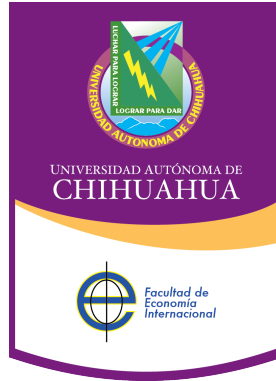


**UNIVERSIDAD AUTÓNOMA DE CHIHUAHUA**

**Facultad de Economía Internacional**



**“Análisis de los determinantes de la adquisición de cuentas en redes sociales: el caso de Twitter”**

Tesis que para obtener el grado de

**Maestro en Economía Empresarial**

Presenta:

**Ing. Jorge Fabián Contreras Navarro**

Director:

**Dr. David de Jesús González Milán**

Octubre 2014

# Índice

Introducción.....	3
Marco Contextual.....	5
Marco Teórico.....	10
Metodología y Resultados.....	13
Conclusiones.....	17
Bibliografía.....	19

## Introducción

En el contexto de crecimiento, un producto con un perfil en una red social busca incrementar la cantidad de seguidores que tiene, pues un perfil representa un canal de distribución para su producto. Chen (2012) apunta que dos productos idénticos pueden tener dos resultados muy diferentes, sólomente basados en qué tan bien se integran en plataformas como Craigslist, Twitter y Facebook. Así mismo Johns menciona la importancia de aumentar los suscriptores a un canal como parte del crecimiento: “¿Cómo incremento la tasa de adquisición, por ejemplo, tener más suscriptores?” (2012).

Esta búsqueda de adquisición de suscriptores es descrita por el embudo de conversión del consumidor, donde McClure (2007) explica que se comienza por la 'adquisición', que son los canales para hacer llegar el producto al usuario y que éste provee una manera de seguir llegando a él, seguido de la 'activación', cuando el usuario ha tenido una experiencia satisfactoria, para después generar renta y referencia a otros clientes potenciales. Palihapitiya (2013) redacta preguntas claves para el crecimiento, basado en el modelo del embudo de conversión, en ¿Cómo llevas gente a la puerta? ¿Cómo los llevas a un momento 'ahá' tan rápido como sea posible? ¿Cómo entregas valor de producto tan frecuentemente como sea posible?.

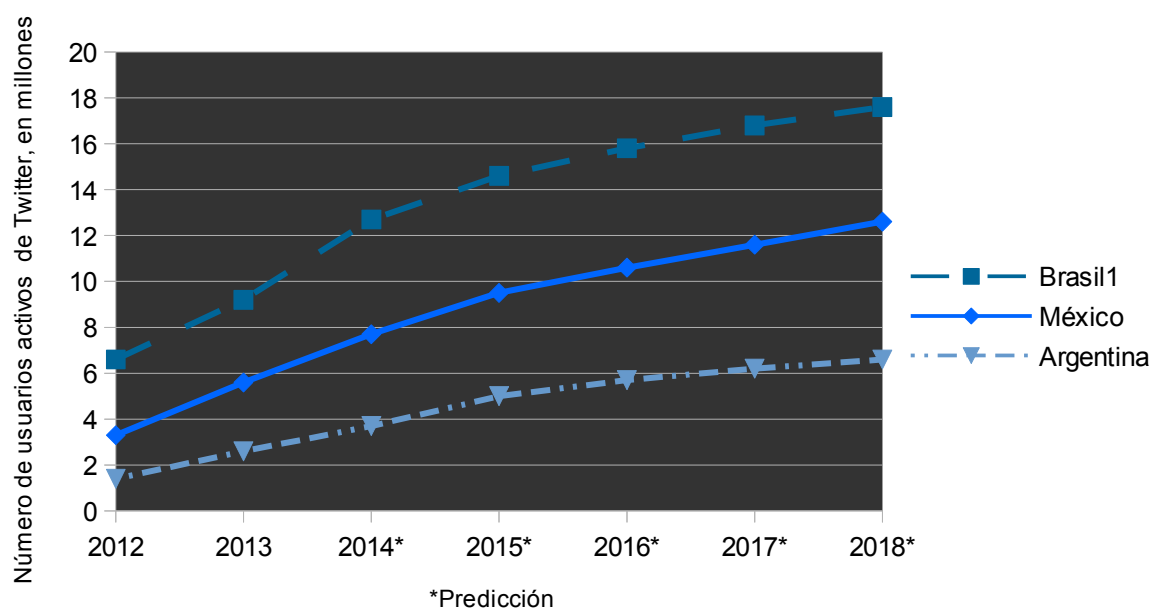
Dentro de las estrategias que se buscan para crecer rápidamente, está el lograr un alto índice de conversión de usuarios. “Puedes ir por el camino de identificar productos que impacten cierta métrica (por ejemplo ¿qué productos ayudan a dirigir adquisición?) y entonces averiguar cómo ya sea (1) optimizar esos canales para producir más valor O (2) construir nuevos canales de adquisición desde cero para agregar más volumen de adquisición.” (Johns, 2012).

Tener una conexión intermedia con otro usuario parece tener una influencia sobre la adquisición. Como mencionan Calvó-Armengol, Patacchini y Zenou (2008), que la gente se puede conocer de manera aleatoria en todo tipo de formas, pero para convertir un encuentro en una conexión, debe haber generalmente algún área en común. Para Bramoullé y Kranton (2007) los elementos del perfil sobre un usuario en una comunidad en línea puede ser un equivalente a la experiencia cara-a-cara para encontrar comunidades compartidas de pertenencia o un fondo en común. Calvó-Armengol *et al.* (2008) también apuntan que la presión social en los enlaces en una red son factores importantes, que proveen fundamento sobre comportamiento económico. Finalmente, Simply Measured (s. f.) en La Guía Completa a la Analítica de Twitter, mencionan sobre enfocarse en usuarios que no son seguidores actualmente, pues estos usuarios podrán no estar al tanto de tu manejo, pero se puede crear una conexión con estos usuarios para alentarlos a seguirte directamente.

El presente trabajo busca medir la influencia, durante la adquisición de usuarios, de los diferentes elementos que componen un perfil de Twitter, considerando que tener conexiones con usuarios en común ha de ser un elemento significativo, dada su influencia social.

