

*Litterio, Arnaldo Mario; Nantes, Esteban Alberto;  
Larrosa, Juan Manuel*

# LA INFLUENCIA EN REDES SOCIALES ONLINE BIMODALES A TRAVÉS DEL CASO DE TRIPADVISOR

---

Informes de investigación IIATA

*2022, vol. 7, no. pp. 225-253*

*Litterio, A. M., Nantes, E. A., Larrosa, J. M. (2022). La influencia en redes sociales online bimodales a través del caso de Tripadvisor. Informes de investigación. En RIDCA. Disponible en:  
<https://repositoriodigital.uns.edu.ar/handle/123456789/6240>*



Esta obra está bajo una Licencia Creative Commons  
Reconocimiento-NoComercial-Sin Derivados 4.0 Internacional (CC BY-NC-ND 4.0)  
<https://creativecommons.org/licenses/by-nc-nd/4.0/>

## LA INFLUENCIA EN REDES SOCIALES ONLINE BIMODALES A TRAVÉS DEL CASO DE TRIPADVISOR

Litterio, Arnaldo Mario\*  
Nantes, Esteban Alberto\*\*  
Larrosa, Juan Manuel\*\*\*

### RESUMEN

El presente trabajo propone una aplicación de herramientas provenientes del análisis de redes sociales y de programación para la explotación de información de una comunidad online, más específicamente una red social bimodal, obtenida de la web a través del uso de la técnica de *web scraping*. Se detallan desde un enfoque teórico y práctico los pasos seguidos en el proceso de extracción y procesamiento de la información obtenida de *www.tripadvisor.com*, se genera un modelo de red social que relaciona diferentes tipos de actores dentro de la red, y se aplica un modelo para detectar de individuos influyentes propuesto anteriormente por el mismo grupo de investigación. Por último se describe la aplicación de herramientas de análisis cuantitativo a los datos obtenidos como minería de texto, frecuencia y nubes de palabras.

El trabajo aborda un problema de marketing contemporáneo desde los métodos cuantitativos y la teoría de redes sociales combinando técnicas conocidas en una forma novedosa. Su resultado es el descubrimiento de información valiosa no evidente desde otros métodos de análisis.

**PALABRAS CLAVE:** Análisis de Redes Sociales - Marketing Digital – Influenciadores - Web Scraping - Tripadvisor

### ABSTRACT

This work proposes an application of social network analysis tools and programming tools for the exploitation of information from an online community, more specifically a bimodal social network.

The steps followed in the process of extracting and processing the information from *www.tripadvisor.com* is detailed from a theoretical and practical approach. These steps include using web scraping, representing the information through graphs, applying a model

---

\* Litterio, Arnaldo Mario ([litterio@uns.edu.ar](mailto:litterio@uns.edu.ar)). Departamento de Ciencias de la Administración, Universidad Nacional del Sur, ([csadmin@uns.edu.ar](mailto:csadmin@uns.edu.ar)).

\*\* Nantes, Esteban Alberto ([enantes@uns.edu.ar](mailto:enantes@uns.edu.ar)). Instituto de Investigaciones Económicas y Sociales del Sur (IIESS) y Departamento de Economía, Universidad Nacional del Sur, ([dptecono@criba.edu.ar](mailto:dptecono@criba.edu.ar))

\*\*\* Larrosa, Juan Manuel ([jlarrosa@criba.edu.ar](mailto:jlarrosa@criba.edu.ar)). San Andrés 800, Altos de Palihue, Código Postal 8000, Bahía Blanca, Argentina.

to detect influential individuals previously proposed by the same research group and the application of quantitative analysis tools to the data obtained such as text mining, frequency and word clouds.

The work addresses a contemporary marketing problem in the perspective of quantitative methods and social network theory, combining known techniques in a novel way. Its result is the discovery of valuable information not evident from other analysis methods.

**KEYWORDS:** Social network analysis; Digital marketing; Influencers; web-scraping; Tripadvisor

## 1. INTRODUCCIÓN

En el mundo actual es difícil encontrar organizaciones que desconozcan y no incorporen herramientas del marketing digital a su abordaje estratégico y táctico del mercado. El marketing digital superó el rol de canal publicitario que le atribuyó la llamada Web 1.0 y hoy en día el análisis del consumidor, el entorno competitivo y la gestión del conocimiento, posicionamiento y promoción de productos y servicios en el mundo online son funciones claves del marketing moderno.

Esto sucedió como consecuencia del profundo cambio cultural y social que supone un mundo digitalizado, geolocalizado e interconectado en tiempo real a través de dispositivos móviles inteligentes. Esta tendencia no hizo más que ser aumentada y consolidada por el impacto de la pandemia de Covid-19 que encerró a las personas y las llevó a socializar, educarse, informarse y consumir a través de plataformas tecnológicas.

El marketing contemporáneo debe aportar a las organizaciones comerciales, políticas y no gubernamentales un mapa para desarrollar su actividad trazado en un terreno en el que se entrelazan disciplinas otrora no vinculadas como el marketing tradicional, el análisis de las redes sociales, el análisis estadístico y el aprendizaje automático.

Las redes sociales online facilitan el diálogo social online entre personas, marcas e instituciones, proveyendo una plataforma en la que opiniones y comportamientos de los consumidores se ven afectados por complejas redes de influencia social a través del boca a boca digital o *electronic word of mouth (eWOM)*.

El contenido generado por usuarios de Internet a través de publicaciones online como *blogs*, reseñas en sitios especializados y el intercambio directo de opiniones y percepciones tiene la propiedad de resultar más influyente que los contenidos generados comercialmente en la decisión de compra de los usuarios expuestos a ese contenido y potencialmente a su red de contactos con un efecto multiplicador.

Este trabajo está organizado de modo que en la primera parte se repasan conceptos teóricos que entrelazan las disciplinas del marketing y el análisis de las redes sociales,

incluyendo el estudio del boca a boca o *word of mouth*, el rol del influyente en la decisión de compra, el estudio y análisis de redes sociales en general y redes sociales online en particular. Se introducen además los métodos y técnicas utilizadas para obtener y procesar la información obtenida de la web con particular detalle en el *web scraping* implementado a través del lenguaje de programación R, y el análisis y procesamiento de información en ese mismo entorno.

Seguidamente se introduce el caso de estudio y se detallan los pasos seguidos para extraer los datos de la web. Se desarrolla en concreto el estudio de la red conformada por el conjunto de restaurantes y usuarios que publicaron opiniones, o *reviewers* de la ciudad de Bahía Blanca obtenidos del sitio [www.tripadvisor.com](http://www.tripadvisor.com).

En una tercera instancia se ofrecen análisis preliminares sobre la información obtenida, se aplica y analiza la red a la luz del modelo de detección de influyentes a partir de métricas de centralidad propuesto en *Marketing and social networks: a criterion for detecting opinion leaders* (Litterio, Nantes, Larrosa & Gomez, 2017), y se contrasta su utilidad a través de la correlación entre los resultados obtenidos por la aplicación del método y una métrica relevante.

Seguidamente se demuestran elementos de análisis de minería de texto con R.

Finalmente, se ofrecerán las conclusiones a las que se arribó y sus limitaciones, en busca de promover la discusión y posibles extensiones a este trabajo.

## 2. MARCO TEÓRICO

### El boca a boca y el marketing online

La comunicación tradicional de las marcas con sus potenciales consumidores es unidireccional y deja fuera de cuestión la interacción con los consumidores, rol que ha quedado reservado en el mejor de los casos a la investigación e inteligencia de mercados (Bacile, Ye & Swilley, 2014; Benedetti, 2015).

El boca a boca o *word of mouth*, ha sido extensivamente estudiado y reconocido como un factor influenciador clave en las decisiones de compra de los consumidores (Lang & Hyde, 2013; King, Racherla & Bush, 2014), y tiene un efecto positivo en la relación del consumidor con la marca, y en otros resultados del marketing (Hudson, Huang, Roth & Madden, 2016; Wang & Gon Kim, 2017).

Hewett (Hewett, Rand, Rust & van Heerde, 2016) postula que la naturaleza de la comunicación de marca ha cambiado con el advenimiento de las tecnologías online, y cuantifica su efecto en el sentimiento del consumidor y resultados de negocio.

Hoy el proceso de compra de un producto o servicio está fuertemente permeado por la experiencia del potencial consumidor dentro del ecosistema de Internet en general y las redes sociales en particular. A través de la web se pueden descubrir nuevos productos,

analizar opiniones y comparar alternativas, ver videos de su uso, concretar la transacción y finalmente compartir la experiencia de uso con el resto del mundo conectado.

La comunidad de potenciales consumidores de una marca comienza a dialogar entre sí y con la marca en lugar de ser simple receptora de un mensaje, y esta interacción afecta profundamente sus percepciones y decisiones de compra (Benedetti, 2015).

Las redes sociales online tienen un denominador común que es su dependencia del contenido generado por los propios usuarios, el cual generalmente se encuentra relacionado con marcas y tiene el potencial de influenciar la percepción de la marca por parte de los consumidores (Smith, Fischer & Yongjian, 2012; Nam & Kannan, 2014).

El “apego” a los medios sociales online está relacionado positivamente con conductas de consumo y apoyo a las marcas, lo que convierte a algunas personas en objetivos naturales de iniciativas de marketing (VanMeter, Grisaffe & Chonko, 2015).

Varios estudios previos señalan una mayor capacidad del contenido generado por usuarios para generar interés sobre un tópico superando los contenidos generados comercialmente (Bickart & Schindler, 2001); el efecto del boca a boca en la confianza, lealtad e intención de compra (Awad & Ragowsky, 2008; Ekran & Evans, 2016; Chen, Fay & Wang, 2011; Pavlou & Ba, 2002) y recompra (Gauri, Bhatnagar & Rao, 2008). La gente común confía en las opiniones desinteresadas publicadas online, lo que sugiere que las empresas deberían enfocarse en los mecanismos que faciliten la dispersión del boca a boca (Duan, Gu & Whinston, 2008).

