



REPORTE DE CONVERSACIÓN DE MIGRACIÓN Y XENOFOBIA

PANAMÁ

1 DE ABRIL 2020 A 1 DE ABRIL 2021



#XENOFOBIAZERO
ALZA TU VOZ CONTRA LA DISCRIMINACIÓN



ÍNDICE

1. RESUMEN EJECUTIVO	3
2. DATOS Y METODOLOGÍA	4
3. LA CONVERSACIÓN DE XENOFOBIA	4
3.1 EVENTOS MÁS RELEVANTES EN LA CONVERSACIÓN DE XENOFOBIA	4
4. CLASIFICACIÓN DE LA CONVERSACIÓN DE XENOFOBIA	8
4.1 MODELO LDA	8
4.2 RESULTADOS	9
5. CUANTIFICACIÓN DE XENOFOBIA POR NACIONALIDAD	12
6. CONCLUSIONES Y RECOMENDACIONES DE POLÍTICA	13
ANEXO 1. NOTICIAS Y XENOFOBIA	14
ANEXO 2. ESTIMACIÓN A PARTIR DE LDA	16

1. RESUMEN EJECUTIVO

El presente informe muestra un análisis de la conversación en línea de xenofobia y lenguaje de odio hacia la población migrante en Panamá. Para realizarlo, se construyó una base de datos que recoge todas las publicaciones de xenofobia y lenguaje de odio en línea hacia la población migrante venezolana, colombiana y nicaragüense entre el 1 de abril de 2020 y el 1 de abril de 2021. Durante este periodo se generaron un total de 2.877 mensajes de este tipo que tuvieron un alcance de alrededor de 6.729.942 usuarios. Esto incluye retweets y compartidos.

El mes con mayor cantidad de publicaciones con contenido xenófobo fue mayo, luego de que un ciudadano venezolano agrediera a un policía. De hecho, la mayoría de las publicaciones de xenofobia en línea estuvieron dirigidas hacia la población migrante venezolana. Al separar las publicaciones de xenofobia por la nacionalidad a que iban dirigidas, se encuentra que el 46% estaban orientada a población migrante venezolana, 26% a población migrante colombiana, 10% a población migrante nicaragüense y 19% no indicaba directamente una nacionalidad.

Durante los brotes más importantes de xenofobia durante el período analizado, las publicaciones con contenido de xenofobia se agruparon en dos conversaciones principales. Primero, la asociación de la migración venezolana al robo de trabajo de personas panameñas. Segundo, se repite constantemente el discurso de que la población migrante venezolana deberían devolverse a su país, estos mensajes van en su mayor parte acompañados de insultos.

Para estudiar a profundidad los distintos componentes de la conversación de xenofobia, se utilizó un modelo de clasificación de Machine Learning (LDA). Este permite clasificar las principales temáticas dentro de todas las publicaciones de xenofobia. Los resultados indican que hubo tres temáticas principales que se diferenciaron del resto. 1. El rechazo a migrantes de nacionalidad venezolana y colombiana indicando que llegan al país a robar. La mayoría de estas publicaciones mostraban una percepción negativa sobre las actividades ilegales que los migrantes estaban haciendo al llegar a Panamá. 2. La idea de que las condiciones a las que se enfrentan las personas migrantes al llegar a países como Panamá son culpa de las malas decisiones que tomaron eligiendo a los gobernantes de su país. En adición, aquí se capturan los mensajes según los cuales la migración colombiana y venezolana está quitándole las oportunidades de empleo a la población panameña. 3. El descontento sobre la situación migratoria del país y la exigencia que hacen los autores de los mensajes al Gobierno Panameño para que tome medidas de mano dura, como un mayor control fronterizo.

Por último, a partir de los hallazgos de este informe se sugieren las siguientes recomendaciones de política:

1. Combatir la percepción de competencia por empleo entre la población migrante y la población de acogida.
2. Tomar en cuenta la población migrante venezolana y colombiana en cualquier esfuerzo por reducir las actitudes en contra de la migración.
3. La identificación oportuna de brotes de xenofobia para alertar a las autoridades y a la población.
4. Proveer herramientas a la ciudadanía para reaccionar ante el aumento del discurso de odio en línea.

El resto del informe sigue de la siguiente manera: En la sección 2 se describen los datos y la metodología utilizada para procesarlos. En la sección 3 se presentan los eventos más importantes en la conversación de xenofobia durante el periodo. En la sección 4 se clasifican y analizan los tópicos dentro de la conversación de xenofobia. En la sección 5 se estudia la nacionalidad a la que van dirigidas las publicaciones de xenofobia. En la sección 6 se concluye y se proponen unas recomendaciones de política.

2. DATOS Y METODOLOGÍA

Los datos utilizados en este informe provienen de la información pública de las redes sociales Facebook, Twitter, Instagram y Youtube en Panamá, entre abril de 2020 - abril de 2021. Para obtener y filtrar los datos se realizó un ejercicio de Web Scraping a través de la plataforma Meltwater. En este sentido, se organizaron y filtraron todos los posts públicos asociados a migración en Panamá. La organización se llevó a cabo a través de una selección de palabras claves, frases y asociaciones que permiten capturar la conversación de migración. Estas palabras y frases clave captan las conversaciones sobre distintas poblaciones migrantes y fueron construidas en conjunto con el equipo local de la OIM en el país y pueden encontrarse en el Anexo 1. En el caso de Panamá, las poblaciones migrantes incluidas en el análisis fueron la venezolana, nicaragüense, colombiana y cubana por ser las mayores comunidades de migrantes en el país.

Una vez capturada y delimitada la conversación de migración, el segundo paso fue cualificar los mensajes para poder identificar las publicaciones de discriminación hacia las poblaciones migrantes. Para esto, se capturó la conversación de discriminación a partir de palabras, frases y expresiones que mostraban una actitud xenófoba hacia la población migrante. Estas palabras, frases y expresiones también fueron trabajadas con el equipo local de OIM. Además, para asegurarse de capturar solo publicaciones de discriminación, se eliminaron todas las publicaciones que denunciaban el comportamiento xenófobo. Los datos finales corresponden a la totalidad de publicaciones en línea que muestran actitudes de discriminación, rechazo o lenguaje de odio hacia alguna de las poblaciones migrantes en Panamá durante el período de estudio. En las siguientes secciones se procede a analizar los contenidos de estas publicaciones.

