

Mensajes, Posiciones Relacionales y Rutas: Una lectura computacional de la Campaña Presidencial 2016 en el Perú

José Manuel Magallanes

eScience Institute and Evans School of Public Policy and Governance
University of Washington
magajm@uw.edu

Department of Social Sciences
Pontificia Universidad Católica del Perú
jmagallanes@pucp.edu.pe

July 20, 2017

Abstract

El Twitter representa un espacio donde muchos usuarios vuelcan ideas cortas sobre diversos asuntos. La variedad de tales asuntos puede dificultar la lectura de lo que esté sucediendo en general, pero hay ciertas circunstancias que atraen un tema particular. Si bien muchos eventos pueden volverse ‘trending’ en las redes, las elecciones presidenciales son un evento que va llamando la atención progresivamente, donde los mensajes se volverán más interesantes y concretos a medida que las encuestas de opinión van informando quiénes se perfilan como favoritos conforme avanza la campaña.

Asumiendo ello, se procedió a recoger diversa información sobre la actividad de los usuarios (candidatos y personajes políticos relevantes). Este trabajo organizó y analizó los Tweets (usando el API del Twitter) durante la época de campaña presidencial de las elecciones de Perú 2016 (entre enero y junio 2016); datos que permitieron organizar computacionalmente las similitudes relacionales entre los candidatos, las rutas de campaña a partir de los mensaje, así como los mensajes a los seguidores en diversos cortes espaciales y longitudinales. Los resultados encontrados presentan patrones que reflejan características que pueden ser consideradas cualitativamente válidas, aun cuando el Twitter no es un medio masivo ni representativo de la población en Perú.

Keywords— Twitter, redes, Peru, campaña electoral presidencial, distancia, similitudes.

1 Contexto

Si bien muchos eventos pueden volverse ‘trending’ en las redes, las elecciones presidenciales son un evento que va llamando la atención progresivamente, donde los mensajes se volverán más interesantes y concretos a medida que las encuestas de opinión van informando quiénes se perfilan como favoritos conforme avanza la campaña.

En el Perú, las elecciones volvieron en 1980, marcando el retorno a la democracia luego del gobierno militar instaurado en 1968. Desde aquella elección, el Perú viene mostrando una variada oferta de candidatos a la Presidencia y al Congreso, no dando el sistema señales de dirigirse hacia un bipartidismo (5 partidos diferentes han ganado las elecciones Presidenciales desde 1980, con dos reelecciones, la de Alberto Fujimori y la de Alan García).

El Perú tuvo *Elecciones Generales* el 2016, un mismo proceso donde se eligió al Presidente del Ejecutivo y a 130 Congresistas para los siguientes cinco años. La elección fue el 10 de abril, pero hubo una segunda vuelta el 5 de junio, sólo para definir la presidencial, ya que ninguno alcanzó la mitad más uno de los votos válidos. Como es costumbre, aún cuando se trata de dos elecciones, la atención mayoritaria de los medios y los analistas es la elección Presidencial, lo cual forzaría a que la discusión on-line y off-line del ciudadano sea mayoritariamente sobre ese mismo tema.

Para estas elecciones presidenciales, se inscribieron inicialmente 19 candidatos. La oferta tenía hasta tres partidos que apostaban por un cambio al modelo económico, un mayor rol del Estado, una política medioambiental más eco-sostenible y una política energética para el beneficio del consumo interno, de los cuales sólo Verónica Mendoza, la candidata del “Frente Amplio”, estuvo luchando hasta el final para pasar a segunda vuelta (siendo la candidata que más creció pues enero tenía menos de 4%, llegando hasta cerca del 19% en las elecciones). Por otro lado, teníamos dos partidos claramente opuestos al concepto de cambio que representaba la izquierda: “Peruanos por el Cambio”, con Pedro Pablo Kuczynski (PPK), como candidato presidencial; y “Fuerza Popular”, que llevaba de candidata a Keiko Fujimori (hija de Alberto Fujimori, presidente del Perú 1990-2000, actualmente recluso en Perú). Luego estaban varios partidos tratando de ubicarse en el “centro”, algún lugar que los diferencie de las dos posiciones antes mencionadas.