## Marco Contextual

Lampe, Ellison y Steinfield (2007) mencionan que las comunidades en línea tienen diferentes fines, pero que una común e importante es formar conexiones entre usuarios. Existe alto interés en las superplataformas, porque dan acceso a más de 100 millones de consumidores, advierte Chen (2012), y las cifras de Twitter (2014) parecen confirmar reportando 271 millones de usuarios activos al mes, y 500 millones de 'Tweets', o mensajes públicos, al día. Además Larson en 2013 y con datos de Global Web Index, sostiene que Twitter es la plataforma con mayor crecimiento, seguido de Facebook, Google+, LinkedIn y Pinterest, y que además México ocupa el lugar número 15 dentro de esta red, medido en crecimiento de cuentas, con 16%. Según Lay (2012), 9 de cada 10 internautas mexicanos usan las redes sociales, y el 62% del total de usuarios se encuentran en los rangos de 12 a 34 años de edad.



**Figura 1. Número de usuarios de Twitter en Latinoamérica 2012-2018. Fuente: Elaboración propia basado en Statista 2014.**

Tomando estas cifras es notoria la importancia de crecer el producto en sus vías digitales: En el reporte de 2011 Social Media Marketing Industry Report: How Marketers Are Using Social Media to Grow

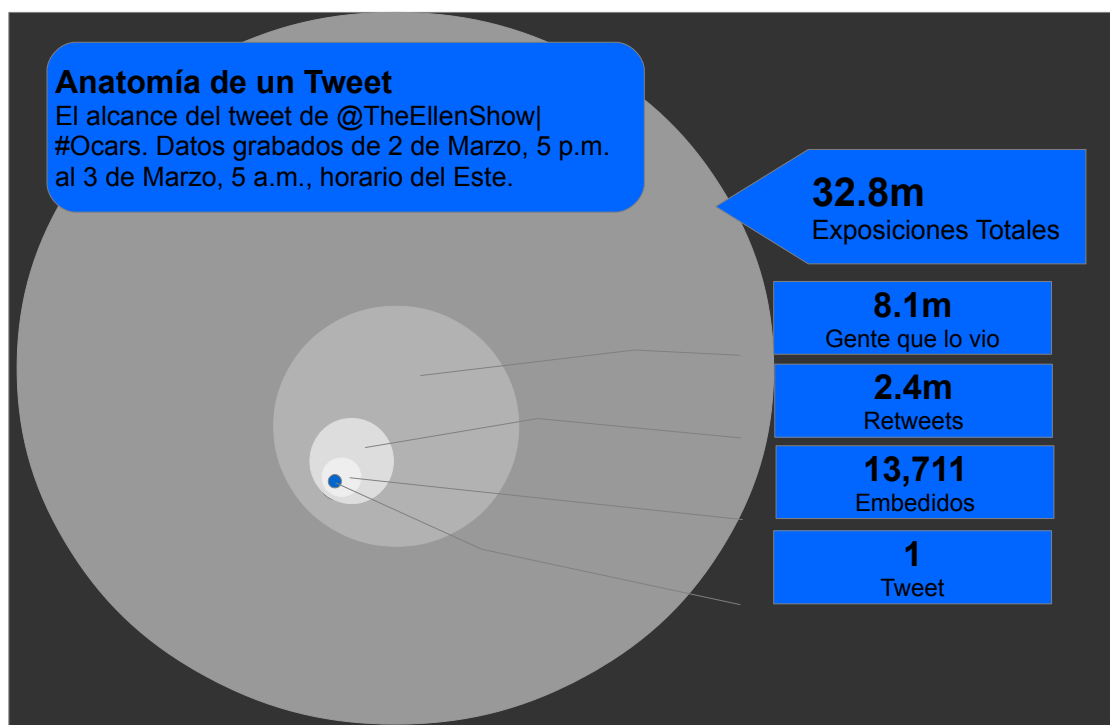
Their Business, Stelzner (2011) hacen notar que entre mercadólogos de redes sociales con experiencia, 72% reportan cerrar más negocios como resultado de esfuerzos en las redes sociales y 62% reportaron generación líder de beneficios con redes sociales.

“La manera más rápida de esparcir tu producto es distribuyéndolo en una plataforma usando APIs [interfaces de programación de aplicaciones], no MBAs [maestrías en administración de negocios]. El desarrollo de negocios es ahora API-céntrico, no gente-céntrico.” (Chen, 2012). Stelzner también menciona que los canales de redes sociales como Twitter y Facebook son la elección creciente de consumidores para recoger información y conectar con marcas.

Sin embargo Chen (2013) considera que las plataformas pierden viralidad, como Facebook donde los desarrolladores están migrando a plataformas más novedosas para poder crecer pues la novedad en un canal de distribución se apaga rápidamente. Twitter parece ser una plataforma con mucha actividad: Rogers (2014) describe la Copa Mundial de fútbol FIFA 2014 como el número más alto de actividad reportado por ellos mismos con 672 millones de tweets relacionados. Otros eventos importantes también han recibido alto flujo de publicaciones, según la misma fuente, como las Olimpiadas de Verano 2012 con 150 millones de tweets (Fitzgerald, 2012), el Super Tazón 48 de NFL con 24.9 millones (Fraser, 2014), Premios de Video Musical 2014 de MTV con 12.6 millones en los Estados Unidos (Charles, 2014) y los Premios Oscars 2014 con 19.1 millones (Fleischman, 2014).

Una sola publicación o tweet genera impresiones a través de una cadena de re-publicaciones que hacen los usuarios y lo transmiten a sus seguidores: 'Retweet' es el mecanismo por el que un usuario menciona a otro, y es uno de los principales medios de propagación y formación de opinión en Twitter (Hansen, Arvidsson, Nielsen, Colleoni y Etter, 2011); esto es factor de viralidad, pues actúa como versión digital de la palabra de boca en boca, analizan Hoang, Lim, Achananuparp, Jiang y Zhu (2011), Goel, Watts y Goldstein (2012) y Berger y Milkman (2011) . Durante los Premios Oscars sucedió el tweet con más