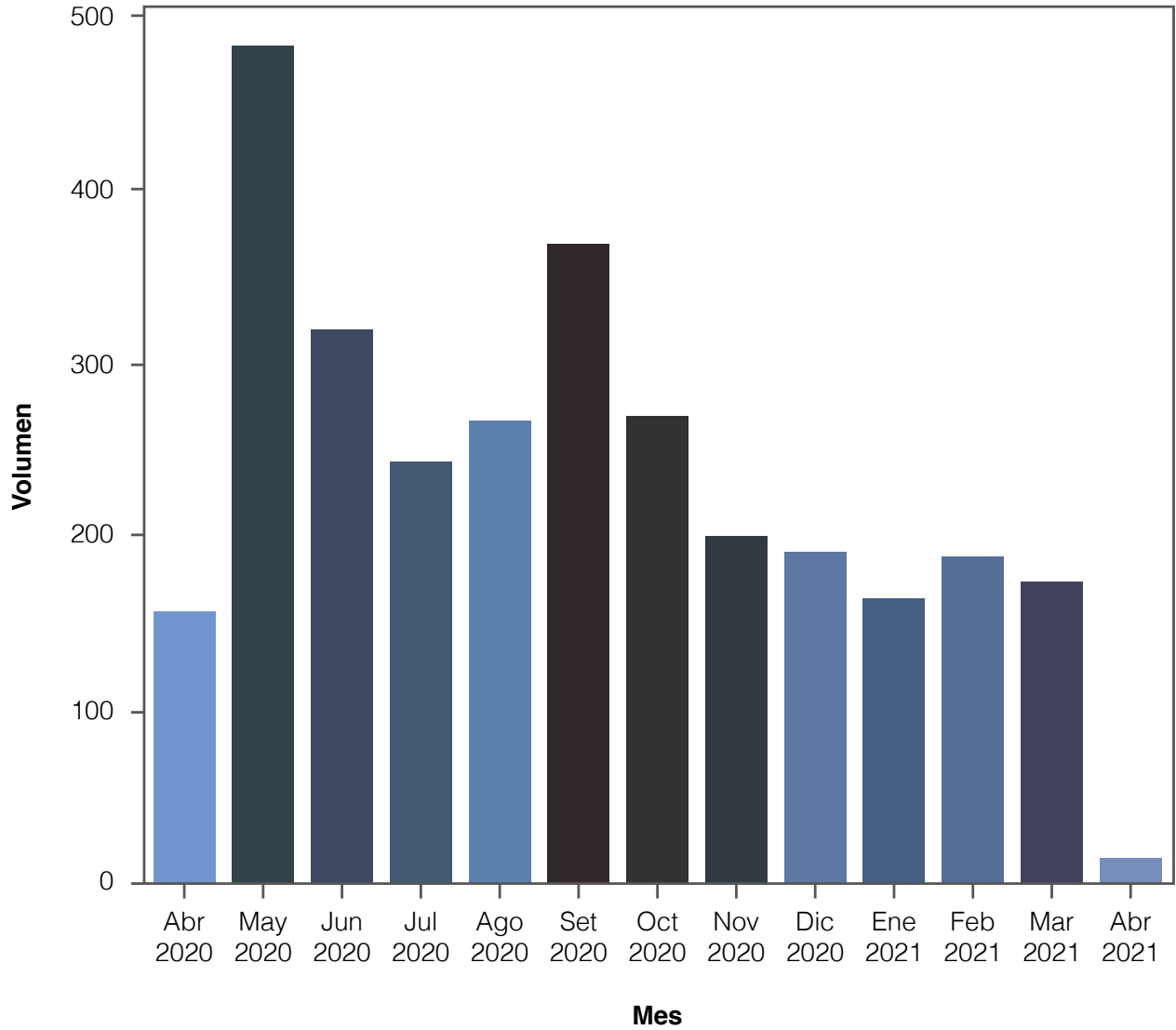
3. LA CONVERSACIÓN DE XENOFOBIA

En esta sección se estudia el comportamiento de la conversación de xenofobia en línea entre el 1 de abril de 2020 y el 1 de abril de 2021. Durante este periodo se generaron un total de 2.877 mensajes de xenofobia en línea. Con el propósito de explicar este comportamiento, primero se discuten los principales eventos que incidieron en esta conversación; segundo, se estudian los tópicos más relevantes dentro de esta conversación; y tercero, se discute la composición por nacionalidad dentro de la conversación de xenofobia.

3.1 EVENTOS MÁS RELEVANTES EN LA CONVERSACIÓN DE XENOFOBIA

Una forma de entender los impulsores de la publicaciones de xenofobia es a partir de los eventos en donde más se generaron ese tipo de publicaciones. Durante el año de estudio, se encontró que hay dos meses que contienen publicaciones de xenofobia por encima del promedio del período. El gráfico 3.1 muestra que los meses con mayor incidencia de xenofobia en línea fueron mayo y septiembre de 2020. A continuación se hace una descripción de la conversación de xenofobia durante estos meses.

Gráfico 3.1 Volumen de Publicaciones de Xenofobia por Mes.



Meltwater (Construcción Propia), 2021.

Mayo 2020

Durante el mes de mayo la conversación de xenofobia se centró en contra de las personas migrantes venezolanas. El día 31 de mayo se registró un incremento de menciones, el cual se originó luego de que un ciudadano venezolano agrediera a un policía, como puede observarse en el cuarto y quinto ejemplo del gráfico 3.2, donde se demuestra un alto nivel de violencia y desprecio hacia la comunidad venezolana. Esta situación hizo que se generaran mensajes xenofóbicos de distinta índole, pudiendo observarse que existe una percepción de que las personas migrantes venezolanas son maleantes.

Es importante mencionar el hecho de que la migración venezolana es vista en este tipo de mensajes como una plaga. Esta situación no se observa solo en el pico registrado el 31 de mayo, sino que a lo largo de todo el mes aparecen mensajes que interpelan a personas de nacionalidad venezolana con la intención de que vuelvan a su país, apelando a que nadie en el mundo les quiere y que le quitan el trabajo a la ciudadanía panameña.

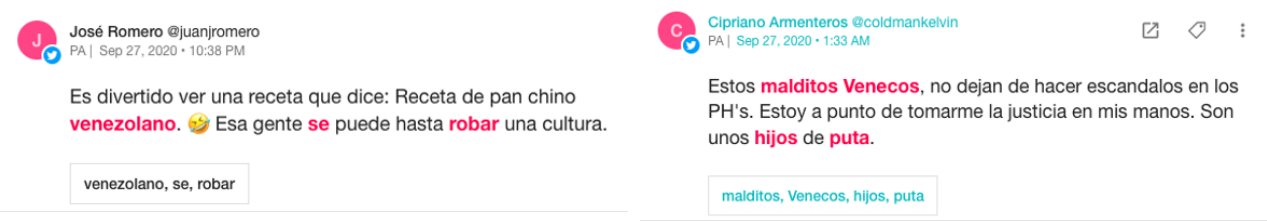
Gráfico 3.2 Ejemplo Mensajes Mayo de 2020



Meltwater, 2021.

Las publicaciones de este mes no tuvieron un estímulo en específico. Sin embargo, siguen la misma línea discursiva de los mensajes observados en mayo, centrándose en contra de la población proveniente de Venezuela. La gráfica 3.3 muestra los mensajes con mayor difusión durante el mes y que resumen las principales temáticas de las que se hablaron.

Gráfico 3.3 Ejemplo Mensajes de Xenofobia Septiembre de 2020



Meltwater, 2021.

Sobre los temas abordados durante la conversación en este periodo, los ejemplos expuestos en el primer, tercer y cuarto mensaje conectan a las personas migrantes con la falta de trabajo y con el hecho de que debieran volver a su país. La primera publicación hace alusión a una empresa venezolana que logró ser proveedora de equipos médicos de la Presidencia de Panamá, mientras que la cuarta se refiere a que las personas migrantes de Venezuela se roban incluso la cultura de un país. También se repiten ofensas hacia esta población, las cuales suelen acompañar los mensajes que buscan que se devuelvan a su país de origen.

También hubo mensajes que se dirigieron a otras comunidades migrantes, como la segunda publicación que se muestra en el gráfico 3.3, la cual apunta a varias nacionalidades que debieran ser deportadas. Sin embargo, cabe destacar que los mensajes hacia personas venezolanas son mayoritarios.