### **Influyentes y líderes de opinión**

Las empresas deben aprovechar la fuerza de la comunicación de consumidor a consumidor, atrayendo a grupos de usuarios que conecten con la marca y actúen posteriormente a su favor. Este grupo no necesariamente debe ser grande, pero sí influyente (Peters, Chen, Kaplan, Ognibeni & Pauwels, 2013; Risselada, Verhoef & Bijmolt, 2014).

Un trabajo seminal, el modelo de comunicación de flujo de dos pasos, postula que las personas siguen a los líderes de opinión quienes a su vez son influenciados por los medios de comunicación (Katz & Lazarsfeld, 1955).

Desde ese trabajo a la actualidad el fenómeno de la influencia fue abordado desde diferentes enfoques y disciplinas. Los términos influyente, influenciador, líder de opinión, *market maven*, *hub* o *alfa user*, se utilizan a menudo indistintamente para referirse a individuos que tienen en común la característica de ser referentes para su red respecto al la adopción y uso en general de productos.

Dentro de las redes sociales surgen naturalmente figuras que toman especial porque tienen el potencial de influenciar un comportamiento de compra en su red de contactos.

La utilidad para el marketing de identificar influenciadores dentro de una red social es indudable y abarca:

- Mejorar la eficacia de la investigación de mercados.
- Bajar costos y lograr efectos multiplicadores en el muestreo y prueba de productos.
- Publicidad directa (Hawkins, Best, Coney & Carey, 1995).

A estos se pueden agregar:

- Relaciones públicas y eventos: Involucrar a estos actores como motor para generar *word of mouth* positivo, manteniendo bajo control el gasto en estas acciones con el máximo retorno sobre la inversión posible.
- Control de daños: Puede resultar de utilidad operar sobre los actores más influyentes con el objeto de moderar y minimizar los daños a la imagen de la marca en caso de eventualidades con ese potencial.

El accionamiento sobre individuos seleccionados mediante algún criterio particular para lograr un efecto multiplicador se conoce como sembrado o *seeding* y se da a través de la interacción de dos mecanismos: expansión del mercado y aceleración del consumo (Libai, Muller & Peres, 2013). El concepto de *seeding* ha sido abordado en varios trabajos. Como contribuciones que estudian el problema de optimización desde una perspectiva experimental se pueden mencionar a Chen, Wang & Wang (2010); Aral, Muchnik & Sundararajan, (2013); Kempe, Kleinberg & Tardos (2015) y Aghdam & Navimipou (2016). Para determinadas redes se han elaborado métricas derivadas de atributos y actividades propios de los usuarios para estimar, entre otros, indicadores de popularidad e influencia (Grossek & Holotescu, 2009).

### **Análisis de redes sociales**

Las redes sociales online son en sí mismas entidades sociales que funcionan como un agregado de las conductas de sus componentes indivisos. Las redes son analizables principalmente en cuanto a su morfología (Smith, Rainie, Himmelboim & Shneiderman, 2014), y a su vez cada uno de los integrantes de la red puede analizarse a partir de métricas que describen su posición dentro de la estructura de relaciones de la red (Hansen, Schneiderman & Smith, 2011).

Estudios previos sugieren que el análisis de la situación estructural de un actor dentro de una red es un buen indicador de liderazgo de opinión (Van der Merwe & van Heerden, 2009).

En *Marketing and social networks: a criterion for detecting opinion leaders* (Litterio et al., 2017) se propone una metodología para detectar actores influyentes dentro de redes sociales basado en la combinación de las métricas llamadas centralidad de intermediación y de vector propio para clasificar los actores de una red social en cuanto a su nivel de influencia tal se observa en figura 1.



Figura 1. Matriz para detectar potenciales influenciadores dentro de una red social (Litterio et al, 2017).

Esta matriz hace uso de dos medidas globales de la importancia de un nodo en relación al resto de la red:

- Centralidad *eigenvector* o de autovector, es una medida estándar de influencia en una red que pondera las conexiones y la calidad de las conexiones de cada individuo.
- Centralidad de intermediación muestra en qué medida un actor se encuentra en el camino más corto entre otros dos nodos. Puede ser pensado como un “puente” dentro de la red, en que el actor toma una posición de intermediario o *broker*.

La matriz permite mapear los individuos que componen una red y de acuerdo con un criterio relevante para determinar los cortes, clasificarlos en cuatro segmentos con características de difusión diferenciadas.

Los líderes no necesariamente deben estar en el medio de cada red importante ya que esto sería a costa de una posición marginal en otra red. Existe un trade off en la construcción de este capital social (Balkundi & Kilduff, 2006).

**Grafos, redes unimodales, redes bimodales y redes implícitas**

Las redes sociales pueden ser representadas mediante grafos. Un grafo es un grupo de nodos o vértices que representa los individuos y otros elementos que componen la red. Estos a su vez se encuentran unidos por enlaces o aristas que representan las relaciones entre ellos.

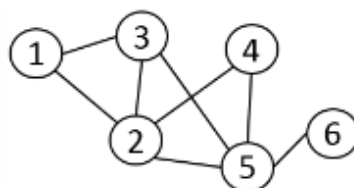


Figura 2. Un grafo con 6 vértices y 8 aristas. Elaboración propia.

El análisis de redes sociales se sirve de la teoría de grafos para analizar la topología de la red social, el modelado estructural, la fuerza de los lazos, la detección de comunidades, cohesión de grupos y cálculo de métricas (Alamsyah & Rahardjo, 2021).

La relación entre dos vértices puede ser enunciada de cualquier forma e incluso puede asignársele un peso o valor. Así, a partir de un grupo de personas podría modelarse una red que enlace personas que son amigas, otra red que enlace personas que comparten gustos de helados y otra que enlace personas según sus preferencias musicales.

Una red bimodal o bipartita es una red que conecta dos variedades de cosas diferentes. Se pueden tomar por ejemplo de red bipartita a los alumnos de una universidad y modelar un enlace entre los alumnos y el profesor de cada materia que están cursando. La red resultante tendría alumnos y profesores. Otros ejemplos habituales de redes bipartitas son las redes de colaboración y las redes de opinión, en las que se conectan personas con contenidos y son ampliamente utilizadas en sistemas de recomendación automática (Zhou, Ren, Medo & Zhang, 2007).

Una red unimodal es una red que conecta una única variedad de entidades entre sí. Siguiendo el ejemplo anterior, la red de alumnos y profesores podría separarse en dos redes unimodales: una que conecta los alumnos que se encuentran cursando alguna materia en común, y otra de profesores que tienen alumnos en común.

El proceso de convertir una red bimodal en una red unimodal se llama proyección bipartita y puede verse en la figura 3.

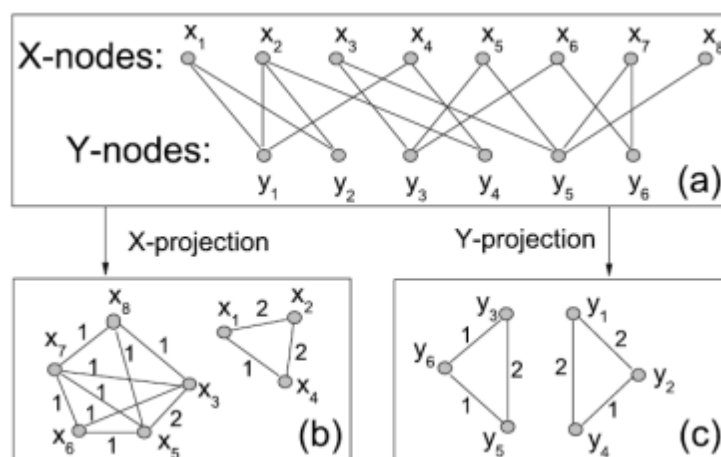


Fig 3. Ilustración de una red bipartita con nodos de tipo X e Y (a) y su proyección X (b) y proyección Y (c), en las que cada enlace toma como peso el número de vecinos de la red original (Zhou, Ren, Medo & Zhang 2007).

Pueden concebirse redes de más modos, pero su utilidad práctica es limitada: La visualización de redes con muchos tipos diferentes de nodos puede ser un ejercicio interesante para la heurística, pero es malo para el análisis cuantitativo debido a que los cálculos de métricas de redes asumen en general que todas las redes son unimodales. Si



bien existen técnicas complejas para analizar redes n-modales, una solución aceptable es generar sus proyecciones unimodales antes de ejecutar cualquier cálculo estadístico estándar sobre ellas.

Hay redes en las que los enlaces o aristas son evidentes ya que se encuentran declarados y definen la estructura de la red y las reglas de la interacción entre los individuos que forman parte de esas comunidades. En *Facebook* por ejemplo existe la relación no direccional de “amistad” y en Twitter o Instagram existen relaciones unidireccionales de “siguiendo”.

Por otra parte existen redes en la que las relaciones entre los individuos no son obvias. Sitios agregadores de opiniones o *reviews* de productos o servicios contienen redes sociales que vinculan a los usuarios con una característica común como compartir afluencia a un mismo local, o consumir un mismo producto. Estas redes se conocen como bimodales y son utilizadas en conjunto con sistemas de recomendación (Gupte & Eliassi-Rad, 2011; Ma, 2013). Su utilidad radica en que agrupan individuos afines pero que no necesariamente tienen que conocerse o haber declarado formalmente algún tipo de relación. Esto las diferencia de las redes sociales como *Facebook*, en la que los vínculos pueden generarse por razones triviales como la reciprocación, cortesía, o presión de pares. Las relaciones en las redes bimodales no necesitan ser declaradas por los usuarios, sino que se basan en una afinidad real.

### **Extracción de datos web**

Existen diferentes métodos para extraer información de redes sociales online y de la web en general. Cuando no se puede bajar la información directamente o no se tiene acceso a una interfaz de aplicaciones (API) que permita extraer la información necesaria una alternativa es hacerlo a través de una técnica llamada *web scraping*. Esta técnica permite emular una consulta hecha por un humano mediante una automatización o *bot* que requiere las páginas a un servidor web, las descarga, y luego las reexpresa en un formato útil.