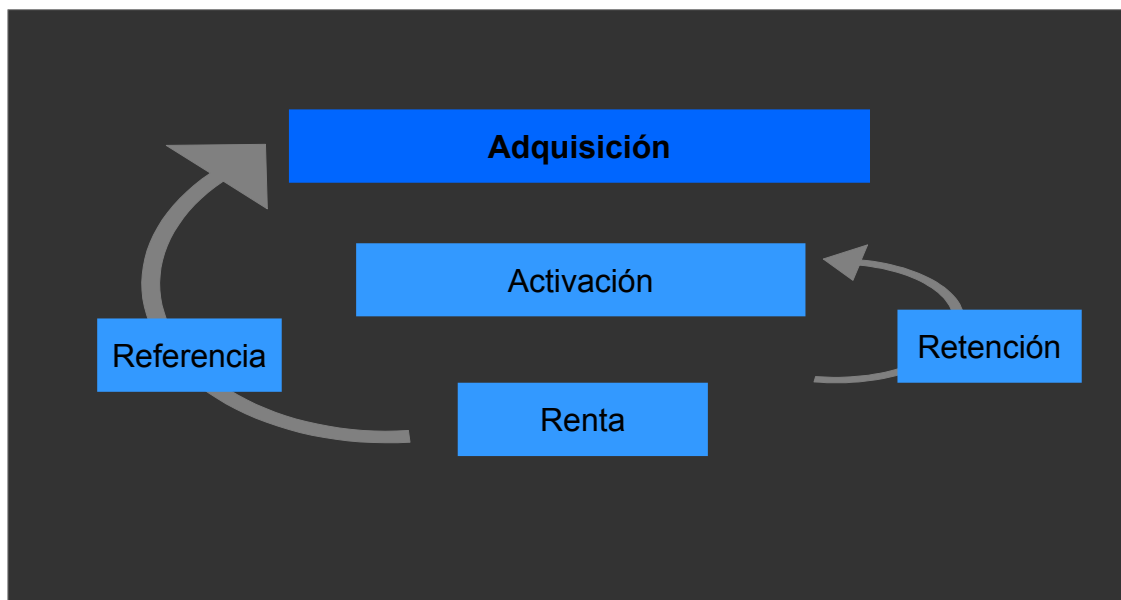
alcance en la historia de Twitter con un alcance de 32.8 millones de impresiones (Fleischman, marzo 2014), siendo retransmitido por retweet 2.4 millones de veces.



**Figura 2. Alcance de un tweet.** Fuente: Elaboración propia basado en Fleischman, 2014.

Puesto que los productos o compañías quieran tener impacto en las redes sociales, buscarán tener la mayor cantidad de seguidores posibles. Para modelar el proceso de afiliación de seguidores a una cuenta de Twitter se utiliza el embudo de conversión de clientes: “la idea de un embudo de conversión que ha estado al centro de la literatura de mercadotecnia por varias décadas (Strong, 1925; Howard y Sheth, 1969; Barry, 1987).” (Abhishek, Fader y Hosanagar, 2013).

La afiliación de un seguidor significa el paso de la adquisición a la activación, pues permite al contenido de la cuenta seguida alcanzar de nuevo su atención, llevándolo al proceso de activación. Johns (2012) menciona algunas componentes que pueden tener parte en este proceso de adquisición: mandar invitaciones, página de inicio, los pasos a la afiliación, afiliación.



**Figura 3. Embudo de conversión.** Fuente: elaboración propia con base a McClure 2007

Cuando un usuario recibe una invitación, analizará rápidamente el perfil público de la cuenta y decidirá seguirla o no. Lampe *et al.* (2007) parafrasean que “Los elementos del perfil proveen pistas acerca de amigos participantes en una comunidad en línea, que podrían actuar como la sutil entrevista hecha en ambientes cara-a-cara para encontrar comunidades compartidas o un trasfondo en común. (Bramoullé y Kranton, 2007)”. Por su parte Gabe (2010) menciona que “Cuando otro usuario recibe una notificación que lo has seguido en Twitter, hay una buena oportunidad que vayan a través de ese correo electrónico para revisar tu perfil. Cuando llegan a tu perfil en Twitter, tienes cerca de 10-15 segundos para impresionarlos lo suficiente como para que te sigan en reciprocidad”.

Dentro de la información inmediatamente visible en un perfil de Twitter se encuentran una foto de perfil, una imagen de encabezado, nombre, descripción, localización geográfica, estadísticos de actividad como cantidad de Tweets enviados y número de seguidores, últimos Tweets, vista en miniatura de fotos y videos publicados, e información de seguidores en común. Toda esta información es la primera impresión de la cuenta hacia el usuario y son los factores de los que habla Gabe, e



interesa el factor de los seguidores en común. Twitter lo describe como “seguidores que conoces” y son los usuarios que actualmente siguen a la cuenta o perfil, y que el usuario a su vez los sigue a ellos. Gandlgruber y Ricaurte (2013) definen que “transitividad” mide la probabilidad de la existencia de las relaciones entre actores que tienen nodos en común.

Lampe *et al.* (2007) argumentan que los individuos forman impresiones de otros para decidir si buscar o continuar una relación, y que en la formación de una impresión inicial, los individuos forman impresión muy rápido, tan pronto como 3 minutos en un ambiente cara-a-cara. Los mismos autores resumen que la pertenencia a una comunidad es una importante característica para evaluar el nivel de conocimiento en común entre individuos, y que este tipo de trasfondo común es importante en las interacciones porque facilita la comprensión y fomenta la cooperación entre participantes, especialmente en casos donde miembros del grupo no se conocen muy bien unos a otros o están interactuando a través de tecnología de información y comunicaciones. Un posible factor puede ser la presión social ejercida por dicho sentido de pertenencia, como estudia Calvó-Armengol *et al.* (2008) quienes afirma que la presión social de enlaces en una red es importante, pero que la presión social es baja en redes poco aglomeradas, dispersas y amplias, y que por tanto es necesario el aglutinamiento para efectos de presión social. Otro factor a considerar es el de la confianza, pues estudian Donath y Boyd (2004) que una muestra pública de conexiones es una verificación implícita de identidad: dado que las propias conexiones están vinculadas al propio perfil, quienes lo han presuntamente visto, lo habrán verificado, y debe asegurar una auto-presentación honesta.

Dada la importancia del crecimiento de cuentas de las redes sociales, se analizan los factores que componen una cuenta de Twitter para determinar la importancia de cada elemento, tomando como principal factor el tener conexiones intermedias con otros usuarios, bajo el supuesto que forma confianza y familiaridad entre la cuenta y el usuario.