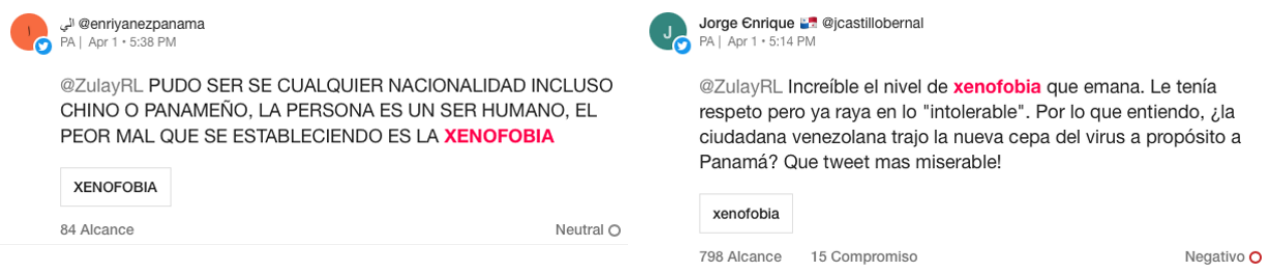
Xenofobia y COVID-19

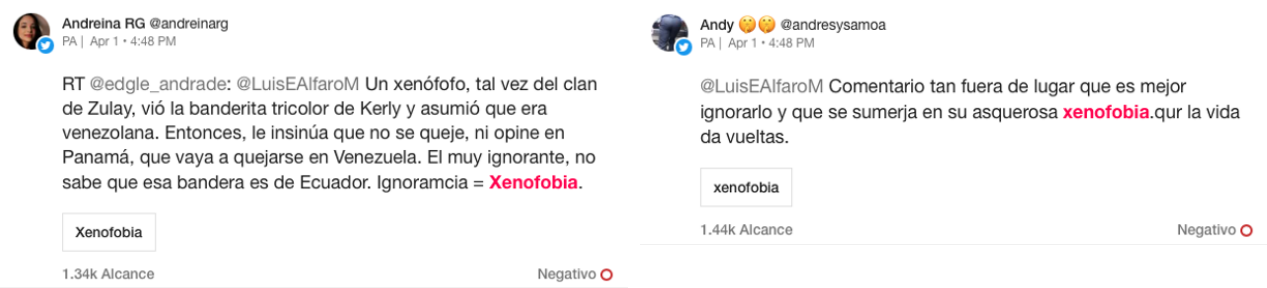
Un posible foco de discriminación hacia la población migrante es la asociación de esta con temáticas relacionadas a la crisis sanitaria originada por el COVID-19. Existen varias razones por las que la crisis sanitaria puede estar asociada a una postura de rechazo hacia la población migrante. Por ejemplo, se puede tener una percepción de que traen consigo la enfermedad desde su país de origen o, por otro lado, una percepción de escasez por la crisis sanitaria y competencia de recursos. En total, los mensajes de discriminación que vincularon a la población migrante con temáticas relacionadas al COVID-19 o la crisis sanitaria representaron un 3% del total de mensajes de xenofobia. Si bien este es un porcentaje relativamente bajo, se requiere más investigación para entender cómo surge y qué forma tiene el vínculo entre los procesos migratorios y esta coyuntura.

Publicaciones de Integración

A lo largo del período de estudio también se generaron publicaciones que denunciaron el comportamiento xenófobo. Estas publicaciones mostraron dinámicas de integración hacia la población migrante y rechazaron su estigmatización. Durante el período se generaron 45 mil publicaciones de integración. Esto representa más de diez veces la cantidad de publicaciones con contenido xenófobo. El gráfico 3.5 muestra ejemplos representativos de las publicaciones de integración.

Gráfico 3.5 Ejemplo de publicaciones de integración





4. CLASIFICACIÓN DE LA CONVERSACIÓN DE XENOFOBIA

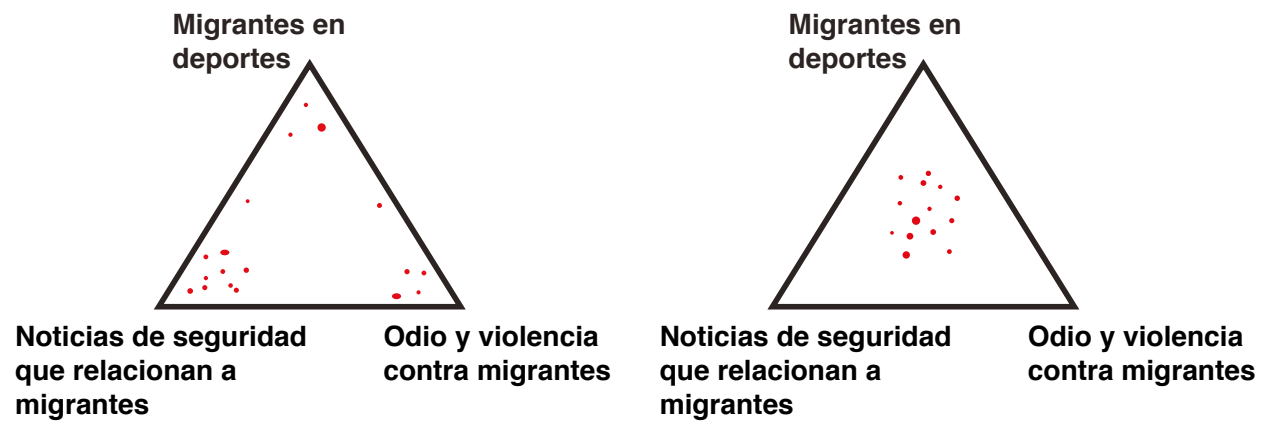
En la sección anterior se discutieron los momentos más importantes de la conversación de xenofobia, así como los impulsores de estos. Sin embargo, otra aproximación para entender el contenido dentro de las publicaciones de xenofobia es estudiando los tópicos o temas utilizados dentro de esta conversación. Existen modelos de aprendizaje de máquinas que permiten identificar las temáticas utilizadas dentro de una conversación. En esta sección se sugiere la utilización de un modelo LDA (Latent Dirichlet Allocation) y se muestran los resultados para la clasificación de la conversación de xenofobia en Panamá.

4.1 MODELO LDA

Para poder estudiar las temáticas más frecuentes dentro de la conversación de xenofobia en Panamá, se utiliza un modelo de tópicos. Este modelo también es conocido bajo el acrónimo LDA (Latent Dirichlet Allocation) en honor a la distribución de Dirichlet, la cual permite estimar la probabilidad de que cada publicación o documento pertenezca a un tópico determinado. La intuición detrás de este modelo resulta de tomar una colección de publicaciones (**corpus de documentos**) en la que pueden existir varios tópicos que se repiten entre ellas. En este sentido, se espera que las palabras que aparecen frecuentemente juntas tienen significados parecidos. A estos *significados parecidos* se les llama tópicos. Cada **tópico** es una combinación de diferentes **palabras** y cada **documento** es una combinación de diferentes **tópicos**.