Algunos candidatos renunciaron durante la campaña, al ver que sus posibilidades de ganar eran poco alentadoras. Pero no sólo hubo renunciadas, sino dos exclusiones de candidatos, los cuales, según las encuestas, tenían el apoyo popular suficiente para ser uno de los dos en pasar a una inevitable segunda vuelta (Keiko Fujimori estaría de todas maneras en la segunda vuelta). Estos candidatos eliminados fueron César Acuña (“Alianza para el Progreso”) y Julio Guzmán (“Todos por el Perú”) (como su exclusión no se dio hasta un mes antes de las elecciones, se les ha incluido en algunos cálculos de este trabajo). La salida de estos candidatos facilitaría no sólo la subida de PPK y Verónica Mendoza (Keiko siempre mantuvo su primer puesto en las encuestas, en promedio con 35%), sino además de Alfredo Barnechea, el candidato presidencial de “Acción Popular”. Llamaba la atención que los dos ex-presidentes que postulaban, Alejandro Toledo (“Perú Posible”) y Alan García (“APRA”) mantuvieron siempre posiciones rezagadas (por debajo del 8%); y además, el “Partido Nacionalista”, cuyo líder era ese momento Presidente saliente del Perú, Ollanta Humala, no pudo lograr que su candidato a la presidencia, Daniel Urresti, salga del grupo “otros” (debajo del 5%) en las encuestas (los detalles oficiales de los demás candidatos pueden además encontrarse en la web de *Voto Informado*¹). Al final de la primera vuelta, Keiko logró el 40% de los votos válidos, PPK el 21%, Verónica 19%, y todos los demás por debajo del 7%. Así, Keiko y PPK compitieron por la Presidencia en la segunda vuelta.

Habiendo pasado esta etapa primera etapa, me pareció pertinente compartir algunos análisis

¹<http://www.votoinformado.pe/voto/candidatos.aspx>

básicos que hicimos desde el eScience Institute de la Universidad de Washington, donde estamos organizando diversos datos accesibles públicamente sobre Perú, con el apoyo de mi equipo local en la Pontificia Universidad Católica del Perú².

2 Midiendo el uso de Twitter

Por lo general, las encuestas han sido la mayor fuente de información para evaluar el avance o retroceso de un candidato, y justamente deben capturar ese desempeño en el momento en que éstas se aplican; buscar ver en ellas fuente de pronóstico, es una aventura diferente, que el mejor diseño no garantiza, porque tampoco es su objetivo. A pesar de ello analistas, opinólogos y wishful-thinkers en Perú buscan en las encuestas lo que ellas no pueden decir. Por ello, decidimos aproximarnos a los datos de las redes sociales, no para reemplazar a las encuestas, sino para buscar otra información que ellas tampoco brindan. Así, decidimos recolectar y organizar datos del mundo on-line, en particular de Facebook, de Twitter y de YouTube. Estos medios on-line tienen diferentes características que los candidatos deben saber aprovechar. De manera simplista, podríamos argumentar que Facebook podrían utilizarse para mensajes más elaborados que abren posibilidad a debate; lo cual demanda, a la vez, que el candidato (o el que administre la cuenta) esté atento a los comentarios y participe del debate creado. Twitter estaría para los mensajes, en texto o imágenes, claros y concisos y esperas que el rebote ('retweet') se vuelva 'viral'. Y Youtube, es de hecho, el más exigente, pues requiere que prepares material que hay que saber capturar, editar y post-producir. Para todos los casos, lo deseable es que el medio sirva para que un candidato logre más seguidores, y que luego estos seguidores se conviertan en votos favorables. Pero a la vez, estos medios exponen no sólo tus mensajes, sino tus errores (los cuales suelen propagarse mucho más rápido que las ideas de tu plan de gobierno). Las tres figuras a continuación muestran nuestra primera exploración de la actividad de los siete candidatos que no estaban en la lista "otros" en las encuestas; incluyendo a los dos excluidos.

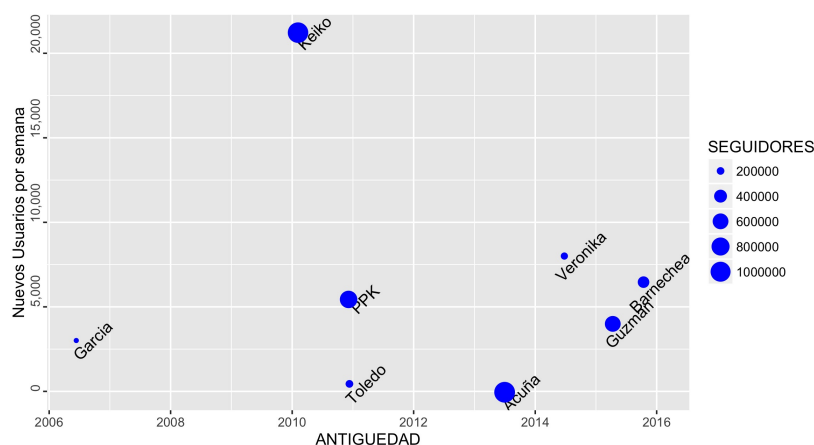


Figure 1: Uso del Facebook.