## **Marco Teórico.**

Torres (2013) menciona la cibermetría como el estudio mediante el análisis cuantitativo de las diversas clases de información que coexisten en ciber espacio, empleando técnicas bibliométricas, cienciométricas, infométricas y de análisis de redes sociales. Para poder medir adquisición, una de las métricas más sencillas es la cantidad de subscriptores que se adquieren. Genes Interactive (s.f.) mencionan en su Libro de Twitter que una de las formas más simples y básicas de medir el nivel de éxito de armado y mantenimiento de una comunidad, es el número de seguidores. Así mismo mencionan en su sección de herramientas de métricas otras formas, como la cantidad de retweets, favoritos, o clicks a los enlaces.

En mercadotecnia de medios digitales es muy popular el método de pruebas A/B donde se expone a un grupo de usuarios a un formato, y a otro grupo otro formato con un cambio, manteniendo lo demás constante. Posteriormente se miden las respuestas y se compara su impacto. Varios autores abordan estas técnicas, como Lu y Liu (2014), Regalado (2014), Christian (2012) y Vaughan (2011), pues han sido una forma efectiva tomar decisiones a base de pruebas directas con los usuarios. Regalado(2014) lo describe como “carreras de caballos estadísticas entre dos o más versiones de un sitio web”. Christian (2012) resalta sus orígenes al crecimiento de Google y su implementación exitosa en la campaña política del presidente de los E.U.A., Obama, así como su amplio uso en grandes compañías como Amazon.com, Netflix y eBay. Sin embargo este mismo autor apunta hacia el peligro de usar pruebas A/B porque puede dirigir hacia máximos locales y estancarse el desarrollo de un producto, además de comprometer las grandes ideas si se quiere utilizar esta estrategia para todo tipo de decisiones.

Goel, Watts y Goldstein (2012) Analizan la difusión en varias redes digitales, entre ellas Twitter. En su

trabajo encuentran que en la mayoría de los casos, las redes de difusión cesan después de un grado de conexión entre usuarios. Para ello rastrearon el impacto de una publicación de un usuario semilla hacia sus conexiones y en subsecuentes conexiones generadas:

“Rastreamos la difusión de 80,000 noticias publicadas en el sistema de microblogueo Twitter durante Noviembre de 2011, donde el artículo original fue distribuido por uno de cinco sitios de noticias populares: The New York Times, CNN, MSNBC, Yahoo! News, y The Huffington Post. Se consideró que los individuos “adoptaron “ un artículo si fue publicado (i.e., “tweeteado”) un enlace a la historia.” (Goel *et al.*, 2012)

Así mismo los autores mencionan haber establecido ventanas de tiempo de dos semanas debido a la escala de tiempo de difusión en Twitter de tan sólo algunos días en la mayoría de los casos. Aunque el trabajo de Goel apunta hacia la difusión y no a la adquisición en específico, se puede observar la metodología orientada a Twitter.

Por su parte, Lampe *et al.* (2007) analizan la red de Facebook para encontrar cómo influyen los elementos de un perfil la cantidad de amigos o conexiones que tiene dicho usuario. Basados en teoría de costos de transacción, estiman que ciertos elementos del perfil dan señales de un trasfondo en común o afinidad, reduciendo el costo de la negociación que requiere para los participantes el engancharse en una relación, y por tanto facilitando la adquisición de conexiones. Para su análisis descargaron masivamente con algoritmos información del sitio, crearon índices del contenido de los usuarios (i.e., si contaban con información pública de los diferentes campos) y usaron regresión múltiple.

“Usamos un acercamiento por regresión múltiple para ver cuáles clústeres de elementos de perfil eran los más fuertemente asociados con el número total de amigos. Las principales variables independientes fueron los índices representando los grupos teóricamente divididos de campos de perfil, como fueron descritos anteriormente – evaluación, convencional, y contacto.

Usamos el registro de número de amigos como una medida dependientes en el análisis de regresión, porque la distribución de amigos estaba altamente sesgada con una larga cola creada por grupos más pequeños de usuarios con un número extremadamente alto de amigos.” (Lampe *et al.*, 2007)

Los resultados de Lampe et al. (2007) son que el coeficiente más alto pertenece al índice de referentes, es decir los campos que permiten a los usuarios tener contacto fuera del sitio, representando campos del perfil más difíciles de falsificar, proveyendo seguridad y reduciendo el costo de transacción.

En el caso de que la variable dependiente sea de naturaleza dicotómica o binaria, esto es, que su valor puede ser solamente 0 o 1, como el caso de esta investigación, los autores (González, Ávila y Aguilar, 2013; Katchova, 2013; Nagler, 1994; O'halloran, s. f. ) se basan en los modelos de probabilidad lineal, Logit y Probit. Para estos modelos se usa una función de enlace, que para el primero es una función de probabilidad y para el segundo es una función logística. González *et al.* (2013) mencionan sobre la distribución, que “Al no tener certeza de que la población que da origen a la muestra siga una distribución normal, el modelo logístico resulta el más adecuado”. La regresión logística usará un estimado de máxima similitud, en vez de un estimado de mínimos cuadrados, usado en regresión múltiple tradicional (O'halloran, s. f.).

Cabe mencionar que dentro de las practicas de crecimiento automatizado se genera el fenómeno de spam o acoso, de cuentas que utilizan mensajes genéricos y acciones masivas para llamar la atención en su producto, siendo generalmente molesto y que desarrolla un grado de reticencia y desconfianza en el usuario, y por tanto es buscado evitarlo, advierten Lampe *et al.* (2007), Palihapitiya (2013), y Donath y Boyd (2004). Simply Measured hace notar que dar favorito se ha convertido en una forma popular de enganchamiento en Twitter porque permite al usuario engancharse al contenido sin dar seguir al usuario (práctica muy socorrida por técnicas de spam). Twitter, además, cierra cuentas que practican spam.

## **Metodología y Resultados.**

Para la obtención de la muestra se tomó una cuenta activa de Twitter siguiendo a 280 usuarios y con 170 seguidores. De los 170 seguidores se sumaron los seguidores de cada uno de esos usuarios, obteniendo una fuente de 4,668,962 usuarios. De éstos se tomaron aleatoriamente 2,805 usuarios objetivo con quienes interactuar, para un primer grupo (A). En este grupo los usuarios siguen a alguien que a su vez sigue la cuenta que se estudia. De manera similar, pero tomando un grado más de separación, se obtiene otro grupo (B) de 2,700 usuarios objetivo, que no tienen un usuario intermedio que los conecte con la cuenta que se estudia. Esta muestra recogida, de 5,505 individuos, se logró descargando con algoritmos los datos desde Twitter a través de su API, buscando registrar la mayor cantidad de usuarios para el tiempo disponible, pues Twitter limita a 350 el número de llamadas por hora, donde cada dato requerido en este modelo es una llamada a Twitter. Se compara este número de muestra satisfactoriamente con los sugeridos en Hsieh, Bloch y Larsen (1998) para varias pruebas con diferentes modelos para tamaño de muestra en covarianza binaria, donde en sus pruebas de baja prevalencia, esperan tamaños de muestras de 1,833 a 2,648 individuos, con potencia o poder de 95% y nivel de significancia de dos lados de 5%.