El LDA es un modelo que intenta determinar el mecanismo subyacente que genera las publicaciones de xenofobia y sus tópicos. Una manera de interpretar el modelo es imaginando que todas las publicaciones de xenofobia son generadas por una máquina, pero no podemos observar dicha máquina ni sus criterios, solo lo que ella produce. LDA simula un conjunto de máquinas con diferentes criterios y selecciona la que mejor se acomoda a los resultados de las publicaciones de xenofobia evidenciados. El gráfico 4.1 muestra la forma en que se pueden separar los tópicos dentro de una conversación con dos ejemplos. En uno los tópicos están separados entre sí (izquierda) y en otro los tópicos se encuentran más juntos (derecha). Entre más separados los tópicos más fácil es analizar su contenido como una temática única.

Gráfico 4.1: Ejemplo de la separabilidad entre tópicos

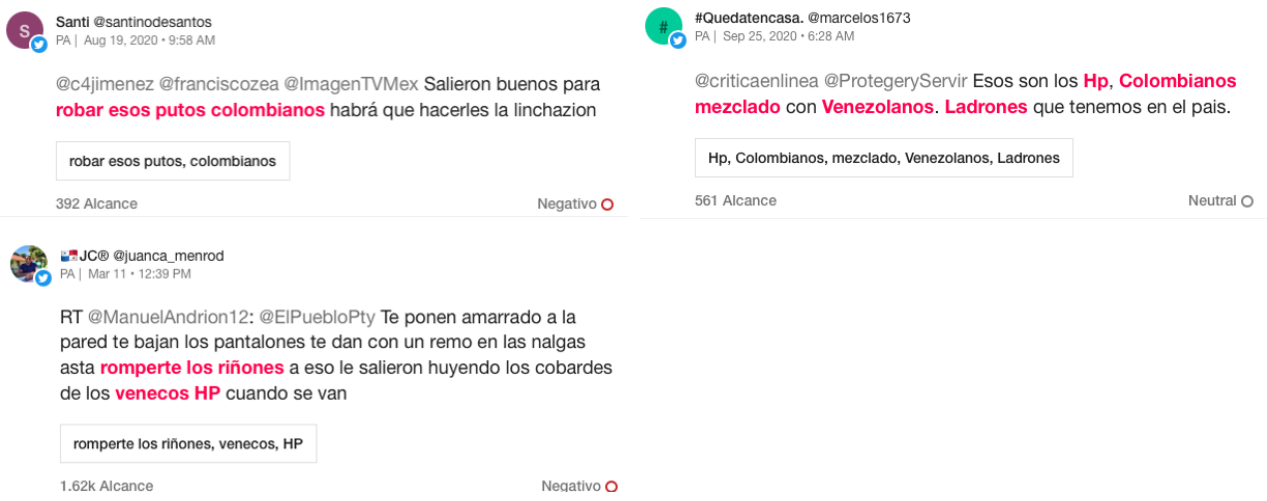


4.2 RESULTADOS

Los resultados de la metodología propuesta sugieren que la conversación de xenofobia estuvo organizada en ocho tópicos diferentes. El primer tópico relacionaba a las personas migrantes con el crimen, en particular el robo. El segundo tópico tuvo que ver con la migración venezolana y la idea de que las decisiones políticas son aquellas que les llevan a afrontar dificultades y cometer delitos en países como Panamá. El tercer tópico tuvo que ver con una exigencia de mayores controles migratorios en las fronteras y una respuesta más fuerte por parte del gobierno panameño ante la llegada de población migrante al país. El contenido del resto de los tópicos no se diferencia entre sí e incluye temas varios como participación de migrantes en crimen, burlas, ofensas a la población migrante y ofensas contra mujeres migrantes. El gráfico 4.6 del anexo muestra la forma en que se distribuyen los tópicos dentro de esta conversación.

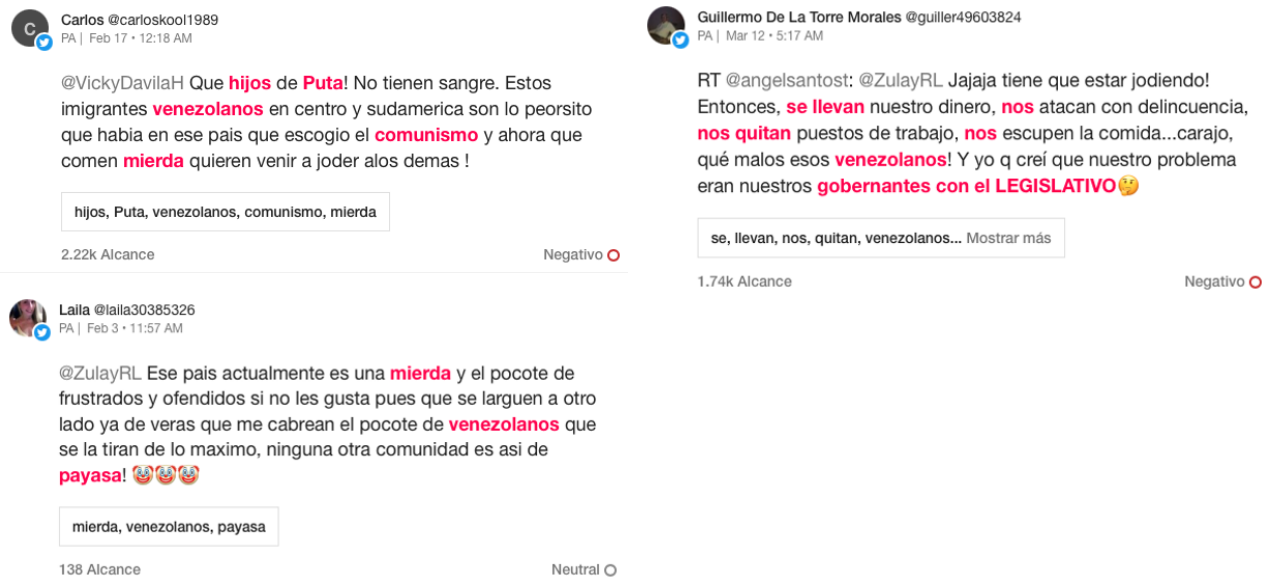
El tópico que más agrupó publicaciones de xenofobia (tópico 1) tuvo que ver con mensajes que asociaban a migrantes de Venezuela con hechos delictivos. Esta parte de la conversación se centró principalmente en un rechazo a las personas migrantes venezolanas indicando que llegan al país a robar. Ahora bien, aun cuando los mensajes de este tópico no responden a una noticia o hecho en particular, se observa que este tópico agrupa los mensajes cuyo contenido corresponde a ofensas directas o insultos. En el gráfico 4.7 se pueden observar los principales términos que contiene el tópico en cuestión.