²<https://projectsuw.github.io/peru2016elections/>

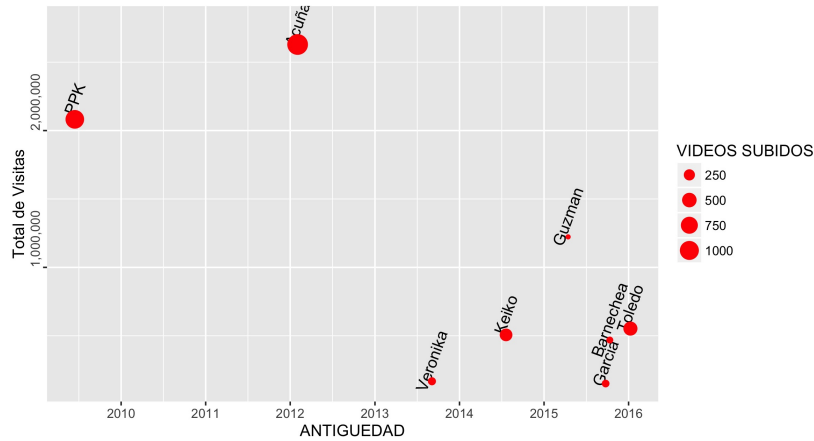


Figure 2: Uso del YouTube.

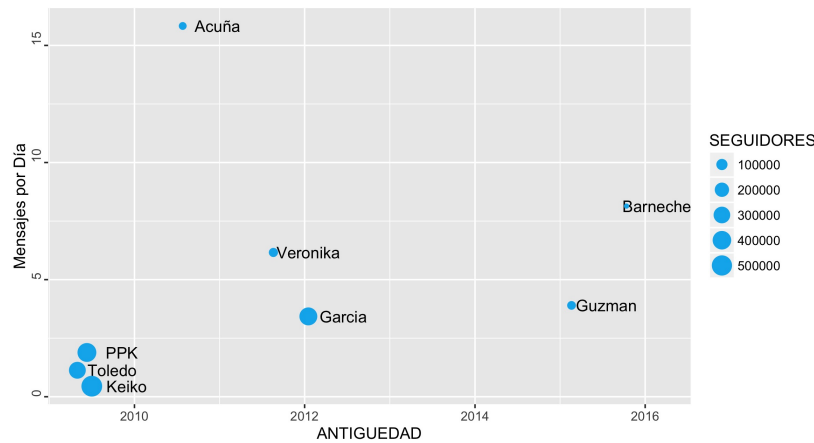


Figure 3: Uso del Twitter.

De los cuadros anteriores podemos observar que no hay candidato que resalte en estos tres medios en simultáneo; lo que daría muestras de tener una estrategia integral, si la tuvieran, aún endeble. Se debe tener en cuenta, además, que, en Perú, el uso de Twitter, Facebook o YouTube está sesgado hacia poblaciones con acceso a Internet, especialmente en sus teléfonos móviles, que se concentran más en las zonas urbanas, donde los usuarios de la costa y especialmente de Lima provocarían un sesgo que dificultaría una interpretación nacional.

Un tema importante a destacar a partir de los cuadros mostrados es que hemos realizado simples conteos. Y aunque intuitivamente parece adecuado, podría hacer pensar que la posición que cada uno ocupa resulta de una relación unívoca con sus contactos, cuando eso no es así. Un usuario cualquiera puede seguir a todos los candidatos en simultáneo, un seguidor no es necesariamente un simpatizante o al menos es difícil de ser contado como tal si sigue a más de uno de los candidatos. Hay algunos rankings circulando en Perú sobre gente influyente en la web, como el elaborado por la consultora Llorente y Cuenca³, pero creemos que modelos de conteos no son los más apropiados,

³<http://www.desarrollando-ideas.com/2014/09/top-50-de-influenciadores-en-la-web/>

considerando el tipo de estructura de datos (red) que se tiene. De todo ello, pasemos más bien a explorar cuantitativamente la naturaleza relacional de los datos.