El *bot* actúa directamente sobre el código fuente de la página y opera siguiendo un conjunto de instrucciones programadas. El método es muy flexible, pero requiere una programación específica para cada sitio web, adaptada a la estructura de las páginas que se quieran explotar.

El proceso de *web scraping* finaliza cuando los datos se encuentran ya estructurados en el formato deseado. El proceso puede visualizarse de la siguiente manera:



Fig 4. Proceso simplificado de web scraping. Elaboración propia.

Es importante aclarar que muchos sitios web disuaden el uso de *web scraping* a través de sus términos de servicio y algunos toman además medidas activas para evitar los *bots*. No obstante, en muchas ocasiones en que se requiere la consulta de grandes cantidades de información el *web scraping* es la única alternativa disponible para recabar los datos que de otra forma sería imposible o prohibitivo en tiempo y costo.

### 3. DESARROLLO DE UN CASO DE ESTUDIO: TRIPADVISOR

Tripadvisor ([www.tripadvisor.com](http://www.tripadvisor.com)) es una plataforma global basada en opiniones generadas por usuarios sobre servicios y productos relacionados con el turismo y gastronomía. En febrero de 2020, el sitio declaró en su *newsroom* 463 millones de visitantes únicos por mes y 859 millones de opiniones acerca de hospedajes, restaurantes, experiencias y transportes (Tripadvisor US Press Center, 2020). La herramienta es muy popular y de gran relevancia para organizaciones relacionadas con el rubro.

El contenido de Tripadvisor es fundamentalmente generado por usuarios del sitio en forma de opiniones o *reviews*. Las opiniones contienen entre otros elementos una valoración de la experiencia a través de una puntuación entre 1 y 5, respondiendo a la siguiente escala semántica: 1: Horrible, 2: Malo, 3: Normal, 4: Muy bueno y 5: Excelente. El agregado de los puntajes obtenidos en todas las opiniones da como resultado la valoración única del local. Tripadvisor ofrece diferentes visualizaciones de conjuntos de locales, conjuntos de opiniones, u opiniones individuales. Los contenidos se exponen ordenados de acuerdo a un algoritmo propietario que toma en cuenta la valoración agregada del local y la cantidad, calidad y fecha de publicación de las opiniones.

Si bien no existen trabajos similares sobre Tripadvisor, un estudio realizado sobre *Yelp!* (sitio agregador de opiniones y competidor directo de Tripadvisor en varias localidades de EE.UU.) concluye que una diferencia de un aumento de una estrella en las reseñas de un local puede significar un aumento en los ingresos de ese local de entre un 5% y 9% (Luca, 2016).

Para este trabajo se definió como caso y objetivo de estudio la red definida por el conjunto de restaurantes de la localidad de Bahía Blanca, Argentina, y los usuarios que publicaron

opiniones de esos locales en Tripadvisor hacia junio de 2019 (Nantes, Litterio & Larrosa, 2019).

La extracción de datos de la web se realizó en forma mayormente automatizada mediante la técnica de *web scraping*, a través de programas o *scripts* programados ad-hoc en lenguaje R. La implementación se hizo en RStudio (<https://www.rstudio.com/>) y RStudio Cloud (<https://rstudio.cloud/>) y se hizo uso de librerías o *packages* adicionales de código abierto que extienden las funcionalidades básicas del lenguaje.

La extracción se organizó en 4 *scripts* de ejecución secuencial. Su contenido, estructura, entradas y salidas de estos se resume en anexo I.

### Análisis preliminares sobre las extracciones

El producto del proceso de *web scraping* fue una base de datos y metadatos de todas las opiniones sobre restaurantes de las ciudades ordenadas en una tabla, limpias y preparadas para ser analizadas. Esta base posee 7019 reseñas sobre 99 locales que poseen al menos una opinión y poseen en promedio 70,9 opiniones cada uno. La distribución de restaurantes en función de las opiniones recibidas se puede ver en el siguiente histograma.

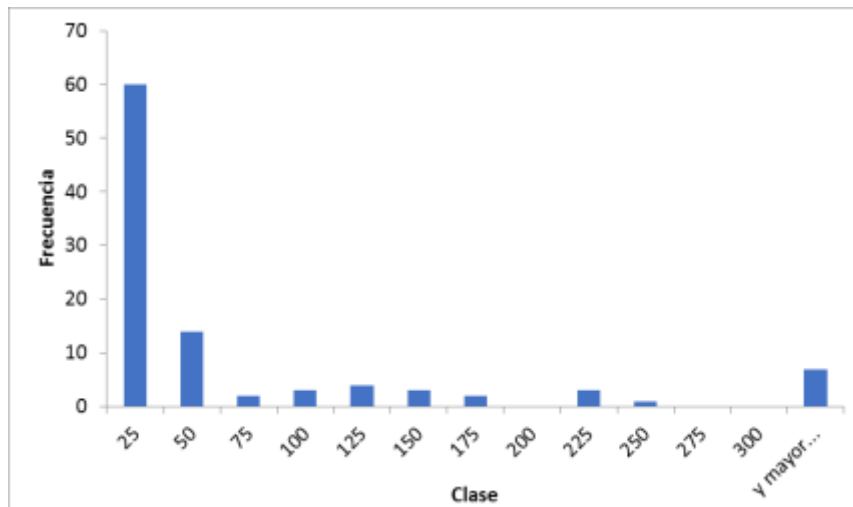


Fig 5. Distribución de restaurantes por cantidad de opiniones recibidas. Elaboración propia.

A su vez se detectaron 2.544 usuarios individuales con un promedio de publicación de 2,76 opiniones cada uno distribuidos como se muestra en figura 6.

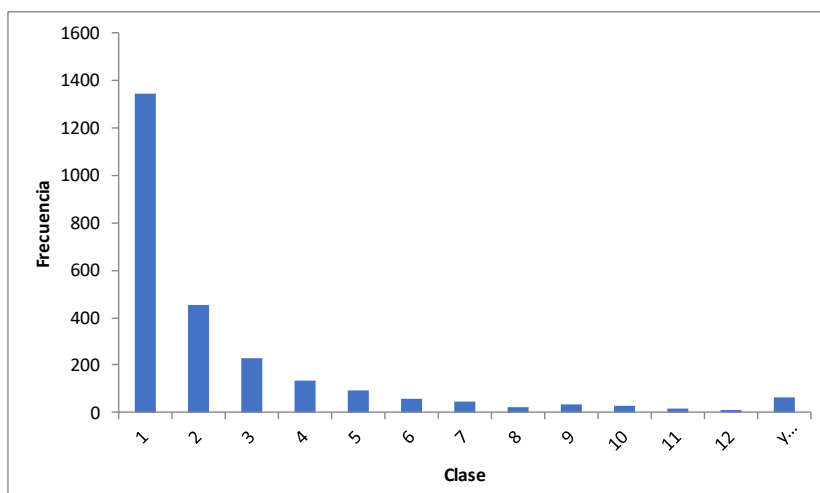


Fig 6. Distribución de usuarios por cantidad de opiniones publicadas. Elaboración propia.

Los ratings o valoraciones individuales de cada una de estas opiniones se concentran en torno al valor 4 (“Muy bueno”) tal como puede verse en figura 7.

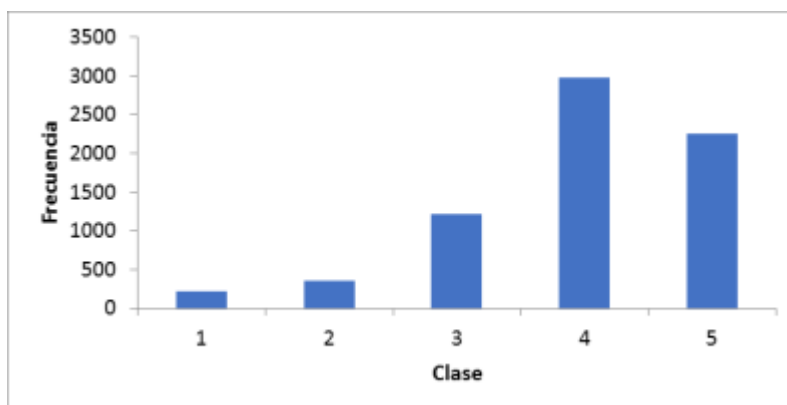


Fig 7. Distribución de ratings por cantidad de opiniones publicadas. Elaboración propia.

### Análisis de la red social: visualización en GEPHI

Para generar las visualizaciones de la red en forma de grafos y el análisis de redes en general se utilizó Gephi (<https://gephi.org/>). Gephi es un software de código abierto para análisis y visualización de redes creado en Java.

La importación de los datos de la base completa se realizó en Gephi a través del módulo de importación.

Una forma habitual para representar redes en una tabla es a través de una lista de enlaces. Se trata de una tabla que debe tener al menos dos columnas (TARGET y SOURCE) donde se listan los nodos entre los que existe un enlace.

Dado que la base de trabajo ya está organizada de esa manera, la importación es directa: cada una de las 7019 opiniones es interpretada como un enlace entre un usuario y un restaurante. En la figura 8 se puede ver una de las pantallas de este proceso.