Gráfico 4.7: Distribución general de términos en el tópico 1



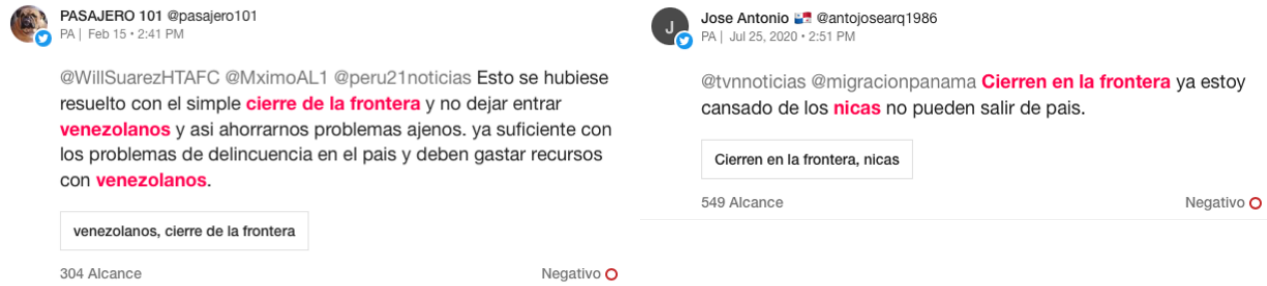
El segundo tópico que más agrupó publicaciones (tópico 2) engloba una conversación que relaciona la migración venezolana con la política. En particular, los mensajes aquí encontrados responden a que algunos autores tienen la idea de que las condiciones a las que se enfrentan las personas migrantes al llegar a países como Panamá son culpa de las malas decisiones que tomaron eligiendo a los gobernantes de su país. Adicional a esta conversación, se encuentra que este tópico condensa aquellos mensajes según los cuales la migración colombiana y venezolana está quitándole las oportunidades de empleo a la población panameña. En el gráfico 4.8 se pueden observar algunos ejemplos representativos del tópico en cuestión.

Gráfico 4.8: Distribución general de términos en el tópico 2



Por último, otro tópico interesante (tópico 3) mostró el descontento sobre la situación migratoria del país y la exigencia que hacen los autores de los mensajes al gobierno panameño para que tome medidas de mano dura, como un mayor control fronterizo. Las publicaciones dentro de este tópico exigían políticas de control hacia la situación migratoria en general, sin embargo, es importante resaltar que en este tópico se incluyeron publicaciones contra la nacionalidad nicaragüense, puesto que en los tópicos analizados hasta el momento las publicaciones se habían centrado sobre las nacionalidades venezolana y colombiana. En el gráfico 4.9 se pueden observar ejemplos representativos de publicaciones en este tópico.

Gráfico 4.9: Distribución general de términos en el tópico 3



Palabras y frases más frecuentes en las publicaciones de xenofobia

Además de los tópicos dentro de la conversación de xenofobia, un aspecto interesante a detallar es la forma que tienen los comentarios de xenofobia. Es decir, si existen algunas expresiones o palabras que sean usadas recurretemente dentro del lenguaje de odio y discriminación. El gráfico 4.10 es una nube de palabras que clasifica las palabras según la frecuencia con la que fueron usadas a lo largo del periodo de análisis. Entre mayor sea el tamaño de la frase o palabra dentro de la nube, fue más alto el número de veces que se usó dentro de la conversación.

Gráfico 4.10 Nube de palabras de la Conversación de Xenofobia

Meltwater, 2021.

Las frases más utilizadas fueron “pedazo de idiota”, “colombianito de mierda” y “venecos de mierda”. Esto muestra la forma en que se mueve el discurso de odio hacia la población migrante. Un hecho interesante es que la mayoría de los mensajes encontrados se desarrollan en torno al descontento que hay por la percepción de que migrantes de Colombia y Venezuela llegan a invadir el país, quitarle recursos a la población y cometer actos delictivos en Panamá.

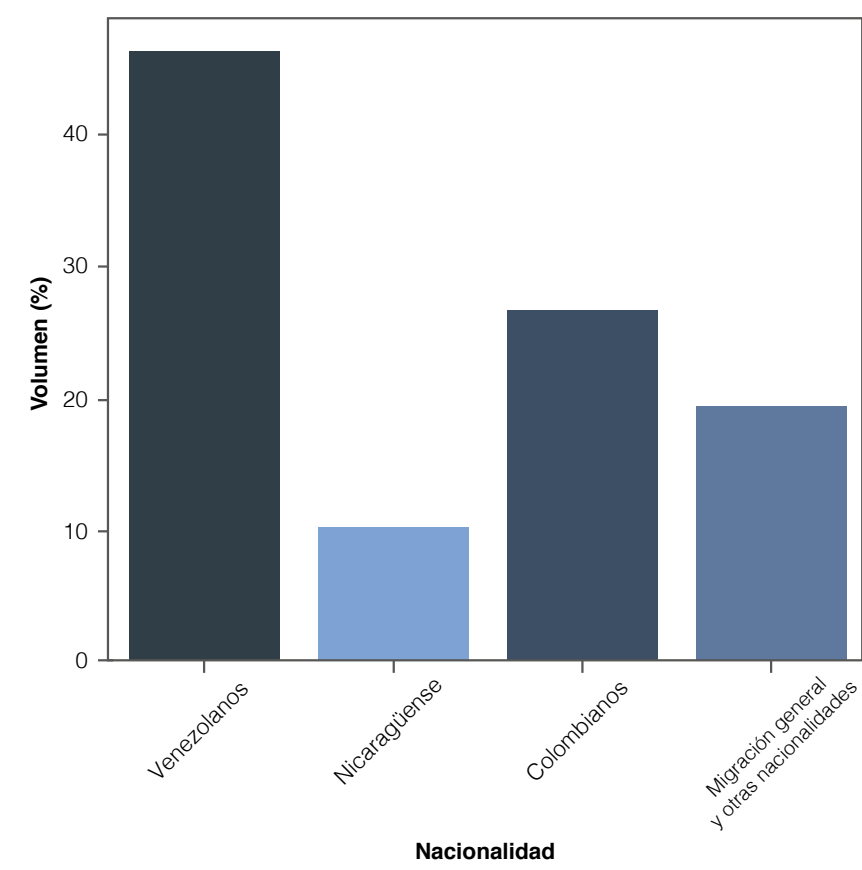
Es importante destacar la relación que los autores de estos mensajes hacen entre las personas migrantes y los problemas del país. En particular, la frase “graves problemas” es ampliamente utilizada y genera la percepción de que son las personas colombianas y venezolanas quienes tienen la culpa de los problemas sociales y de seguridad que puede estar experimentando el país. Por otro lado, la frase “maldita lacra” genera, de manera despectiva, la idea de que las personas migrantes están quitando oportunidades a la población y empeorando su calidad de vida. Ligado a esta frase se encuentra “muertos de hambre”. Bajo este concepto los autores de los mensajes generan la idea de que la población migrante que llega al país no tiene recursos para poder suplir necesidades alimentarias y esto despierta la percepción de mayor pobreza asociada a estos.

5. CUANTIFICACIÓN DE XENOFOBIA POR NACIONALIDAD

Por último, un elemento importante en el análisis de la conversación de xenofobia en línea es entender la población a la que van dirigidos. En específico, dentro de Panamá existen migrantes de distintas nacionalidades, siendo las poblaciones migrantes venezolanos, nicaragüenses y colombianos las más importantes. En esta sección se estudia la distribución de las publicaciones de xenofobia por nacionalidad a la que van dirigidos.

Mediante la aplicación de análisis de texto y, particularmente expresiones regulares, es posible establecer la nacionalidad a la que van dirigidas las publicaciones de xenofobia. La forma de hacer esto es mediante la búsqueda de palabras o frases que denoten nacionalidad dentro del mensaje de xenofobia. Cabe resaltar que algunas publicaciones discriminatorias no expresan directamente la población a que van dirigidas, es decir, están orientadas a la población migrante en general, o van dirigidas a nacionalidades que no son relevantes dentro del estudio; estas publicaciones son clasificadas como “Sin Definir”.