3 Mirando relaciones

Para explorar la naturaleza relacional de los datos se requiere información de las conexiones. Para estos caso nos limitaremos a Twitter dada la flexibilidad que sus servicios para desarrolladores permite (el Twitter API). Una manera de evidenciar justamente que los candidatos no están separados en el mundo on-line es construir una red en la que estén ellos y sus contactos o ‘amigos’. Es a través de sus amigos que podemos detectar algunas cercanías entre ellos, así como distancias, como vemos en la Figura 4.

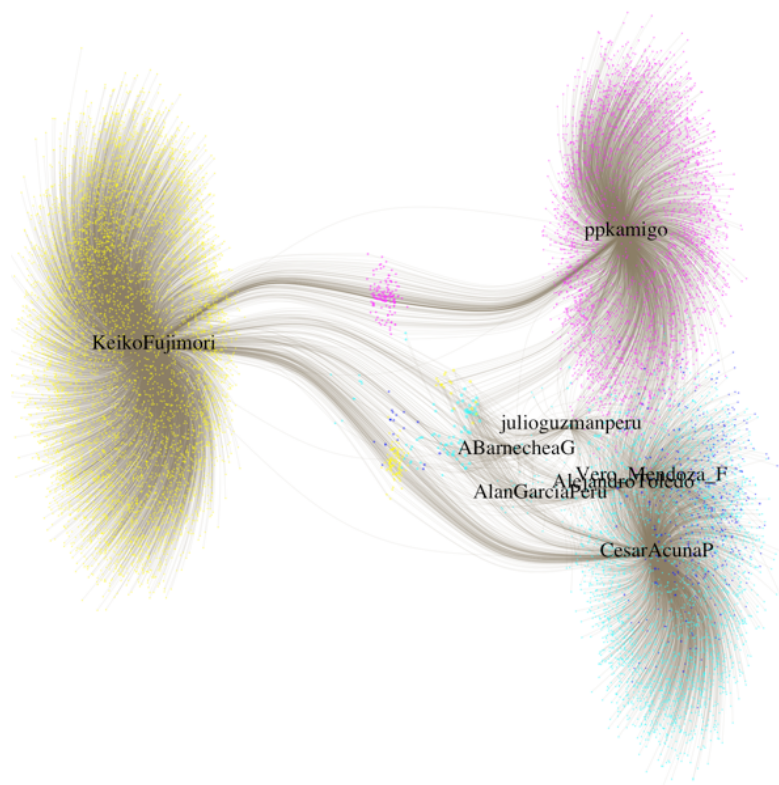


Figure 4: Relaciones compartidas entre candidatos presidenciales en Twitter.

Si las encuestas fuerzan a que uno defina a quien sigue, el mundo de las redes permite un seguimiento multidireccional como se muestra arriba. La posición (layout) de los nodos siempre puede ser manipulada para ver mejor alguna característica de la red, por lo que el layout elegido (el Fruchterman-Reingold) se utilizó para validar la existencia de comunidades y a la vez para mostrar que hay, en efecto, intersección entre los amigos de los candidatos. Sin embargo, la gráfica anterior no permite decir más. Fue así que, inspirados en un artículo de Ernesto Calvo ⁴ y un trabajo de

⁴<http://elestadista.com.ar/?p=4112>

Pablo Barberá ⁵, decidimos procesar los datos recogida para poder detectar, en efecto, distancias y proximidades. La idea sencilla detrás, es encontrar una manera de acercar más a los candidatos que tengan más contactos en común; así, menos contactos en común alejarían a los candidatos. Las posiciones que se calcularon se ubicaron en dos dimensiones. Aquí, lo importante es validar si Twitter puede reflejar on-line lo que la percepción general tiene en claro del mundo off-line; es decir, que dos candidatos (Keiko Fujimori y Pedro Pablo Kuczinsky) se distanciaban de los demás por representar la continuidad ; y que la mayoría de los candidatos estaban entre el centro, y menos hacia la izquierda. La Figura 5 muestra la posición de los candidatos que figuraban en las encuestas fuera del grupo “otros” (se ha incluido además, nuevamente, a los dos candidatos excluidos).