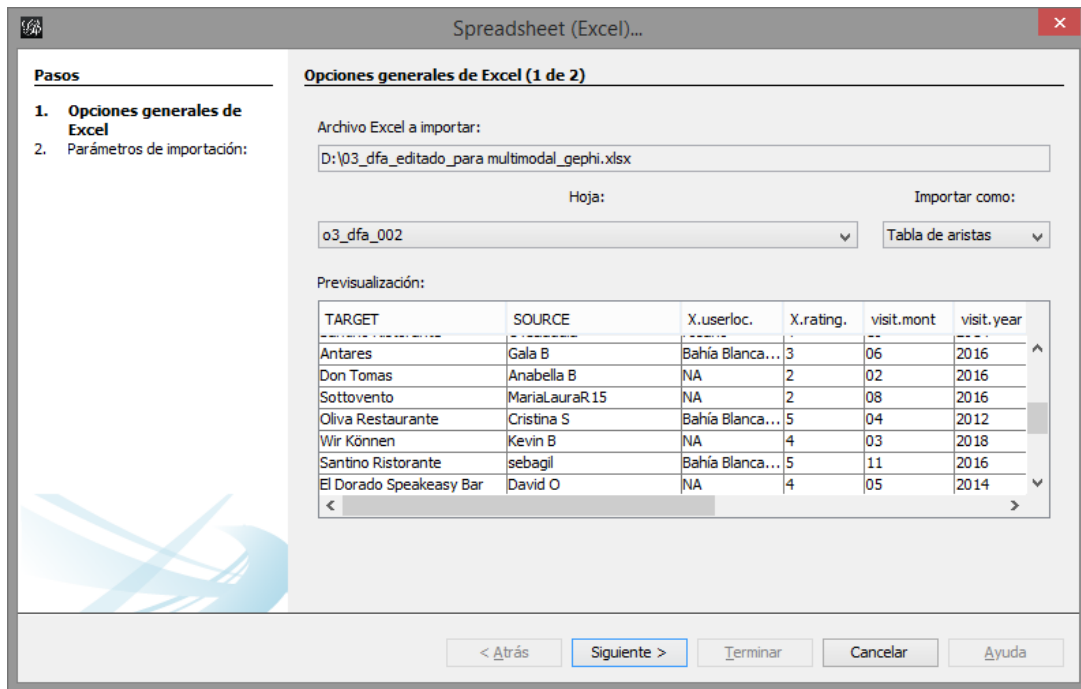
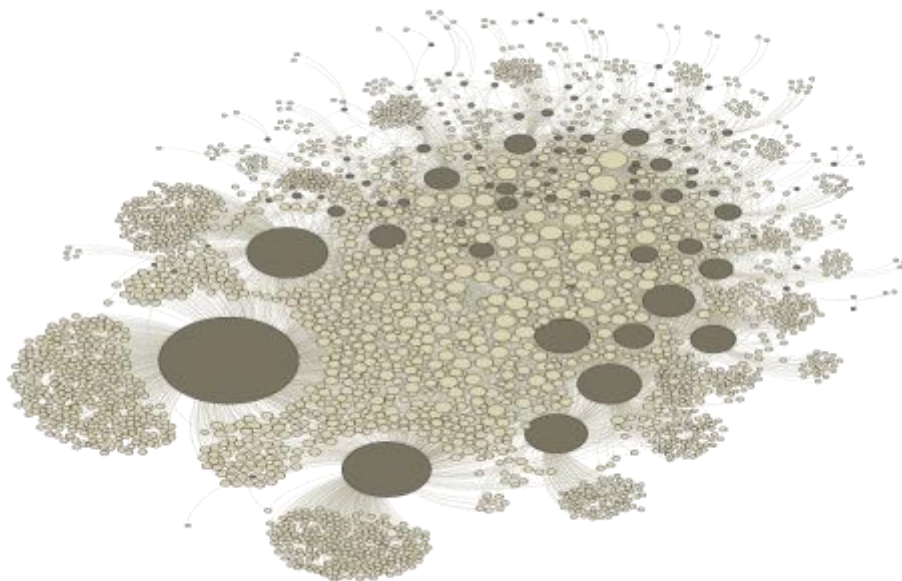


Fig 8. Importación de una hoja de cálculo en Gephi. Elaboración propia.

Al importar una lista de enlaces con esta modalidad Gephi infiere los nodos que conforman la red y genera una tabla de nodos, sobre la cual se puede trabajar para calcular métricas propias de cada individuo o indicar atributos de los nodos que permitan una mejor visualización. En este caso se generó una columna AttributeColumn con los valores “REST” o “PEOPLE” que indica si el vértice es un restaurante o una persona y se utilizó para colorear los vértices de la figura 9. Se calculó además la centralidad de vector propio para asignarle un tamaño a cada uno de los vértices.



*Fig 9. Red social bipartita con restaurantes como nodos oscuros y usuarios como nodos claros. El tamaño de los nodos corresponde a la centralidad de autovector calculada por Gephi sobre el total de la red, como red no dirigida. Arreglados con el algoritmo ForceAtlas2 (Jacomy et al., 2014) Elaboración propia con Gephi.*

## PROYECCIÓN DE LA RED Y APLICACIÓN DEL MODELO DE DETECCIÓN DE INFLUYENTES

La figura anterior sirve para de visualizar y comprender en forma rápida y desde una perspectiva heurística el ecosistema de la gastronomía local, pero antes de aplicar otros tipos de análisis es necesario proyectar la red anterior para obtener una red unimodal.

El modelo detección de influyentes de Litterio et al. (2017) utiliza cálculos de centralidad que asumen una red con entidades de un solo tipo, por lo que se hace necesario obtener una red conformada únicamente por los usuarios presentes en la base (neta de los restaurantes).

Conceptualmente, este procedimiento transforma una red bimodal (que relaciona restaurantes con personas) en una nueva red unimodal que tiene solo personas donde los enlaces se encuentran determinados por la relación “ambos individuos generaron una opinión sobre el mismo local”. Para cada restaurante entonces, se debe generar una red que contiene todos los usuarios que publicaron opiniones unidos entre sí. En el presente trabajo se realizó una proyección simple es decir asignando el mismo peso en todos los enlaces.

Gephi no dispone de una herramienta para hacer esa transformación, por lo que se resolvió en una planilla de Excel mediante programación en Visual Basic for Applications (VBA). La cantidad posible de combinaciones para cada restaurant puede calcularse mediante la siguiente fórmula, donde n es el número de individuos que presentaron una opinión sobre el restaurant y k la cantidad de ítems en cada combinación, 2 en este caso:

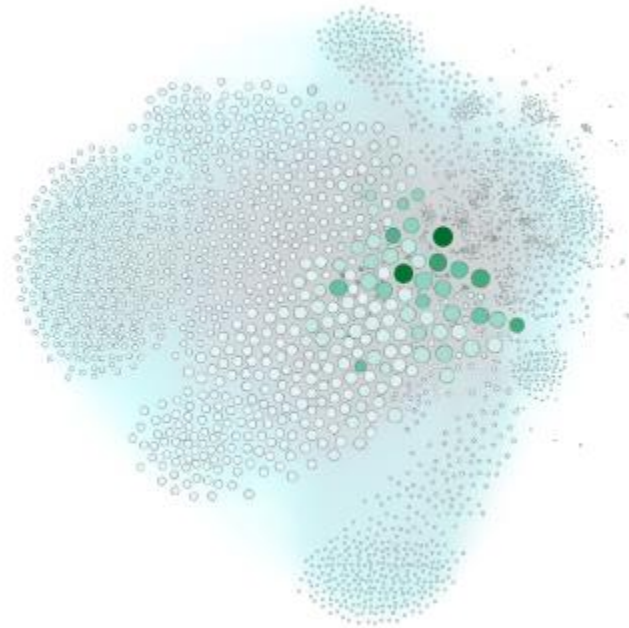
$$\binom{n}{k} = \frac{P_{k,n}}{k!} = \frac{n!}{k!(n-k)!} \quad \text{donde:} \quad P_{k,n} = \frac{n!}{(n-k)!}$$

Por ejemplo, para un restaurante sobre el que opinaron 10 usuarios, la cantidad de posibles combinaciones entre los usuarios es de 45. Para un local con 100 usuarios es 4.950, y para 1000 usuarios es de 499.500. Rápidamente puede verse que el crecimiento es exponencial, de hecho en el presente trabajo las combinaciones posibles antes de eliminar duplicados superan el millón de enlaces.

El diseño del algoritmo utilizado se muestra en anexo II.

Una vez generadas todas las combinaciones posibles se eliminaron los pares duplicados. La tabla de aristas unimodal resultante resultó tener 2.544 usuarios y 926.955 enlaces únicos. Sobre esta tabla se calcularon las centralidades de vector propio y de intermediación. Se muestra en la figura 10 el grafo resultante. Nótese que dada la cantidad

de enlaces que se representan en ese grafo, los mismos no se visualizan como líneas individuales sino como áreas de densidad.



*Fig. 10 Visualización de la red unimodal de usuarios, donde cada punto corresponde a un usuario, el tamaño es proporcional a su centralidad de vector propio y la intensidad del coloreado es proporcional a la centralidad de intermediación, ambas métricas calculadas como red no dirigida. Arreglados con el algoritmo ForceAtlas2 (Jacomy et al., 2014) Elaboración propia con Gephi.*

Sobre esta red se calcularon las centralidades de vector propio y de centralidad de intermediación que sirven para aplicar el modelo de detección de influyentes. En figura 11 puede verse una representación de ambas métricas en un diagrama de dispersión.