Para cada usuario objetivo con que se interactuó, se descargaron secuencialmente los siguientes datos de su perfil de Twitter: cantidad de amigos o gente que sigue, cantidad de seguidores, cantidad de tweets publicados, cantidad de tweets a que ha dado favorito, cantidad de veces que el usuario ha sido enlistado, edad de la cuenta en días y desplazamiento UTC (tiempo universal coordinado). Además de estos datos obtenidos directamente de Twitter, se calcularón los siguientes índices:  $f2f$  (seguidores/seguídos), es un indicador de la proporción de qué tantos seguidores tiene contra cuántos sigue;  $tfr$  (tweets/edad), es la frecuencia promedio de sus tweets por días;  $ffr$  (favoritos/edad), es la frecuencia promedio de favoritos que da por días.

Variable	Descripción
friends	Cantidad de usuarios a los que sigue
followers	Cantidad de usuarios que lo siguen
tweets	Cantidad de tweets publicados por el usuario en total
favorites	Cantidad de favoritos que ha marcado el usuario en total
listed	Cantidad de veces que ha sido enlistado el usuario
since	Edad de la cuenta en días
fic	Amigos en común: 0 para ninguno, 1 para al menos uno
f2f	followers / friends - Compara seguidores contra seguidos
tfr	tweets / since – Cantidad promedio de tweets por día
frf	favorites / since – Cantidad promedio de favoritos por día
adq	Adquisición: 0 si no dio seguir, 1 si dio seguir a la cuenta

**Cuadro 1. Variables de estudio. Fuente: elaboración propia.**

Se excluyó el valor de UTC offset del análisis porque no es un dato disponible para todos los usuarios.

User	@user	tid	friends	followers	tweets	favorites	listed	since	f2f	tfr	ffr	UTC offset	FIC	adq
<a href="#">LivFree_DieRich</a>	@LivFree_DieRich	28158379	397	264	1011	7	0	1	0.6650	1011.0000	7.0000	0	1	0
<a href="#">Ross Quintana</a>	@Ross_Quintana	411845007	55379	62956	42123	58	1081	1009	1.1368	41.7473	0.0575	-25200	1	0
<a href="#">Justin Mares</a>	@jwmares	139193277	1937	2932	1266	509	76	1570	1.5137	0.8064	0.3242	-18000	1	0
<a href="#">Gema follows back!</a>	@unbecoming_	1941470528	15924	17759	417	2	2	317	1.1152	1.3155	0.0063	-14400	1	0
<a href="#">Shey Stitt</a>	@SheyPhoto	2350651844	2909	2754	64	157	30	181	0.9467	0.3536	0.8674		1	0
<a href="#">Rafael Martinez N.</a>	@rafaelmartinezn	62319454	11440	10598	122458	39379	83	1842	0.9264	66.4810	21.3784		1	0
<a href="#">CL0mydia</a>	@Clo_bo13	2150662871	1855	1325	1352	434	1	299	0.7143	4.5217	1.4515		1	1
<a href="#">Annabel ID !!</a>	@presumablyz9	2308041367	3745	4779	259	3	0	207	1.2761	1.2512	0.0145	-10800	1	0
<a href="#">Kelly Falardeau</a>	@kellyfalardeau	101845750	1626	1276	2084	34	25	1687	0.7847	1.2353	0.0202	-21600	1	0
<a href="#">Chama</a>	@melle_chama	982541827	390	440	2050	1252	0	1	1.1282	2050.0000	1252.0000	0	1	0
<a href="#">agah uğurel</a>	@agahugurel	256541480	542	177	643	5	2	1273	0.3266	0.5051	0.0039	-7200	1	0
<a href="#">Joselin</a>	@CofrePea	2172461989	502	359	1194	244	2	289	0.7151	4.1315	0.8443		1	0
<a href="#">Ariana4837v</a>	@KottaYuno	361613520	736	526	865	20	2	1090	0.7147	0.7936	0.0183	-16200	1	0
<a href="#">Que De Monadas</a>	@quedemonadas	2436733152	50150	62170	2706	2	8	131	1.2397	20.6565	0.0153		1	0
<a href="#">zahijS</a>	@IvanaZahira	2565008367	170	146	132	20	2	85	0.8588	1.5529	0.2353		1	0
<a href="#">dany98house</a>	@danytaccouse	1238593050	1098	406	5517	3873	3	534	0.3698	10.3315	7.2528	7200	1	1

**Cuadro 2. Muestra de los datos por usuario. Fuente: elaboración propia.**

Para la interacción con dichos usuarios se buscó entre sus últimas publicaciones alguna que no comenzara con “@” o “RT” , pues estas son respuestas o re-publicaciones, y se dio “favorito” a dicha publicación para generar atención en el usuario. Hubo un tiempo de espera de 10 días y se registraron

cuáles de todos los usuarios habían dado seguir a la cuenta. Éste es el número de usuarios adquiridos, que sumaron 146, una tasa de adquisición de 2.7%

Se utilizó el modelo logit dada la naturaleza binomial de la variable dependiente, donde ésta puede ser 0 o 1, y la probabilidad de ocurrencia que  $Y = 1$  está dada por:

$$Prob(Y = 1) = F(\beta x') \quad (1)$$

que es una función  $F$  que depende de un vector de variables explicativas  $x'$  y el vector de coeficientes a estimar  $\beta$ ; su función de distribución acumulada está dada por:

$$F(\beta x') = \Lambda(\beta x') = \frac{e^{\beta x'}}{1 + e^{\beta x'}} \quad (2)$$

Se realizaron estimaciones de regresiones logísticas para diferentes especificaciones con combinaciones de las variables definidas para encontrar su significancia en la explicación del modelo.

A continuación se resumen los diferentes escenarios.