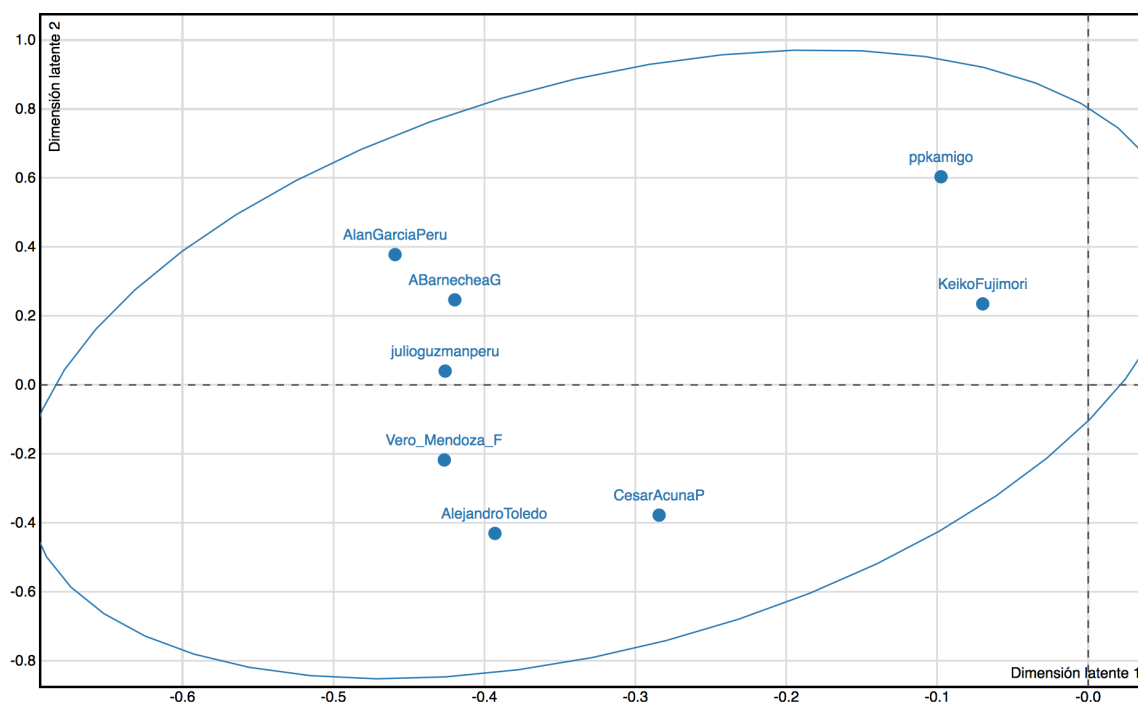


Figure 5: Cercanía entre candidatos presidenciales.

Este gráfico, con sólo siete nodos y la información de amigos en común, muestra como Keiko y PPK se alejan de los demás. Por lo visto la información no es suficiente para separar claramente a los otros 5 candidatos. En este caso, queda claro que la red está muy incompleta. Por ello, decidimos añadir más actores. Veamos que sucede al incluir a todos los 19 candidatos presidenciales con sus candidatos a vice presidentes (si es que tenían cuenta de Twitter). El resultado se ve en la Figura 6.

El gráfico anterior muestra ahora mejores detalles. Se han introducido más actores, que han traído consigo también más conexiones que han ayudado a diferenciar a los que se agrupaban al centro. Con la ayuda de una técnica de conglomerados (dbscan) hemos además encerrado en una elipse las facciones que se pueden revelar. En particular, se refleja un claro alejamiento del centro de Alan García del APRA y de su candidata a la Vice Presidencia, Lourdes Flores. Otra notoria definición se obtiene al incluir a Marco Arana (vozdelatierra), que acompaña a Verónica como su candidato a