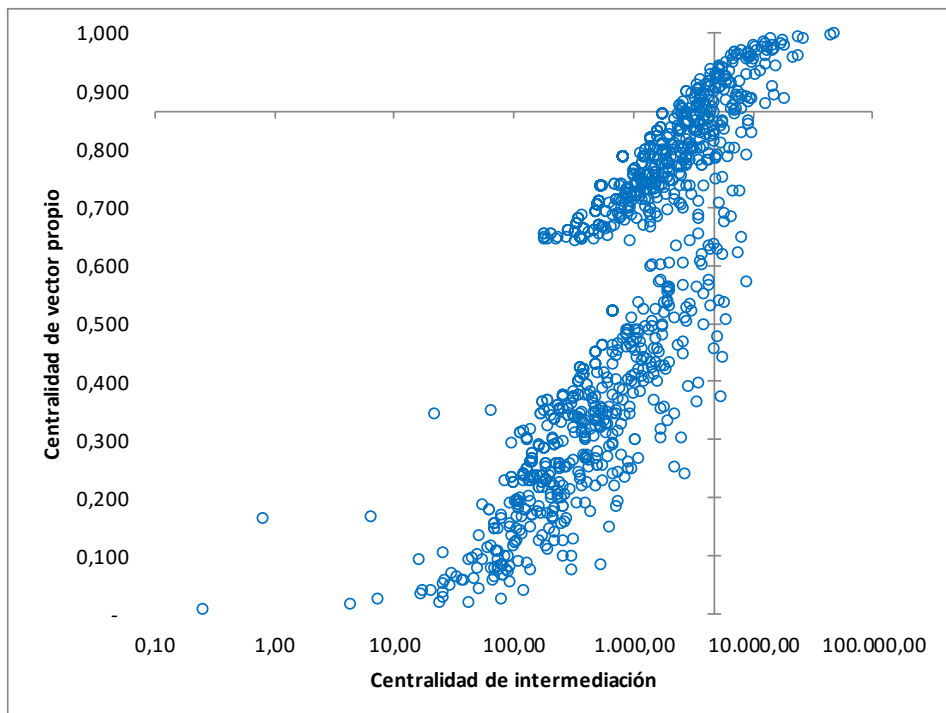


Fig. 11 Diagrama de dispersión que representa los usuarios de la red analizada, donde cada punto corresponde a un usuario, y los ejes corresponden a centralidad de vector propio y centralidad de intermediación. Elaboración propia con Excel.

|                 | Centralidad de intermediación | Centralidad de vector propio |
|-----------------|-------------------------------|------------------------------|
| Media           | 911,42                        | 0,415                        |
| 95th percentile | 4.732,33                      | 0,866                        |
| 75th percentile | 698,34                        | 0,643                        |
| median          | -                             | 0,343                        |
| 25th percentile | -                             | 0,165                        |

Tab. 2 Métricas de la red y definición de criterios de corte para la aplicación del modelo de detección de influyentes en el percentil 95. Elaboración propia en Excel.

La aplicación del modelo con cortes en el percentil 95 para cada una de las métricas determina que 85 usuarios son los seleccionados como potenciales influyentes. Se propone como hipótesis de trabajo que los individuos seleccionados como potenciales influyentes tendrán una mejor performance en relación con el resto cuanto a la cantidad de veces que otros usuarios indicaron que una publicación del usuario es “útil”. Este último dato ya se encuentra disponible en 03\_dfa.

En la tabla 3 pueden verse estadísticas de la variable en cada cuadrante luego de aplicar la segmentación. En principio se ve que los usuarios “puentes” (de alta centralidad de intermediación) y los “influyentes” (de alta centralidad conjunta) tienen una mejor performance que el resto, pero no se diferencian claramente entre sí.

|                         | CONEXIONES IMP | SECUNDARIOS | PUENTES | INFL  |
|-------------------------|----------------|-------------|---------|-------|
| Media                   | 17,02          | 17,78       | 29,05   | 24,68 |
| Mediana                 | 8,5            | 5           | 20,5    | 19    |
| Desviación estándar     | 23,49          | 51,81       | 44,33   | 24,36 |
| Curtosis                | 4,22           | 371,93      | 28,23   | 3,18  |
| Coficiente de asimetría | 2,23           | 15,22       | 4,92    | 1,53  |
| Rango                   | 90             | 1555        | 283     | 118   |
| Mínimo                  | 0              | 0           | 2       | 0     |
| Máximo                  | 90             | 1555        | 285     | 118   |
| Cuenta                  | 42             | 2366        | 42      | 85    |

Tab. 3 Estadísticas de la variable “cantidad de veces que otros usuarios indicaron que una publicación del usuario es “útil” por cuadrante.

En figura 12 se muestra la distribución acumulada de usuarios por segmento, en función de la cantidad de veces que una contribución del usuario se indicó como útil. Las líneas de actores con conexiones importantes y secundarios crecen más rápido lo que indica que esos segmentos tienen más cantidad de usuarios con pocas contribuciones útiles. El 80% de los usuarios tiene 25 votos como máximo.



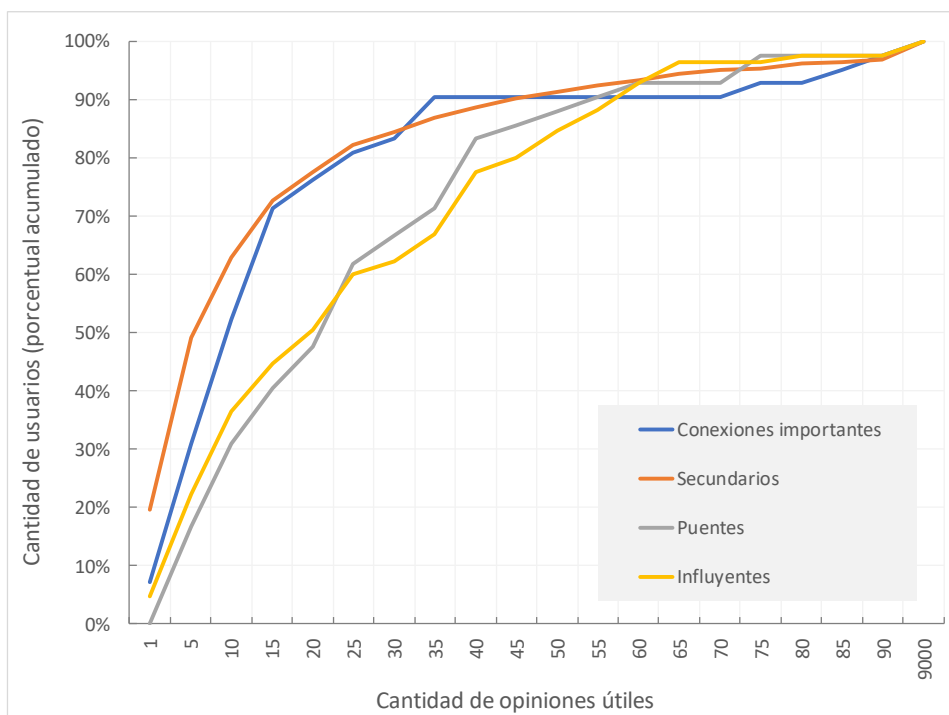


Fig. 12 Distribución acumulada de usuarios por segmento, en función de la cantidad de veces que una contribución del usuario se indicó como útil.

Las líneas de “puentes” e “influyentes” muestran un performance muy similar. Se realizaron varios cruces con las variables disponibles en el set de datos extraído de la web para entender por qué estos dos segmentos no muestran diferencias en su performance, pero ningún análisis fue concluyente. Se ofrecen reflexiones al respecto en las conclusiones del trabajo.

### Minería de texto

De forma de aprovechar todo el cuerpo de datos extraído de la web se propone una aplicación sencilla de minería de texto a través de R. Concretamente se demuestra un análisis de frecuencia de palabras y visualización realizado principalmente con los paquetes “tidytext” (Silge & Robinson, 2016) y “wordcloud2” (Lang & Chien, 2018).

Se trabajó directamente sobre el texto completo de las opiniones, el cual se sometió a las siguientes transformaciones:

- 1.- Limpieza: se eliminaron tildes de palabras, se pasaron todas las palabras a minúsculas, se eliminaron palabras de tipo “stop words” y palabras comunes.
- 2.- *Tokenización*: se derivó una tabla de la original en la que cada registro contiene una palabra individual, o un n-grama, o sea conjunto de n palabras que aparecen en forma seguida en el cuerpo de la opinión.

Luego se analizó frecuencia de palabras o n-gramas para diferentes cortes de la tabla original. En la figura 13 puede verse un el conteo de las palabras más frecuentes para el total de opiniones.



Fig. 13 Conteo de palabras más frecuentes en el total de opiniones.  
Elaboración propia con R.

El análisis de frecuencia de palabras realmente cobra más sentido cuando se analizan diferentes cortes. Por ejemplo: ¿Qué es lo que hace a una buena opinión? ¿Cuáles son los términos más relevantes utilizados en las opiniones de 4 y 5 estrellas?



Fig. 14 Nube de palabras que representa la frecuencia de palabras en opiniones de restaurantes con calificación 4/5 y 5/5. Elaboración propia con R.



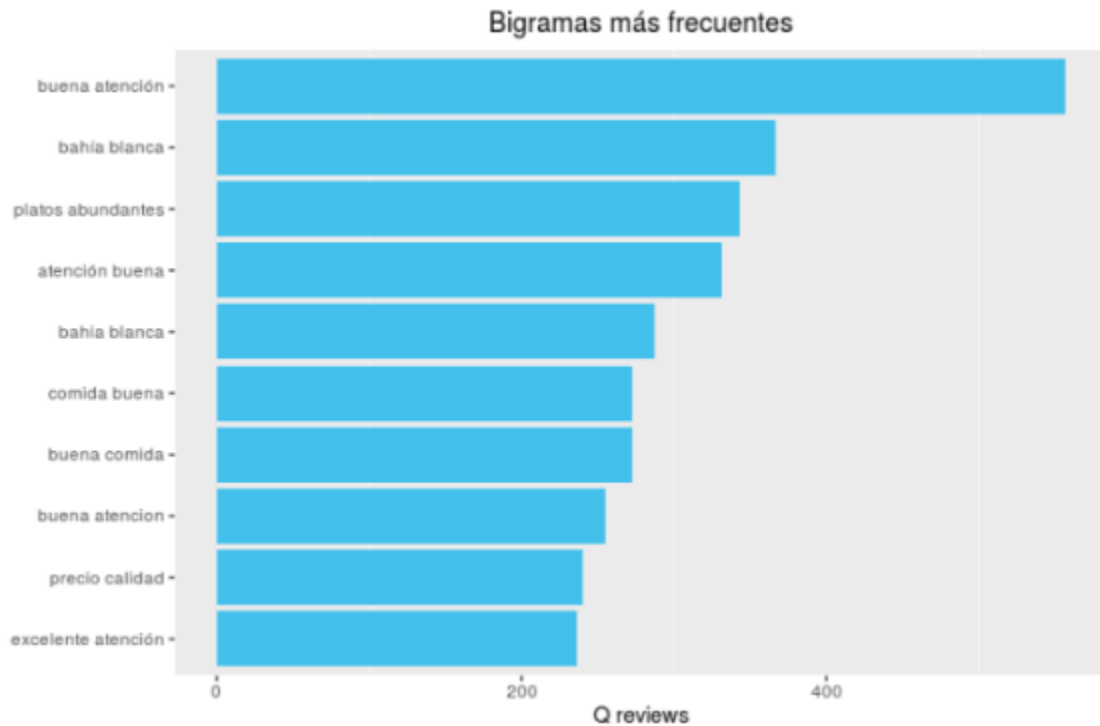


Fig. 16 Frecuencia de bigramas en el total de opiniones. Elaboración propia con R.