Variable	Escenario A		Escenario B		Escenario C	
	Coef.	Prob.	Coef.	Prob.	Coef.	Prob.
tweets	-6.20E-006	0.6242				
tfr	-0.000368	0.6675				
since	-0.0004	0.0067	-0.000412	0.0043	-0.000478	0.0007
listed	-0.002959	0.2310	-1.82E-003	0.3237		
friends	1.05E-005	0.5136				
followers	-2.75E-006	0.8355				
fic	-0.122624	0.4760	-0.135159	0.4303		
ffr	0.000193	0.8465				
favorites	-7.26E-005	0.3264	-7.05E-005	0.3345		
f2f	-0.228659	0.1260	-0.240971	0.0895	-0.317079	0.0312
c	-3.024567	0.0000	-3.006663	0.0000	-22.44892	0.0000
McFadden R2		0.023973		0.022404		0.018931
Log Verosimilitud		-657.788		-658.8454		-661.1857

**Cuadro 3. Regresiones logarítmicas para 3 escenarios diferentes.**

Fuente: elaboración propia en base a resultados.

En el escenario A se incluyeron las variables en su totalidad para tratar de describir el modelo. Se puede observar que las probabilidades para las diferentes variables son altas, excepto para f2f y since. Para el escenario B se corrieron las 5 variables de menor probabilidad del escenario A, y se volvieron a observar f2f y since como las únicas variables significativas. Para el escenario C se utilizaron solamente las dos variables significativas. Además al observar las R cuadradas de Mc Fadden y los Log de Verosimilitud de cada escenario, se verifica que la mejor descripción está dada por el escenario C.

Se calcularon los efectos marginales de estas dos variables, esto es, el cambio en la probabilidad de ocurrencia de la variable dependiente (que  $Y=1$ ), dado un cambio en una unidad en una variable independiente, representado por la función:

$$\partial p / \partial x_j = \Lambda(\beta x') [1 - \Lambda(\beta x')] \beta_j = \frac{e^{\beta x'}}{(1 + e^{\beta x'})^2} \beta_j \quad (3)$$

obteniendo los siguientes resultados:

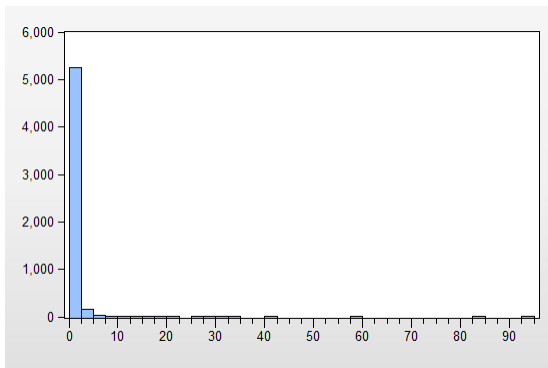
Variable	dy/dx	Prob.
f2f	-0.0071959	0.019
since	-0.0000109	0.001

**Cuadro 4. Efectos marginales.**

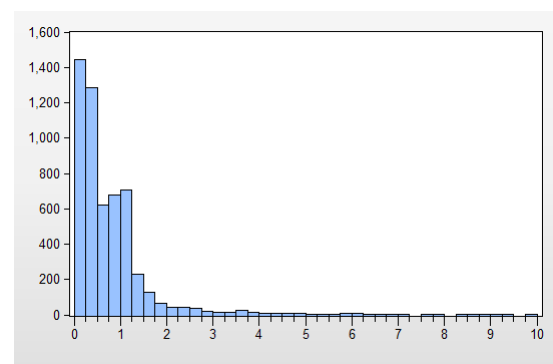
**Fuente:** elaboración propia en base a resultados.

Se realizó un análisis adicional truncando fuertemente la variable f2f, lo cual implica una disminución en la varianza, quitando las colas en un valor de  $0.3 < f2f < 2$ , quedando el centro de la muestra con 56% de la población total. A continuación se muestran histogramas de las variables sobre la muestra completa para apreciar su distribución.

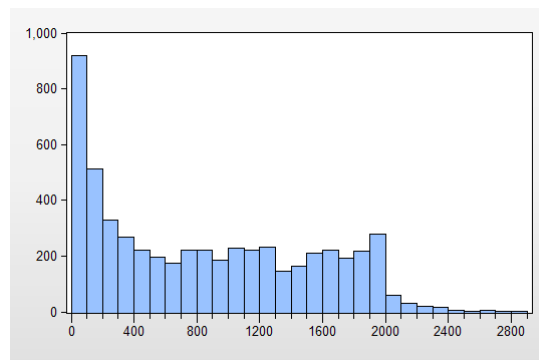




**Figura 4. Histograma para f2f, muestra completa.**  
Fuente: elaboración propia con base a datos.



**Figura 5. Histograma para f2f < 10.**  
Fuente: elaboración propia con base a datos.



**Figura 6. Histograma para since, muestra completa.**  
Fuente: elaboración propia con base a datos.

Este grupo se analizó de la misma manera con una regresión logística y cálculo de efectos marginales para las variables significativas f2f y since, con los siguientes resultados comparativos:

	Muestra completa		0.3 < f2f < 2	
Variable	dy/dx	Prob.	dy/dx	Prob.
f2f	-0.0071959	0.019	-0.0394118	0.000
since	-0.0000109	0.001	-0.0000179	0.000

**Cuadro 5. Efectos marginales comparativos entre muestras.**  
Fuente: elaboración propia en base a resultados.

Al eliminar los extremos, dy/dx cambia notoriamente dado que los efectos marginales se calculan sobre la media de la variable. Este hecho es señal de una varianza grande en los datos, aunque no tanta como para afectar la significancia de los coeficientes estimados.

## **Conclusiones.**

Varios autores han estudiado la importancia de los perfiles de los productos o compañías en las redes sociales, y de cómo son canales para llegar a más consumidores, al poder ampliar su ventana de adquisición en plataformas con cantidades muy grandes de usuarios y masividad en conexión y comunicación. Las compañías deben poner atención en el crecimiento de seguidores de sus cuentas de redes sociales, y por tanto es importante estudiar cómo impactan diferentes factores la toma de decisión de un usuario a seguir una cuenta o no seguirla.

Para el presente estudio se analizó la información disponible en el perfil público de Twitter de cada individuo; dicha información es usada por los usuarios como referencia para entender qué persona es con la que están interactuando, y por tanto esta información debe contribuir en la decisión de seguir a ese perfil. La información que se usó como variables obtenidas de los perfiles fue: cantidad de seguidores, cantidad de perfiles seguidos, cantidad de tweets, cantidad de favoritos, veces que ha sido enlistado, amigos en común y edad de la cuenta. Adicionalmente se modelaron 3 índices: relación de seguidores entre seguidos, cantidad de tweets entre la edad de la cuenta y cantidad de favoritos entre la edad de la cuenta.