⁵<http://pan.oxfordjournals.org/content/23/1/76.full>

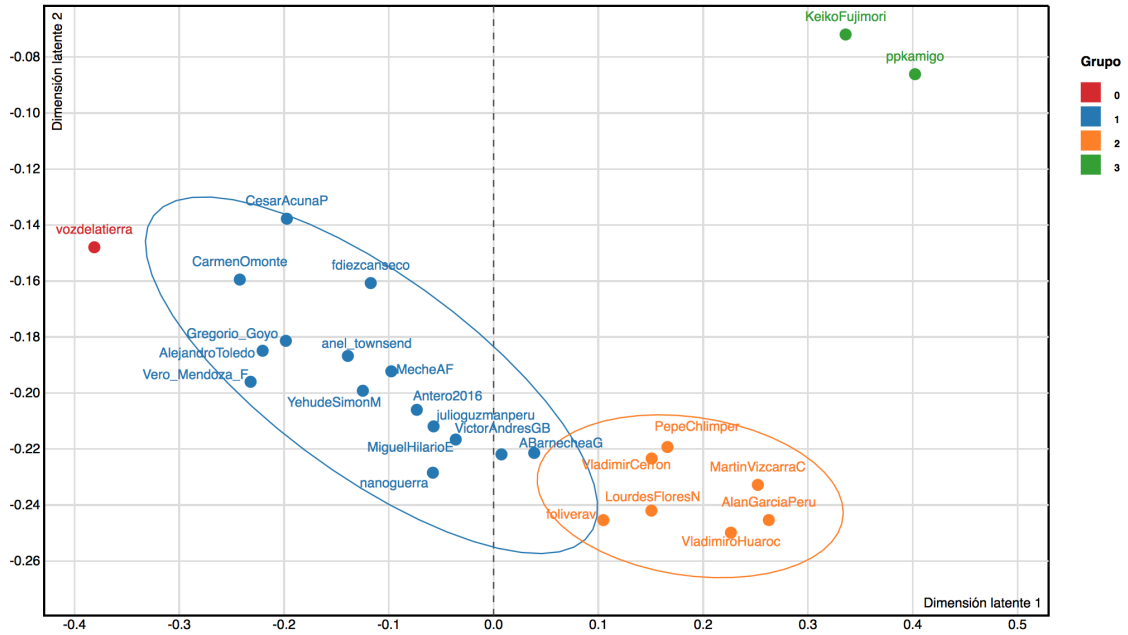


Figure 6: Cercanía entre candidatos presidenciales. Incluye a los vice presidentes.

la vice presidencia. Arana se aleja mucho del bloque naranja y verde, y queda en las inmediaciones del azul, donde sigue Verónica Mendoza, candidata presidencial de su agrupación. ¿Qué habría que hacer para desordenar esta red? Con esto quiero decir que mientras se añadan personas del mundo político, la red debe tomar forma similar, pero si añadimos personas de otra dimensión (jugadores de fútbol, cocineros, etc.) es probable que las posiciones halladas pierdan sentido, pues estaríamos mezclando múltiples dimensiones, cuando lo que se busca es entender un concepto político particular. Por ejemplo, si añadimos periodistas políticos, la red toma otras características pero aún sigue siendo válida. En la Figura 7 incluiremos a tres comunicadores: Cesar Hildebrant (ensustrece) que tiene una postura muy crítica al modelo; a Aldo Mariátegui que tiene una postura más condescendiente con las posturas de PPK y Keiko; y a Rosa María Palacios (rmapalacios) que busca ser menos extremista que ambos.

Las posiciones de todos siguen siendo prácticamente las mismas, pero las elipses muestran qué candidatos de un grupo estarían en los límites extremos, de ahí que Acuña sigue en el grupo de centro (por el color) pero al estar en la periferia de la elipse indica que es un atípico entre estos. Respecto a los comunicadores, Hildebrant se acerca a la izquierda, Rosa María al centro y Aldo a la derecha, todo ello como se esperaba. Vemos así, que se detecta bien la posición de los comunicadores; pero es ese rol, el de servir de enlace entre políticos, el que disminuye un poco las distancias entre todos y evita que se separe el conglomerado naranja del gráfico anterior. Hasta aquí, con las limitaciones que pueda tener Twitter, vemos que las relaciones que conviven entre sus usuarios sí refleja lo que se percibe o asume como evidente en el mundo off-line.