Contra intuitivamente, puede concluirse entonces a partir del análisis de opiniones que la calidad de la comida es importante, pero se da por sentada y no constituye el factor diferenciador que determina una opinión excelente. La comida determina el *baseline* (nunca se debe descuidar) y el lugar es importante (tiene gran relevancia en las buenas y malas opiniones), pero la calidad de la atención es lo que convierte a un restaurante en excelente. El precio pasa a un segundo plano si el cliente es bien atendido.

#### 4. CONCLUSIONES

En el transcurso del trabajo se introdujeron varios conceptos necesarios para entender el alcance del problema. Se dio un marco a los mecanismos de difusión y adopción de productos y servicios. Se expuso el rol de los influyentes en ese proceso. Se introdujo la teoría de redes sociales y conceptos sobre su análisis y explotación. Se describieron varias tareas necesarias para preparar la información y aplicar técnicas de análisis de redes sociales, de análisis estadístico y de minería de texto.

Finalmente se utilizó la técnica de *web scraping* para descubrir y analizar una red social de un sitio popular de reseñas de lugares y de servicios turísticos.

El proceso no se encontró exento de dificultades: Como se comentó anteriormente, la técnica de *web scraping* requiere programación específica; esto lo hace altamente flexible pero también altamente falible y requiere ajustes y adaptaciones frecuentes. Se experimentó además la ejecución de los scripts desde una instalación local de RStudio y

desde la nube (RStudio Cloud). No obstante esto, la red finalmente se pudo extraer, visualizar y proyectar. Los resultados de la implementación del modelo de detección de influyentes propuesto por Litterio et al. no resultaron a priori los esperados. Es muy probable que la razón sea que ese modelo no es 100% compatible con una red de este tipo ya que fue diseñado para analizar redes unimodales, en las que el vínculo entre individuos permite difusión. Por ejemplo, si A es amigo de B, es probable que si A comparte un excelente comentario sobre su experiencia con el producto X, B se muestre más interesado en consumir ese producto que si no existiera la relación. En el caso de una red bimodal como la analizada en este trabajo, el vínculo entre dos usuarios se define por la afluencia de ambos a un mismo local. No es necesario que se conozcan ni si quiera que asistan el mismo día. En ese sentido es difícil imaginar un proceso de difusión.

Otra razón puede ser la elección de una métrica inadecuada para hacer el análisis. La cantidad de veces que otros usuarios indicaron que una publicación del usuario es "útil" puede ser una métrica engañosa, porque esos votos positivos son por cualquier publicación que haya hecho el usuario en su historia, en la red analizada o fuera de ella. Si un usuario escribió una opinión sobre un hotel en Londres que recibió 1.000 votos positivos y luego ingresó a la red de restaurantes de Bahía Blanca por haber opinado sobre algún local de la ciudad, ese usuario "trae" a esta red los 1.000 votos positivos que generó en otra parte.

Aun no habiendo podido demostrar una relación entre el modelo para detectar influyentes y esta red bimodal, es lícito decir que se descubrieron herramientas potencialmente útiles para empresas del rubro de servicios turísticos: Si bien no se llegó a plantear una metodología de trabajo propiamente dicha, se demostró que el conjunto de herramientas utilizadas incluyendo varias herramientas de análisis cuantitativo (y se ganaron algunos *insights* en el camino), que pueden ser fácilmente replicables, mejorables y sin duda útiles para empresarios del sector.

Tal como se mencionó al principio del trabajo, en la actualidad es clave que las organizaciones entiendan el marketing digital y lo incorporen a su caja de herramientas para abordar el mercado.

Este trabajo introdujo varios conceptos y técnicas aplicados a un problema de marketing contemporáneo y su resultado es el descubrimiento de información valiosa no evidente desde otros métodos de análisis. Creemos y esperamos que estos análisis puedan ser replicados por las organizaciones y aplicados en plataformas web y problemas de negocio diferentes para lograr un mejor entendimiento del mercado.

## REFERENCIAS BIBLIOGRÁFICAS

- Alamsyah, A. & Rahardjo, B. (2021) Social Network Analysis Taxonomy Based on Graph Representation. arXiv preprint arXiv:2102.08888.
- Aghdam S. M. & Navimipou N. J. (2016) *Opinion leaders selection in the social networks based on trust relationships propagation*. Karbala International Journal of Modern Science 2, 2016 88-97.  
DOI: 10.1016/j.kijoms.2016.02.002
- Aral S., Muchnik L. & Sundararajan A. (2013) *Engineering social contagions: Optimal network seeding in the presence of homophily*. Network Science, 1, pp 125-153  
DOI: 10.1017/nws.2013.6
- Awad N. & Ragowsky A. (2008) *Establishing Trust in Electronic Commerce through Online Word of Mouth: An Examination across Genders*. Journal of Management Information Systems, Vol. 24, No. 4, pp. 101-121  
Stable URL: <http://www.jstor.org/stable/40398913>
- Bacile T., Ye C. & Swilley E. (2014) From firm-controlled to consumer-contributed: consumer co-production of personal media marketing communication. Journal of Interactive Marketing Volume 28, Issue 2, May 2014, Pages 117–133  
DOI: 10.1016/j.intmar.2013.12.001
- Balkundi, P. & Kilduff, M. (2006) *The ties that lead: A social approach to leadership*. The Leadership Quarterly 17 (2006) 419-439  
DOI: 10.1016/j.leaqua.2005.09.004
- Bastian M., Heymann S., Jacomy M. (2009). *Gephi: an open source software for exploring and manipulating networks*. International AAAI Conference on Weblogs and Social Media.
- Benedetti, A. (2015) Marketing en redes sociales: Detrás de escena. Buenos Aires: Ed AMDIA
- Bickart B. & Schindler R. (2001) *Internet forums as influential sources of consumer information*. Journal of Interactive Marketing Volume 15, Issue 3 Pages 31–40  
DOI: 10.1002/dir.1014
- Chen W., Wang C. & Wang C. (2010) *Scalable Influence Maximization for Prevalent Viral Marketing in Large-Scale Social Networks*. Proceedings KDD'10.  
DOI: 10.1145/1835804.1835934
- Chen X., van der Lans R. & Phan T. Q. (2017) *Uncovering the Importance of Relationship Characteristics in Social Networks: Implications for Seeding Strategies*. Journal of Marketing Research: April 2017, Vol. 54, No. 2, pp. 187-201.  
DOI: 10.1509/jmr.12.0511

- Crawley M. (2012) *The R book*. 2nd ed. Wiley, 2012. Recuperado de <https://www.cs.upc.edu/~robert/teaching/estadistica/TheRBook.pdf>
- Duan, W., Gu, B., & Whinston, A.B. (2008). *Do online reviews matter?—An empirical investigation of panel data*. *Decision Support Systems*, 45(3), 1007–1016  
DOI: 10.1016/j.dss.2008.04.001
- Ekran I. & Evans C. (2016) *The influence of eWOM in social media on consumers' purchase intentions: An extended approach to information adoption*. *Computers in Human Behavior*. 61, pp 47-55.  
DOI: 10.1016/j.chb.2016.03.003
- Gauri D., Bhatnagar A., & Rao R. (2008) *Role of word of mouth in online store loyalty*. *Communications of the ACM Vol 51 Issue 3* pp 89-91 .  
DOI: 10.1145/1325555.1325572
- Grossek G., & Holotescu C. (2009) *Indicators for the analysis of learning and practice communities from the perspective of microblogging as a provocative sociolect in virtual space*. The 5th International Scientific Conference eLSE - eLearning and Software for Education, BUCHAREST, April 09-10, 2009
- Gupte, M. & Eliassi-Rad, T. (2011). *Measuring Tie Strength in Implicit Social Networks*. *Proceedings of the 3rd Annual ACM Web Science Conference, WebSci'12*.  
DOI: 10.1145/2380718.2380734.
- Hansen D., Shneiderman B., & Smith M. (2011). *Analyzing Social Media Networks with NodeXL: Insights from a Connected World*. Ed. Morgan Kauffman.
- Hewett K., Rand W., Rust R. & van Heerde H. (2016) *Brand Buzz in the Echoverse*. *Journal of Marketing*: May 2016, Vol. 80, No. 3, pp. 1-24.  
DOI: 10.1509/jm.15.0033
- Hawkins D., Best R., Coney K. & Carey K. (1995) *Consumer behavior: Implications for marketing strategy*. McGraw-Hill
- Hudson S., Huang L., Roth M. S. & Madden T. (2015) *The influence of social media interactions on consumer–brand relationships: A three-country study of brand perceptions and marketing behaviors*. *International Journal of Research in Marketing* Volume 33, Issue 1, March 2016, Pages 27-41  
DOI: 10.1016/j.ijresmar.2015.06.004
- Jacomy M., Venturini T., Heymann S. & Bastian M. (2014) *ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software*. *PLoS ONE* 9(6): e98679.  
DOI: 10.1371/journal.pone.0098679
- Katz, E. & Lazarsfeld, P.F. (1955) *Personal influence: The part played by people in the flow of mass communications*, New York: The Free Press.