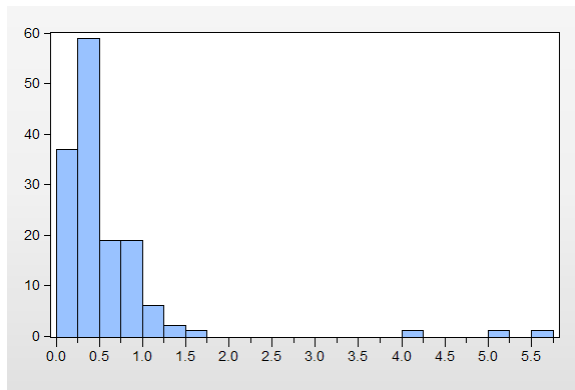
En base a estudios sobre la presión social y la búsqueda de círculos en común que permiten establecer lazos de confianza, incluso probados en medios digitales, se hipotetizó que el factor más importante dentro de los analizados sería el de tener amigos en común, es decir, que el usuario siguiera a otro usuario que ya seguía la cuenta. Sin embargo los resultados demostraron lo contrario, teniendo la variable (fic) un valor de probabilidad bastante alto (0.4760), en el escenario A, no aportando a la descripción del modelo.

Del los coeficientes restantes, la mayoría mostraron niveles de probabilidad bastante altos como para ser considerados significativos para el modelo. Los únicos dos coeficientes que obtuvieron un valor bajo de probabilidad fueron  $f2f$  (seguidores entre seguidos) y  $since$  (edad de la cuenta en días) con valores de 0.0312 y 0.0007 respectivamente, y también ambos coeficientes resultaron de signo negativo.

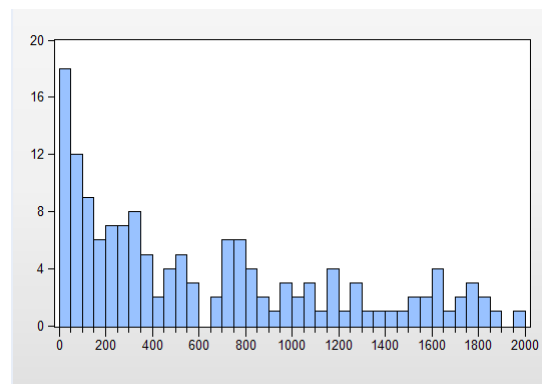
Tomando las dos variables,  $2f2$  y  $since$ , se obtuvieron sus efectos marginales de -0.0071959 y -0.0000109 respectivamente. Ambos coeficientes son de signo negativo, lo que implica que conforme la cuenta del usuario es más vieja o es más alta su tasa de seguidores/seguídos, disminuye la probabilidad de que decida dar seguir a la cuenta. Por cada unidad que  $f2f$  aumenta, es decir, que aumenta en un tanto de seguidos la cantidad de seguidores que tiene, cae aproximadamente 0.7% la probabilidad de que el usuario de seguir. Por cada día que acumula la edad de la cuenta, la probabilidad de seguir la cuenta cae 0.0000109; ésto es aproximadamente 0.4 % por año de antigüedad.

Poniendo lo anterior en otras palabras, es más probable que un usuario que tiene poco tiempo con su cuenta de Twitter decida seguir el perfil del producto, que una persona con bastante antigüedad. De la misma manera, si el usuario sigue a muchos más usuarios que los que lo siguen a él, es más probable que decida seguir el perfil del producto, comparado con un usuario que tiene muchos seguidores pero que sigue a pocos.

Para la cuenta actual se pudieron observar los siguientes histogramas, sobre los valores de  $f2f$  y  $since$  de los usuarios que fueron adquiridos:



**Figura 4. Histograma de f2f para usuarios adquiridos.**  
Fuente: elaboración propia con base a resultados.



**Figura 5. Histograma de since para usuarios adquiridos.**  
Fuente: elaboración propia con base a resultados.

donde el 80% de dichos usuarios se encuentran bajo las condiciones  $f2f < 1$  y  $since < 1400$ . Usando esta misma restricción para la muestra completa, se obtiene que el 56% de la población cumple con estas características. Ésta información se podría utilizar para restringir el proceso de dar favorito únicamente a usuarios bajo estas características, permitiendo una mejoría en el proceso, enfocándose sólo a usuarios potenciales. Si se toma una población sesgada bajo estas condiciones, la tasa de adquisición sube a 3.8%

Es notorio que la influencia de las variables analizadas en la toma de decisión del usuario es baja. La mayor parte de la información que permite que un usuario se haga un juicio de la cuenta que está evaluando, es a través del contenido de su descripción, de sus tweets y su contenido gráfico. Todos estos factores, mucho más abstractos, han quedado fuera de esta investigación, y se recomienda estudiar el impacto de ellos en la adquisición de usuarios, para futuros trabajos.