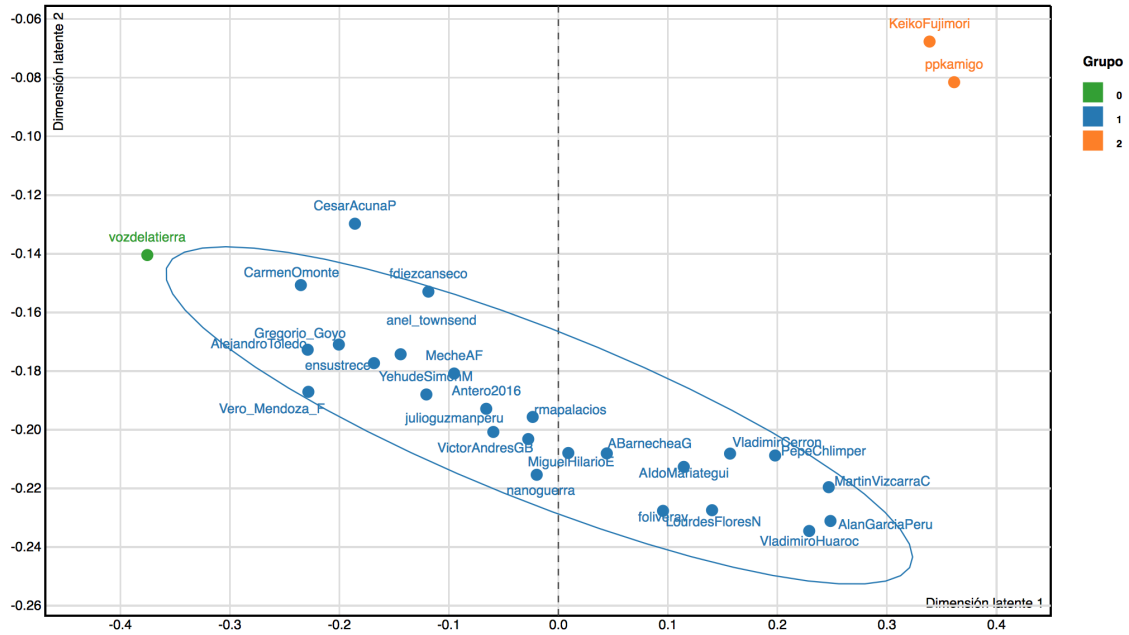


Figure 7: Cercanía entre candidatos presidenciales. Incluye vice presidentes y periodistas.

4 Comparando Discursos

4.1 Entre Candidatos

Para esta sección vamos a usar los mensajes de Twitter. Estando a pocos días de las elecciones, se decidió utilizar los mensajes que cada uno de los 5 primeros candidatos en las encuestas envió vía su cuenta Twitter. La metodología, de manera intuitiva, organiza y contabiliza el vocabulario utilizado. Considérese además que Twitter solo acepta 140 caracteres. Entonces, lo que se hace es comparar grupos de vocabulario. A muchos les gusta, por ejemplo, mostrar nubes de palabras, pero ello sirve poco de manera comparativa. En este caso, el programa que se utiliza (hecho en Python) permite darnos un valor entre 0 y 1. Mientras más cercano a 0, los discursos tienen vocabulario similar, mientras más cercano a 1, son más distantes. Un 0.5 significaría que el lenguaje ha variado en un 50%. Muchos pueden cuestionar que hay palabras con poco significado que eleva la probabilidad que los lenguajes sean similares; pero la técnica utilizada (tf-idf) toma en cuenta ello y pondera adecuadamente las palabras utilizadas. En este caso, se han agregado todos los mensajes de Twitter desde el 1 de enero del 2016 hasta el 2 de abril de 2016. Así mismo sólo se han incluido a los primeros 5 puestos según las encuestas. Los resultados se muestran en la Figura 8.

DE	A	DISTANCIA	RANKING
AlanGarciaPeru	ppkamigo	0.868181	1
AlanGarciaPeru	Vero_Mendoza_F	0.873703	2
AlanGarciaPeru	ABarnecheaG	0.888388	3
AlanGarciaPeru	KeikoFujimori	0.928531	4
ABarnecheaG	ppkamigo	0.852894	1
ABarnecheaG	Vero_Mendoza_F	0.871664	2
ABarnecheaG	AlanGarciaPeru	0.888388	3
ABarnecheaG	KeikoFujimori	0.948659	4
Vero_Mendoza_F	ppkamigo	0.849496	1
Vero_Mendoza_F	ABarnecheaG	0.871664	2
Vero_Mendoza_F	AlanGarciaPeru	0.873703	3
Vero_Mendoza_F	KeikoFujimori	0.930685	4
ppkamigo	Vero_Mendoza_F	0.849496	1
ppkamigo	ABarnecheaG	0.852894	2
ppkamigo	AlanGarciaPeru	0.868181	3
ppkamigo	KeikoFujimori	0.931253	4
KeikoFujimori	AlanGarciaPeru	0.928531	1
KeikoFujimori	Vero_Mendoza_F	0.930685	2
KeikoFujimori	ppkamigo	0.931253	3
KeikoFujimori	ABarnecheaG	0.948659	4

Figure 8: ¿Qué tanto se has diferenciado los discursos (vía Twitter)? (Del 1/1/16 al 2/4/16)

Del cuadro anterior, se puede afirmar que el lenguaje de cada candidato se diferencia claramente entre ellos, pues los valores al compararlos de dos en dos están más cerca a 1 y por encima de 0.5.