- Kempe D., Kleinberg J. & Tardos E. (2015) *Maximizing the Spread of Influence through a Social Network*. Theory of Computing Journal Volume 11, Article 4 pp. 105-147.  
DOI: 10.4086/toc.2015.v011a004
- King R., Racherla P. & Bush, V (2014) *What We Know and Don't Know About Online Word-of-Mouth: A Review and Synthesis of the Literature*, Journal of Interactive Marketing  
DOI: 10.1016/j.intmar.2014.02.001
- Lang B. & Hyde K.(2013). *Word of mouth: what we know and what we have yet to learn*. Journal of consumer satisfaction, dissatisfaction and complaining behavior 26: 1–18.
- Lang D. & Chien G. (2018) *wordcloud2: Create Word Cloud by 'htmlwidget'*. R package version 0.2.1. <https://github.com/lchiffon/wordcloud2>
- Libai B., Muller E. & Peres R. (2013) *Decomposing the Value of Word-of-Mouth Seeding Programs: Acceleration Versus Expansion*. Journal of Marketing Research: April 2013, Vol. 50, No. 2, pp. 161-176.  
DOI: 10.1509/jmr.11.0305
- Litterio A. M., Nantes, E. A., Larrosa, J. M. & Gómez, L. J. (2017) *Marketing and social networks: a criterion for detecting opinion leaders*. European Journal of Management and Business Economics, Vol. 26 Issue: 3, pp.347-366.  
DOI: 10.1108/EJMBE-10-2017-020
- Luca, M. (2016) *Reviews, Reputation, and Revenue: The Case of Yelp.com* Harvard Business School, Working Paper 12-016
- Ma, H. (2013) *An experimental study on implicit social recommendation*. SIGIR '13 Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval. Pages 73-82  
DOI: 10.1145/2484028.2484059
- Nam H., & Kannan P. (2014) *The Informational Value of Social Tagging Networks*. Journal of Marketing: July 2014, Vol. 78, No. 4, pp. 21-40.  
DOI: 10.1509/jm.12.0151
- Nantes, E. A., Litterio, M., Larrosa, J. M. (2019). *Explotación y detección de influyentes en redes sociales online implícitas*. XXXIII Encuentro de Docentes Universitarios de Comercialización de Argentina y América Latina. 3 y 4 de octubre. Disponible en: <https://repositoriodigital.uns.edu.ar/handle/123456789/6047>
- OmaymaS (2019) *Webscraping Tripadvisor data: Cologne\_Restaurants.R*. GitHub repository, recuperado de [https://github.com/OmaymaS/Web-Scraping-TripAdvisor-Data/blob/master/Cologne\\_Restaurants.R](https://github.com/OmaymaS/Web-Scraping-TripAdvisor-Data/blob/master/Cologne_Restaurants.R). Ren junio de 2019.
- Pavlou P. & Ba S. (2002) *Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premium and Buyer Behavior*. MIS Quarterly 26, 3, pp 243-268.  
DOI: 10.2307/4132332



- Serrano Puche J. (2016) *Internet y emociones. Nuevas tendencias en un campo de investigación emergente*. Comunicar: Revista científica iberoamericana de comunicación y educación, Nº 46, pp 19-26.  
DOI: 10.3916/C46-2016-02
- Silge J. & Robinson D. (2016) *tidytext: Text Mining and Analysis Using Tidy Data Principles in R*. Journal of Open Source Software, 1(3), 37  
DOI 10.21105/joss.00037
- Smith M., L. Rainie, I. Himelboim & B. Shneiderman (2014), *Mapping Twitter Topic Networks: From Polarized Crowds to Community Clusters*. Pew Research Center.
- Peters K., Chen Y., Kaplan A., Ognibeni B. & Pauwels K. (2013) *Social media metrics - A framework and guidelines for managing social media*. Journal of Interactive Marketing Vol 27 pp 281-298  
DOI: 10.1016/j.intmar.2013.09.007
- R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Versión 3.6. URL <https://www.R-project.org/>.
- RStudio Team (2018). *RStudio: Integrated Development for R*. RStudio, Inc., Boston, MA  
URL <http://www.rstudio.com/>.
- Simon H. A. (1982) *Models of bounded rationality: Empirically grounded economic reason* (Vol. 3). MIT press.
- Tripadvisor US Press Center (2020) Recuperado de <https://tripadvisor.mediaroom.com/> el 25/02/2020.
- Van der Merwe R. & van Heerden G. (2009) *Finding and utilizing opinion leaders: Social networks and the power of relationships*. South African Journal of Business Management, Vol. 40, pp 65-76.
- VanMeter R., Grisaffe D. & Chonko L. (2015) *Of “Likes” and “Pins”: The Effects of Consumers’ Attachment to Social Media*. Journal of Interactive Marketing 32, pp 70-88.  
DOI: 10.1016/j.intmar.2015.09.001
- Wang Z. & Gon Kim H. (2017) *Can Social Media Marketing Improve Customer Relationship Capabilities and Firm Performance? Dynamic Capability Perspective*. Journal of Interactive Marketing. Volume 39, August 2017, Pages 15–26  
DOI: 10.1016/j.intmar.2017.02.004
- Wickham H. (2019) *rvest: Easily Harvest (Scrape) Web Pages*. <http://rvest.tidyverse.org/>, <https://github.com/tidyverse/rvest>.

Zhou T., Ren J., Medo M. & Zhang Y. (2007) *Bipartite network projection and personal recommendation*. Physical review. E, Statistical, nonlinear, and soft matter physics. 76. 046115.

DOI: 10.1103/PhysRevE.76.046115.

## 1. ANEXO I

### Scripts utilizados para la extracción de datos

#### 1. **proyectofinal00a.R**

Instala y carga las librerías necesarias para extender las funciones básicas de R, fundamentalmente el paquete rvest creado por Hadley Wickham (2019).

```
install.packages("rvest")
install.packages("xml2")
install.packages("dplyr")
install.packages("magrittr")
install.packages("ggplot2")
install.packages("lubridate")
install.packages("tibble")
install.packages("purrr")
install.packages("tidytext") #Text mining
install.packages("tidyr") #Spread, separate, unite, text mining (also included in the
tidyverse package)
install.packages("widyr") #Use for pairwise correlation
install.packages("wordcloud2") #nube de palabras para minería de texto
```

#### 2. **proyectofinal001.R**

Tiene como objeto generar un listado con los nombres, URLs (de Tripadvisor), ratings promedio y cantidad de opiniones recibidas de cada restaurant de la ciudad barriendo todos los restaurantes de la página principal de restaurantes de Bahía Blanca<sup>1</sup> y las n paginaciones que le siguen. Tripadvisor ofrece datos de 30 restaurantes por página en esta vista. En este caso se barrieron 4 páginas y se obtuvieron los datos de 107 restaurantes. El código de este script toma algunos elementos del script publicado por OmaymaS (2019).

OUTPUT: La salida de esta etapa es una tabla 01\_dfa en formato de dataframe con los datos indicados.

#### 3. **proyectofinal002.R**

Toma la tabla 01\_dfa generada en el paso anterior y barre secuencialmente todos los URLs de cada uno de los restaurantes, recuperando para cada caso las URLs de cada opinión individual del URL principal y paginaciones. Tripadvisor ofrece 10 *reviews* por página. Para el primer restaurante del listado, con 931 opiniones al momento de la extracción, se barrieron 94 páginas en total.

---

<sup>1</sup> [https://www.tripadvisor.com.ar/Restaurants-g312744-Bahia\\_Blanca\\_Province\\_of\\_Buenos\\_Aires\\_Central\\_Argentina.html](https://www.tripadvisor.com.ar/Restaurants-g312744-Bahia_Blanca_Province_of_Buenos_Aires_Central_Argentina.html)

Este es el más complejo de los scripts utilizados ya que involucra una rutina para barrer los URLs de cada uno de los restaurantes y a su vez una rutina anidada que barre hasta agotar todas las opiniones que existen para cada uno de ellos.

OUTPUT: Genera una tabla 02\_dfa en formato de dataframe que contiene las URLs de cada opinión individual. Se obtuvieron en total 7019 URLs.

#### 4. proyectofinal003.R

Toma la tabla 02\_dfa y barre secuencialmente todos los URLs de cada una de las opiniones, generando una nueva tabla que contiene para cada registro: (a) Nombre del restaurante, (b) Nombre del usuario o reviewer, (c) Localización declarada por el reviewer, (d) Rating asignado al restaurant (entre 1 y 5 estrellas), (e) Día de la visita, (f) Opiniones totales generadas por el usuario, (g) Cantidad de veces que otros usuarios indicaron que una publicación del usuario es "útil", (h) Título del review, (i) Texto completo del review.

OUTPUT: Genera una tabla 03\_dfa en formato de dataframe que contiene toda la información anterior.

## 2. ANEXO II

### Scripts para el llenado de columnas – proyección unimodal

En "Columna 1" y "Columna 2" de tabla 4 se pueden ver la combinaciones para una tabla modelo de n=4 usuarios. De las columnas adicionales se puede inferir la lógica utilizada en el algoritmo

| N° usuario | Q repeticiones | Columna 1 | Columna 2 | N° usuario | Q repeticiones |
|------------|----------------|-----------|-----------|------------|----------------|
| 1          | 3              | A         | B         | 2          | 1              |
|            |                | A         | C         | 3          |                |
|            |                | A         | D         | 4          |                |
| 2          | 2              | B         | C         | 3          |                |
|            |                | B         | D         | 4          |                |
| 3          | 1              | C         | D         | 4          |                |

Tab 4. Ejemplo de tabla de enlaces para una red de 4 vértices totalmente conectada, donde se verifica un total de 6 combinaciones posibles y se puede observar la lógica con la que se completa cada columna.

Habiendo observado que todas las tablas en general y cada columna en particular se ajustaban a una lógica común, se programaron 2 rutinas independientes, una para completar cada columna. Ambos toman como input el listado de n usuarios que publicaron una opinión en un local y cada uno aplica la lógica que se detalla seguidamente:

#### 1era columna:

Pega nombre de 1er usuario y avanza hasta la siguiente posición (n-1) veces  
 Pega nombre de 2do usuario y avanza hasta la siguiente posición (n-2) veces  
 Pega nombre de 3er usuario y avanza hasta la siguiente posición (n-3) veces

...