Gráfico 5.1 Volumen de Xenofobia por Nacionalidad

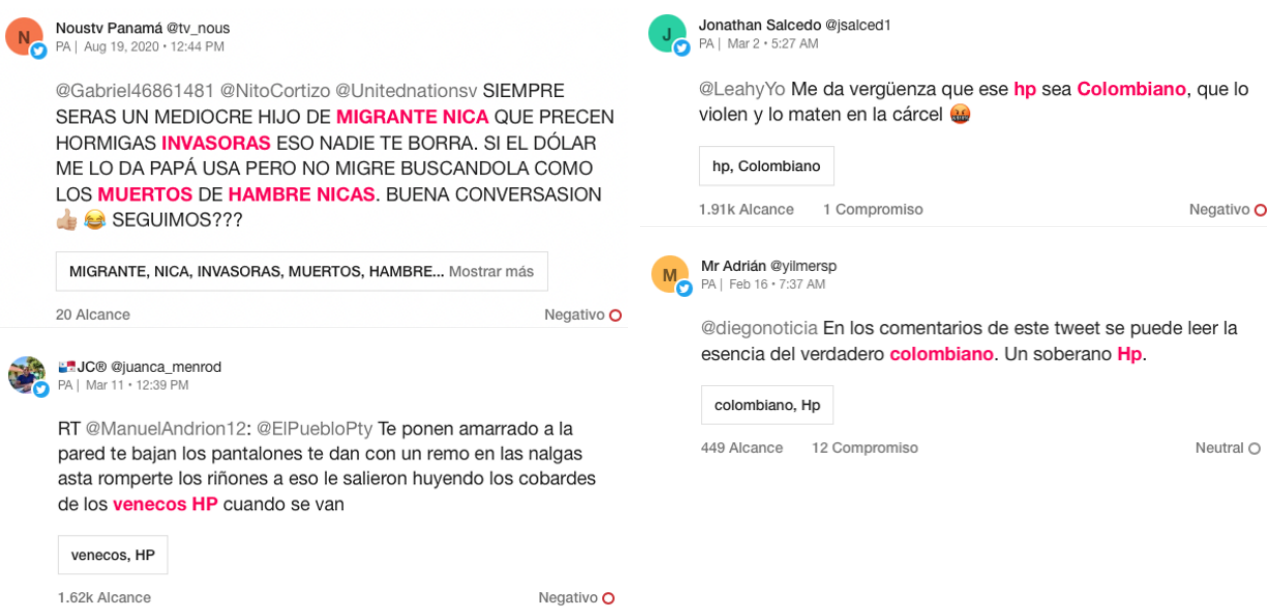


Meltwater (Construcción Propia), 2021.

El gráfico 5.1 muestra el porcentaje de mensajes que se refirieron a migrantes de nacionalidad venezolana, colombiana o nicaragüense. Salta a la vista que los mensajes xenófobos hacia la población venezolana corresponden a un 45.07%, siendo la nacionalidad con la mayor participación entre las incluidas en el análisis. En segundo lugar se encuentra la población colombiana con 26.53%. Adicionalmente, el 10.05%

de los mensajes de xenofobia fueron en contra de la población nicaragüense y cerca del 19% de los mensajes no se definen dentro de ninguna de las nacionalidades anteriormente descritas.

Gráfico 5.2 Ejemplo Mensajes de Xenofobia por Nacionalidad



Meltwater, 2021.

Un ejemplo del porcentaje de la conversación de xenofobia dirigida a cada nacionalidad se puede observar en el gráfico 5.2, así como en la sección 3.1 donde se observaba que los mensajes producidos durante los picos de mayor volumen en la conversación se encuentran dirigidos en su mayoría en contra de personas migrantes venezolanas, por sobre el resto de nacionalidades.

6. CONCLUSIONES Y RECOMENDACIONES DE POLÍTICA

El análisis de las publicaciones de xenofobia y lenguaje de odio en línea arroja varios hallazgos que pueden ser utilizados para mitigar los riesgos asociados al incremento de la xenofobia.

Primero, existen coyunturas en las que las publicaciones de odio hacia la población migrante incrementan significativamente. Durante el período de estudio estas coyunturas estuvieron asociadas a la publicación de noticias que relacionaban a migrantes con crímenes violentos y al anuncio de políticas públicas, como la atención gratuita del parto a mujeres migrantes. Segundo, el rechazo hacia la población migrante parece estar agrupado en prejuicios y mitos. Además del rechazo hacia la atención del parto de las migrantes, otros prejuicios incluyen la percepción de la población migrante haitiana como una amenaza para la seguridad de la nación y el supuesto obstáculo que representa la comunidad internacional para tomar políticas de mano dura. Para mitigar el aumento de la discriminación hay que tomar en cuenta estos prejuicios. Teniendo en cuenta lo anterior, se proponen las siguientes recomendaciones de política.

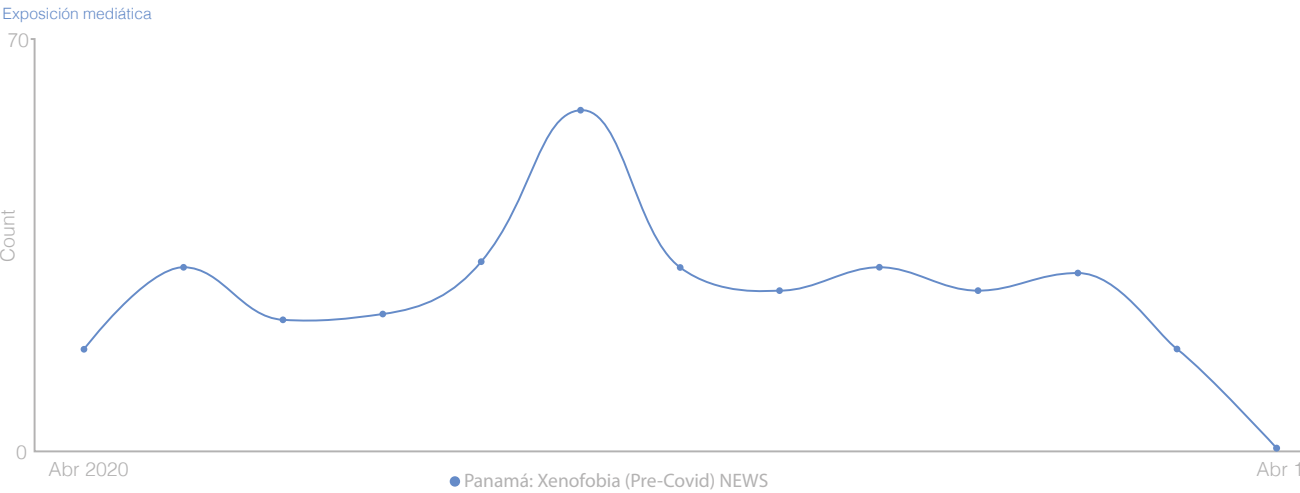
Recomendaciones

- Una parte importante de las publicaciones de xenofobia tienen que ver con la percepción de competencia de empleos. Es importante que estas preocupaciones sean atendidas para disminuir la percepción de injusticia en la atención del fenómeno migratorio.
- Las publicaciones de xenofobia van dirigidas mayoritariamente hacia la población migrante venezolana y colombiana. Los esfuerzos por reducir las actitudes contra las personas migrantes deben considerar ambas poblaciones.
- En coyunturas de brotes de xenofobia, como el ocurrido en septiembre, aumenta la probabilidad de que se vulneren los derechos de la población migrante. Es importante que se identifiquen estos brotes oportunamente para poder alertar a las autoridades y a la población en general sobre los aumentos en el discurso de odio y sus riesgos.
- Deben proveer herramientas a la ciudadanía para reaccionar ante el aumento del discurso de odio en línea. Es importante evitar un escalamiento en la violencia en línea que polarice aún más la conversación sobre migración.