4.2 De un mismo candidato en el tiempo

Finalmente, hemos organizado los discursos de cada candidato por mes. A partir de ello, se ha calculado qué tanto un candidato ha ido variando su propio lenguaje en el tiempo. Esto lo vemos en la Figura 9

BARNECHEA	Enero	Febrero	0.7714
	Febrero	Marzo	0.3785
	Enero	Marzo	0.8579
GARCIA	Enero	Febrero	0.5047
	Febrero	Marzo	0.4521
	Enero	Marzo	0.7286
MENDOZA	Enero	Febrero	0.3922
	Febrero	Marzo	0.1939
	Enero	Marzo	0.5100
PPK	Enero	Febrero	0.5656
	Febrero	Marzo	0.4017
	Enero	Marzo	0.7401
FUJIMORI	Enero	Febrero	0.3397
	Febrero	Marzo	0.8035
	Enero	Marzo	0.8703

Figure 9: ¿Han mantenido los candidatos su mismo discurso en Twitter durante la campaña?

En este caso hay cosas por resaltar:

- Verónica Mendoza y Keiko son las que menos variaron su lenguaje de enero a febrero.
- Alfredo Barnechea es el candidato que más varió su lenguaje de enero a marzo.
- Barnechea y Keiko han tenido las mayores variaciones de lenguaje de enero a marzo.

- Keiko es la que modificó más su lenguaje de febrero a marzo. Es notorio que el día del debate entre candidatos (una semana antes de la elección), ella misma firmó un documento donde expresaba por escrito diversos compromisos que no utilizó durante su campaña y dónde busca comprometerse con un estilo de gobierno que se diferencia del de su padre.
- Verónica es quien menos ha variado su lenguaje, y su lenguaje de febrero es prácticamente el mismo al de marzo (lo cual coincide con la opinión de Steve Levitsky en su artículo “El resurgimiento de la izquierda”).

Recuérdese que los valores van de 0 a 1, por lo que pueden interpretarlos como porcentajes. Al mostrar estas diferencias y resaltar las mínimas y máximas, lo usamos nuevamente como referencia. Tengo en cuenta que esto NO ES NECESARIAMENTE una medida de consistencia o coherencia del candidato.

5 CONCLUSIONES

En este artículo, hemos supuesto que las elecciones presidenciales pueden ser consideradas como uno de esos eventos donde la gran mayoría de los usuarios de las redes sociales pertenecientes a un país enfocan sus mensajes, es decir, sus mensajes revelan sus preferencias y/o intenciones. Así mismo, consideramos que para explotar adecuadamente lo que los medios como el YouTube, Facebook o Twitter pueden informar al analista político se debe aprovechar justamente la estructura de datos, la red, que éstos tienen, en vez de apostar metodológicamente en conteos. Es decir, la red debe servir justamente para detectar patrones que las encuestas de opinión no puedan encontrar. Aprovechando la información de las relaciones disponible vía el API de Twitter, se ha logrado validar, para el caso del Perú, que las distancias y cercanías entre los políticos se reflejan de manera similar a lo asumido en el mundo off-line. Para evidenciar ello, se construyó una red asociando a los candidatos según sus amistades en común, y la vez es la cantidad de esas amistades comunes determina la cercanía y la lejanía entre ellos. Para lograr reflejar la realidad, se tuvo que incluir en diferentes momentos sólo candidatos, los candidatos de su plancha presidencial, y también periodistas políticos. Ello pues los nodos de la red deben pertenecer a la misma dimensión, la política, para que se note claramente cómo se configura este tipo de relaciones sociales. Finalmente, se ha validado también que el lenguaje utilizado por los candidatos en Twitter permite reflejar por un lado que sus mensajes sí los diferencian, y que, por otro lado, ellos varían con diferente intensidad su lenguaje. En particular, la poca variación de Mendoza, y la mayor variación de Keiko reflejaron lo que se apreció en la campaña. Quedan pendiente muchas aplicaciones de este tipo de datos, más aún incluyendo y combinando información de los medios como YouTube o Facebook que no han sido considerados; pero hasta aquí, vemos que el avance de Twitter en los usuarios de Perú permite reflejar situaciones del mundo off-line al aplicar estas dos sencillas aplicaciones, por lo que se espera que podamos obtener cada vez mejor información en los siguientes procesos políticos.