Pega nombre del usuario (n-1) 1 vez

Una vez que finaliza con ese local sigue con el próximo set de usuarios hasta agotar los locales.

**2da columna:**

Pega nombre de 2do usuario 1 vez y avanza a la siguiente posición

Pega nombre de 3er usuario 1 vez y avanza a la siguiente posición

Pega nombre de 4to usuario 1 vez y avanza a la siguiente posición

...

Pega nombre del n usuario 1 vez y avanza a la siguiente posición, donde vuelve a comenzar con toda la escala corrida +1:

Pega nombre de 3er usuario 1 vez y avanza a la siguiente posición

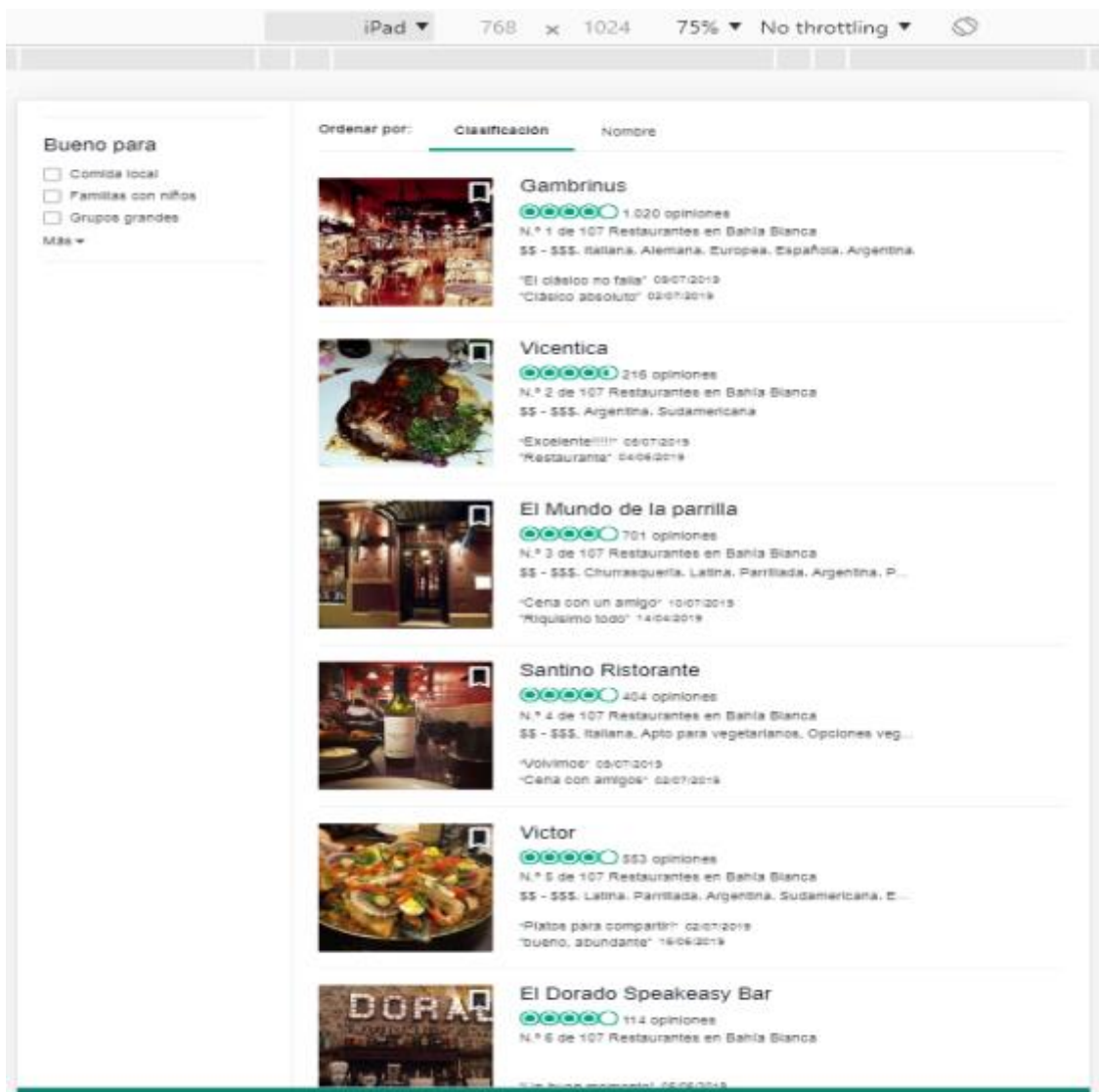
Pega nombre de 4to usuario 1 vez y avanza a la siguiente posición

...

Pegar nombre del n usuario 1 vez y avanzar a la siguiente posición

Cuando se llega a la última posición en la que el único dato que se pega es n ya que no se puede seguir con el loop, se avanza con el siguiente restaurante.


**3. APENDICES**



*Tripadvisor, vista de listado de restaurantes por localidad. Primera página para localidad de Bahía Blanca capturada en julio de 2019.*

iPad 768 x 1024 75% No throttling

Fotos Descripción general **Opiniones** Preguntas y respuestas




**5** Opiniones Escibió una opinión el 19 de marzo de 2019 mediante dispositivo móvil

### Típico restaurant Aleman

Desde hace muchísimos años donde comer el mejor para con papas o snakus con chucrut y la mejor cerveza tirada que no fuera el Gambrinus era imposible. Muchos años lo manejaron algunos ex tripulantes del buque Grand Spee hundido en nuestras aguas y actualmente sigue... [Más](#)

Fecha de la visita: enero de 2019

Agradecece a miguelangelarranz




**5** Opiniones Escibió una opinión el 19 de marzo de 2019 mediante dispositivo móvil

### No es lo mismo

Las personas que han conocido y han degustado sus platos en épocas pasadas sabrán de que hablo. Hoy se darán cuenta que no es lo mismo.

Fecha de la visita: marzo de 2019

Agradecece a Andrea P




**5** Opiniones Escibió una opinión el 2 de marzo de 2019

### Normal

Despues de mucho tiempo volvimos. La comida bien. La atencion normal. Precios normales. Lugar muy ruidoso. Sillas incomodas. La atencion del cajero deja mucho que desear. Sobrevalorado para mi humilde opinion. No vuelvo.

Fecha de la visita: marzo de 2019

Agradecece a fionapettarelli




**5** Opiniones Escibió una opinión el 19 de febrero de 2019

### Muy buena comida y excelente atención..

Exquisita comida, muy bien elaborada: acompañada por buenos vinos mendocinos y la cordialidad y atención de mozo, muy canchero , atento y amable , mostrando que tienen muchos años de buena atención.-

Fecha de la visita: febrero de 2019

Agradecece a daniel0mar



**5** Opiniones Escibió una opinión el 13 de febrero de 2019 mediante dispositivo móvil

### Comida de Fonda, abundante.

Lugar recomendable, se destaca la atención de los mozos, muy predispuestos y la agilidad en los tiempos de entrega de platos. Buena atención, buen bodegón con platos abundantes y de calidad pero sin preparaciones complejas. Precios razonables.

Tripadvisor, vista de listado de opiniones por restaurant. Capturada en julio de 2019.

iPad 768 x 1024 75% No throttling

tripadvisor ARGENTINA Bahía Blanca

Publicar Viajes Buzón Buscar

Bahía Blanca Hoteles Qué hacer **Restaurantes** Vuelos Alquileres temporarios Alquiler de autos

América del Sur > Argentina > Argentina central > Provincia de Buenos Aires > Bahía Blanca > Restaurantes de Bahía Blanca > Gambrinus Comida de Fonda, abundante. - Gambrinus

## “Comida de Fonda, abundante.”

Opinión sobre el Gambrinus

103 fotos

**Gambrinus**  
 Pasaje Arribeños 174, Bahía Blanca B8000LZB, Argentina  
 +54 291 456-2750 Página web Mejorar este perfil

Clasificado como N.º 4 de 138 Restaurantes en Bahía Blanca  
 1.020 opiniones  
 Certificado de Excelencia

madedefe25  
San Antonio de Areco, Argentina  
22 likes

Opinión escrita 13 de febrero de 2019 mediante dispositivo móvil

**Comida de Fonda, abundante.**

Lugar recomendable, se destaca la atención de los mozos, muy predispuestos y la agilidad en los tiempos de entrega de platos. Buena atención, buen bodegón con platos abundantes y de calidad pero sin preparaciones complejas. Precios razonables.

Fecha de la visita: febrero de 2019

Pregunté a madedefe25 sobre Gambrinus

Agradécele a madedefe25

Esta es la opinión subjetiva de un miembro de TripAdvisor, no de TripAdvisor LLC.

Los viajeros están perdiendo la cabeza por...

- Hotel Land Plaza Bahía Blanca 450 opiniones
- Hotel Muniz 141 opiniones
- Hotel Austral Bahía Blanca 224 opiniones

Todos los hoteles en Bahía Blanca (9)

Hacer una pregunta

Obtén respuestas rápidas del personal y los visitantes anteriores de Gambrinus.

Hola, Esteban. ¿Qué te gustaría saber sobre este restaurante?

Recibir notificaciones de nuevas respuestas a tus preguntas.

Preguntar Directrices de publicación

Opiniones (1.020)

Calificación de viajeros

|           |     |
|-----------|-----|
| Excelente | 397 |
| Muy bueno | 444 |
| Normal    | 103 |
| Malo      | 17  |
| Horrible  | 12  |

Tipo de viajero

|           |         |
|-----------|---------|
| Familias  | Mar-may |
| Pareja    | Jun-ago |
| Solitario | Sep-nov |
| Ejecutivo | Dic-feb |
| Amigos    |         |

Epoca del año

Idioma

- Todos los idiomas
- español (973)
- Inglés (28)
- portugués (11)

Más idiomas

Consulta qué dicen los viajeros:

Buscar opiniones

Qué hay cerca

Restaurantes (132) Hoteles (2) Actividades (47)

Tripadvisor, vista de opinión individual. Capturada en julio de 2019.