de amistad en la red. Además, se planea explorar el comportamiento de usuarios que se encuentren dentro de un cierto radio de distancia.

Referencias

1. S. T. Dumais, E. Cutrell, J. J. Cadiz, G. Jancke, R. Sarin, and D. C. Robbins, "Stuff I've seen," *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval - SIGIR '03*, vol. 49, no. 2, pp. 72–79, 2003.
2. P. Borlund, "Interactive Information Retrieval: An Evaluation Perspective," *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, p. 151, 2016. [Online]. Available: <http://doi.acm.org/10.1145/2854946.2870648>
3. F. Barile, J. Mastho, and S. Rossi, "The Adaptation of an Individual 's Satisfaction to Group Context : the Role of Ties Strength and Conflicts," pp. 357–358, 2017.
4. E. Ashley-Dejo, S. Ngwira, and T. Zuva, "A survey of Context-Aware Recommender System and services," in *2015 International Conference on Computing, Communication and Security, ICCCS 2015*, 2016.
5. F. Barile, J. Masthoff, and S. Rossi, "A Detailed Analysis of the Impact of Tie Strength and Conflicts on Social Influence," *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization - UMAP '17*, pp. 227–230, 2017. [Online]. Available: <http://dl.acm.org/citation.cfm?doid=3099023.3099056>
6. E. Quintarelli, E. Rabosio, and L. Tanca, "Recommending New Items to Ephemeral Groups Using Contextual User Influence," *Proceedings of the 10th ACM Conference on Recommender Systems - RecSys '16*, pp. 285–292, 2016. [Online]. Available: <http://dl.acm.org/citation.cfm?doid=2959100.2959137>
7. D. Peña, *Análisis de datos multivariantes*. Editorial desconocida, 2002. [Online]. Available: <https://books.google.com.ar/books?id=TrVIAAAACAAJ>

8. M. Razghandi and S. A. H. Golpaygani, "A Context-Aware and User Behavior-Based Recommender System with Regarding Social Network Analysis," *Proceedings - 14th IEEE International Conference on E-Business Engineering, ICEBE 2017 - Including 13th Workshop on Service-Oriented Applications, Integration and Collaboration, SOAIC 2017*, pp. 208–213, 2017.
9. L. F. Lin and Y. M. Li, "A Social Recommendation Mechanism for Enhancing O2O E-Commerce," *Proceedings - 2017 6th IIAI International Congress on Advanced Applied Informatics, IIAI-AAI 2017*, pp. 401–406, 2017.
10. A. Yadav, S. Chakraverty, and R. Sibal, "A survey of implicit trust on social networks," *Proceedings of the 2015 International Conference on Green Computing and Internet of Things, ICGCIoT 2015*, pp. 1511–1515, 2016.
11. A. Kalai, A. Wafa, C. A. Zayani, and I. Amous, "LoTrust: A social Trust Level model based on time-aware social interactions and interests similarity," *2016 14th Annual Conference on Privacy, Security and Trust, PST 2016*, pp. 428–436, 2016.
12. Z. Wang, Y. Zhang, Y. Li, Q. Wang, and F. Xia, "Exploiting social influence for context-aware event recommendation in event-based social networks," *Proceedings - IEEE INFOCOM*, 2017.
13. M. Wei and D. Wang, "CPERS: Contextual and personalized event recommender system," *Proceedings - 2016 International Conference on Computational Science and Computational Intelligence, CSCI 2016*, pp. 421–426, 2017.
14. A. S. Beavers, J. W. Lounsbury, J. K. Richards, S. W. Huck, G. J. Skolits, and S. L. Esquivel, "Educational Research," vol. 18, no. 6, 2013.
15. Peter J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0377042787901257>