ANEXO 1. NOTICIAS Y XENOFOBIA

Durante el periodo estudiado, pudimos evidenciar que se han generado 23.812 publicaciones sobre migrantes, con un promedio de 65 publicaciones diarias. Sin embargo, no todas estas son catalogadas como xenófobas. Luego de realizar una nueva regla, se pudo apreciar que solo un 1,5% del total de publicaciones fueron con rasgos xenófobos. Es decir, 346 publicaciones, lo que implica un promedio de 0,96 publicaciones diarias.

Gráfico de volumen por mes, publicaciones xenófobas en medios.



Meltwater, 2021.

El peak que se genera en el mes de Septiembre responde a bandas de venezolanos que se dedican al robo de computadores y otras piezas de automóviles.. Otros ejemplos de noticias con rasgos xenófobos son:

Nuestro país, incluidos malandrines | La Prensa Panamá

Desdén, rencor. Pena y hasta vergüenza de habitar este pago. Así percibe este psicólogo y educador la expresión de alta densidad "este país ... la que se ha nacido. 'Este país' puede ser Ticolandia o **Nicaragua**. Alguien admite que no quiere ensuciarse con tanto truhan callejero, sobre

Doce migrantes haitianos son procesados en Panamá por vandalizar un albergue

La defensa de los imputados anunció un recurso de apelación cuya audiencia se realizará el 12 de agosto, de acuerdo con la información ... en el albergue de La Peña luego de que un grupo de **migrantes amenazaran** con prender fuego al lugar si no los dejaban continuar su viaje.

Residentes de La Peña en Darién se sienten amenazados

Los propios residentes y autoridades locales nos han hecho llegar sus testimonios, luego de vivir momentos de angustia frente a las ... de angustia frente a las acciones de **vandalismo** y violencia de los **migrantes** que permanecen en el campamento de **refugiados** de La Peña, en



Imponen medida cautelar a haitianos migrantes por vandalismo

El Ministerio Público MP informó que la Fiscalía Regional de Darién logró la medida cautelar de detención provisional para 12 migrantes ... albergue para migrantes **Migrantes** en Darién dan ultimátum al gobierno y **amenazan** con protestas Panamá podría expulsar a **migrantes** por actos

Colombiano que mató a su cuñada, saldrá de prisión en el año 2058. Lo absolvieron de un tercer delito

El atroz femicidio de la chiricana Leticia Lalyre Serracín en La Locería ya tiene un culpable 70 años de edad, en la actualidad tiene tan solo 32. Este sujeto **colombiano** es el único responsable de la muerte Leticia Lalyre Serracín, en



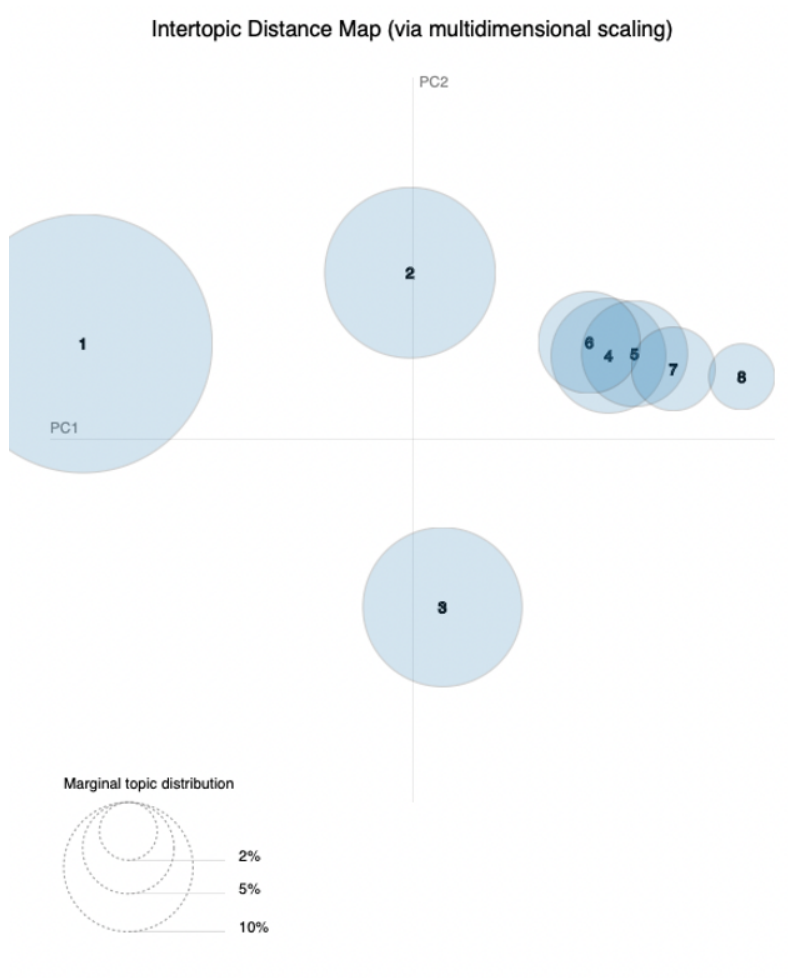
'Colombiana' presa por crimen y tentativa en El Chumical

El Juez de Garantías, Frank Torres Ruiz, decretó la medida cautelar personal de detención provisional a Leydi Guevara, de 31 años y de ... Elida Mc Gittens, respectivamente. El hecho se dio cuando la **'Colombiana'**, supuestamente, ordenó que los sicarios conocidos como 'Russo' y



ANEXO 2. ESTIMACIÓN A PARTIR DE LDA

Gráfico 4.6: Distribución general de tópicos dentro de la conversación de xenofobia