## Bibliografía

- Abhishek, V. Fader, P. S., & Hosanagar, K. (diciembre, 2013). *Media exposure through the funnel: A model of multi-stage attribution*. Recuperado de [http://smgwworld.bu.edu/wise2013/files/2013/12/wise20130\\_submission\\_156.pdf](http://smgwworld.bu.edu/wise2013/files/2013/12/wise20130_submission_156.pdf)
- Awareness. (agosto, 2011). *The social marketing funnel: Driving business value with social marketing*. Recuperado de [http://blog.pyramid.se/wp-content/uploads/2011/08/Social\\_Funnel.pdf](http://blog.pyramid.se/wp-content/uploads/2011/08/Social_Funnel.pdf)
- Berger, J., & Milkman, K. L. (2011). What makes online content viral? *Journal of Marketing Research*. Publicación anticipada en línea. doi:10.1509/jmr.10.0353
- Bramoullé, Y., & Kranton R. (2007). Public goods in networks. *Journal of Economic Theory*, 135, 478-494. doi:10.1016/j.jet.2006.06.006
- Calvo-Armengol, A., Patacchini, E., & Zenou, Y. (2008). Peer effects and social networks in education, *IZA discussion papers*, 3859. Recuperado de <http://ftp.iza.org/dp3859.pdf>
- Charles, S. (25 de agosto de 2014). *2014 #VMAs*. Recuperado de <https://blog.twitter.com>
- Chen, A. (27 de abril de 2012). *Growth Hacker is the new VP Marketing*. Recuperado de <http://andrewchen.co>
- Chen, A. (04 de abril de 2013). *Why developers are leaving the facebook platform*. Recuperado de <http://andrewchen.co>
- Christian, B. (25 de abril de 2014). *The A/B test: inside the technology that's changing the rules of business*. WIRED. Recuperado de <http://www.wired.com/>
- Donath, J., & Boyd, D. (Octubre, 2004). *Public displays of connection*. BT Technology Journal. Vol 22. No. 4.
- Fitzgerald, A. (13 de agosto, 2012). *Olympic (and Twitter) records*. Recuperado de <https://blog.twitter.com>
- Fleischman, M. (5 de marzo de 2014). *The reach and impact of Oscars 2014 Tweets*. Recuperado de <https://blog.twitter.com>
- Fraser, L. (3 de febrero de 2014). *Celebrating #SB48 on Twitter*. Recuperado de <https://blog.twitter.com>
- Gabe, G. (22 de junio de 2010). *The anatomy of a Twitter profile review*. Recuperado de <http://www.hmtweb.com/imd/?p=275>
- Gandlgruber, B., Ricaurte, P. (2013). *La naturaleza de las redes sociales virtuales y su impacto en el desarrollo económico y político*. En Islas, O. y Ricaurte, P., Investigar las redes sociales. Comunicación total en la sociedad de la ubicuidad. (pp. 48-62). Lugar: Razón y Palabra, México. ISBN 978-607-00-7126-3
- Genes Interactive. (s. f.). *El libro de Twitter*. Recuperado de <http://www.librodetwitter.com/>
- Goel, S., Watts, D. J., & Goldstein, D. G. (2012). *The structure of online diffusion networks*. Recuperado de <http://research.microsoft.com>
- González, D. J., Ávila, A., Aguilar, J. G. (2013). *Monotonía microempresarial y posibilidades de crecimiento en microempresas*. Paradigma Económico. Revista de economía regional y sectorial. Año 5 num 2. Lugar: UAEM. Estado de México, México.
- Hansen, L. K., Arvidsson, A., Nielsen, F. A., Colleoni, E., & Etter, M. (4 de enero de 2011). *Good friends, bad news. Affect and virality in Twitter*. Recuperado de <http://arxiv.org>
- Hoang, T., Lim, E., Achananuparp, P., Jiang, J., & Zhu, F. (2011). *On modeling virality of Twitter content*. Singapore Management University.
- Hsieh, F. Y., Bloch, D. A, Larsen, M. D. (1998). *A simple method of sample size calculation for linear and logistic regression*. Statistics in medicine. 17. 1623-1634. CCC 0277—6715/98/141623—12
- Johns, A. (14, mayo, 2012). en respuesta a What are some decisions taken by the “Growth team” at

- Facebook that helped Facebook reach 500 million users?. Quora. Recuperado de <http://www.quora.com>
- Katchova, A. (2013). *Probit and Logit models*. Econometrics Academy. Recuperado de <https://www.youtube.com>
- Lampe, C., Ellison, N., & Steinfield, C. (mayo 2007). *A familiar Face(book): Profile elements as signals in an online social network*. Michigan State University. San Jose, CA.
- Larson, D. (16 de mayo de 2013). *The 15 fastest-growing countries on the fastest growing network: Twitter*. Tweet Smarter. Recuperado de <http://blog.tweet smarter.com/>
- Lay, I. T. (noviembre de 2012). *Legislación, usos e impacto de las redes sociales*. Presentado en XX Encuentro Internacional de Educación a distancia. Universidad de Guadalajara. Guadalajara, México.
- Lu, L., Liu, C. (abril de 2014). *Separation strategies for three pitfalls in A/B testing*. Presentado en The User Engagement Optimization Workshop. Recuperado de [http://www.ueo-workshop.com/wp-content/uploads/2014/04/Separation-strategies-for-three-pitfalls-in-AB-testing\\_withacknowledgments.pdf](http://www.ueo-workshop.com/wp-content/uploads/2014/04/Separation-strategies-for-three-pitfalls-in-AB-testing_withacknowledgments.pdf)
- McClure, D. (agosto, 2007). *Startup Metrics for Pirates: AARRR!*. Presentado en Ignite Seattle 2007, Seattle, WA. Recuperado de <https://www.youtube.com/>
- Nagler, J. (3 de marzo de 1994). *Interpreting Probit Analysis*. New York University. Recuperado de <http://www.nyu.edu/>
- O'Halloran. (s. f.). *Lecture 10: Logistical regression II – Multinomial data*. Columbia University. Recuperado de <http://www.columbia.edu/>
- Palihapitiya, C. (enero, 2013). *How we put Facebook on the path to 1 billion user*. Udemy (Productor). Recuperado de <https://www.youtube.com/>
- Regalado, A. (22 de enero de 2014). *Seeking edge, websites turn to experiments*. MIT Technology Review. Recuperado de <http://www.technologyreview.com/>
- Rogers, S. (14 de julio de 2014). *Insights into the #WorldCup conversation on Twitter*. Recuperado de <https://blog.twitter.com>
- Simply Measured. (s. f.). *The complete guide to Twitter analytics. How to analyze the metrics that matter*. Recuperado de <http://gnip.com/docs/Simply-Measured-Complete-Guide-to-Twitter-Analytics.pdf>
- Statista. (2014). *Number of Twitter users in Latin America from 2012 to 2018, by country (in millions)*. Recuperado de <http://www.statista.com/>
- Stelzner, M. A. (abril de 2011). *2011 Social media marketing industry report. How marketers are using social media to grow thier business*. Recuperado de <http://www.socialmediaexaminer.com//SocialMediaMarketingReport2011.pdf>
- Torres, L. C. (2013). *Cómo analizar redes sociales en internet. El caso Twitter en México*. En Islas, O. y Ricaurte, P., Investigar las redes sociales. Comunicación total en la sociedad de la ubicuidad. (pp. 158-169). Lugar: Razón y Palabra, México. ISBN 978-607-00-7126-3
- Twitter. (agosto, 2014). *About*. Recuperado de <https://about.twitter.com/company>
- Vaughan, P. (18 de noviembre de 2011). *10 dos and don'ts of A/B testing*. HubSpot Inbound hub. Recuperado de: <http://blog.hubspot.com/>