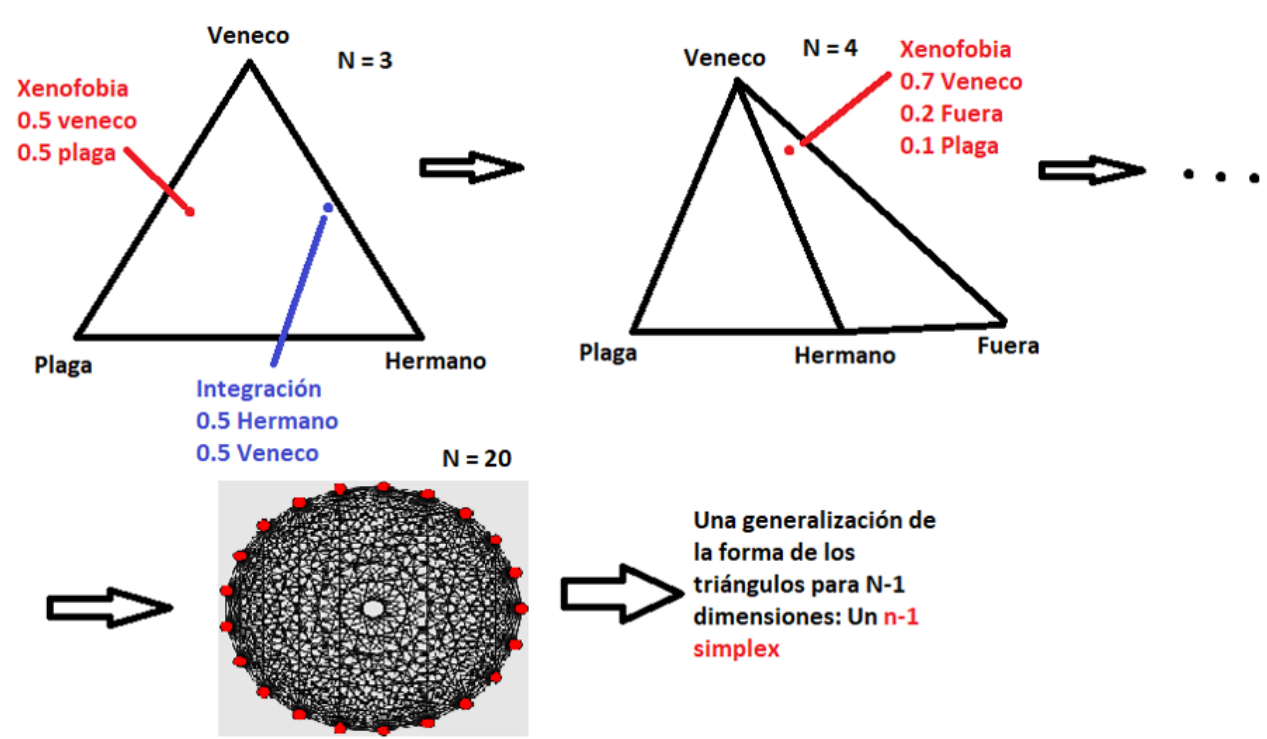
Fuente: Creación propia usando pyLDAvis. Los números representan el orden de los tópicos en términos de número de publicaciones. Es decir, el tópico 1 es el que agrupó más publicaciones y el tópico 8 el que menos. La interpretación del contenido de cada tópico es subjetiva al investigador. Esto quiere decir que, si bien el modelo propone la organización de unos tópicos basadas en las palabras que contienen cada Tweet, es trabajo del investigador estudiar las publicaciones que entraron dentro de cada grupo para determinar *de qué se trata* el tópico.

Una vez identificados los tópicos se puede calcular qué tan diferentes son entre ellos a partir del prior de Dirichlet. Esto es una métrica que nos permite entender qué tan separados o mixtos están los tópicos entre ellos usando como referencia los documentos en su interior. La distribución de Dirichlet suele representarse como un triángulo equilátero.

Como muestra la gráfica 4.1, a la derecha se encuentra un triángulo en donde cada uno de sus vértices es un tópico que puede aparecer en la conversación capturada a través de Meltwater. Los puntos que están al interior de este triángulo son los documentos o publicaciones que resultan de esta captura. Se puede observar como estos puntos están agrupados en el centro del triángulo, esto quiere decir las publicaciones están distribuidas de forma similar entre tópicos y que entre ellas existen diferencias reducidas (A esto le llamamos un alto prior de Dirichlet; $\alpha > 1$). Por su parte, el triángulo ubicado a la izquierda describe una

situación en donde las publicaciones pertenecen con mayor probabilidad a un tópico específico y están separadas fuertemente entre ellas (un bajo prior de Dirichlet; $\alpha < 1$). Existen entonces, por su parte, palabras que pertenecen a varios tópicos (se conoce como el prior de Dirichlet sobre los términos; β); LDA permite establecer cuales son las palabras más usadas por tópico y con esa información entender la clasificación que está realizando sobre las publicaciones. Sin embargo, esto genera un problema de visualización, la idea de poseer una combinación lineal de tópicos por palabra supone que, si el diccionario es lo suficientemente grande, dejemos de visualizar ese triángulo equilátero en donde cada esquina es una palabra (véase gráfico 4.2).

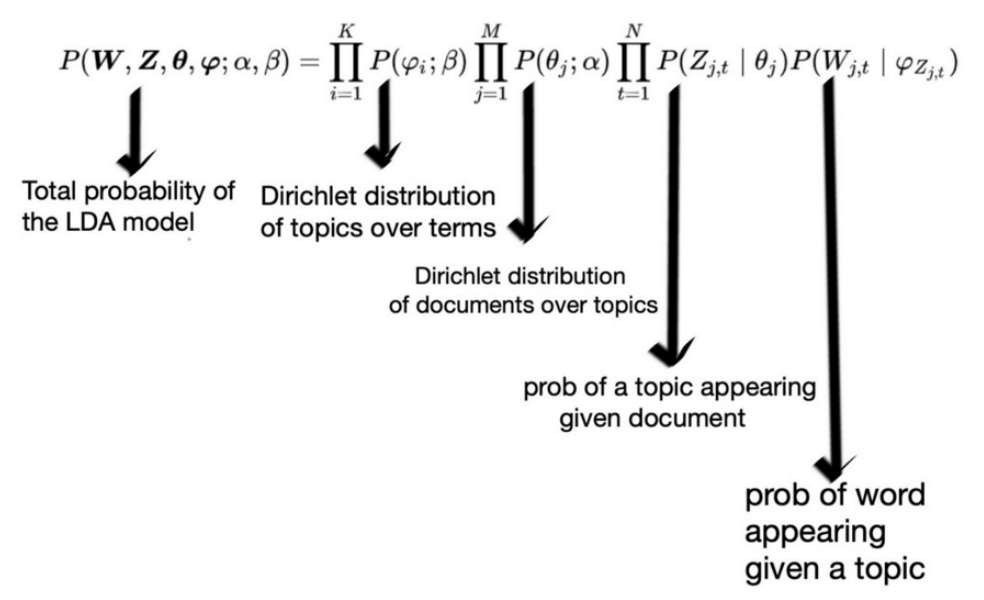
Gráfico 4.2: Clasificación de tópicos con un diccionario creciente



El gráfico 4.2 muestra que sucede con la clasificación de los tópicos una vez el número de palabras totales que constituyen el diccionario crecen. En él se puede encontrar como cada tópico se sitúa en un espacio de términos y como la combinación lineal de esos términos genera un tópico. Por ejemplo, el tópico de xenofobia cuando $N = 3$, se compone de forma equitativa por las publicaciones de 3 palabras que contienen solo “veneco” y “plaga”, una vez el número de palabras incrementa, la complejidad de las combinaciones lineales a su vez incrementa. El concepto de simplex resume esta idea del prior de Dirichlet para $N-1$ dimensiones (en este caso palabras), esto es gracias a que podemos asegurar una generalización de la noción de triángulo o tetraedro para dimensiones arbitrarias.

El modelo LDA en su forma base es una combinación de probabilidades en donde cada una sigue un grupo de distribuciones y un conjunto específico de parámetros que dependen de su jerarquía.

Gráfico 4.3: Ecuación de la probabilidad total modelo LDA



Fuente:<https://towardsdatascience.com/latent-dirichlet-allocation-intuition-math-implementation-and-visualisation-63ccb616e094>

De izquierda a derecha, se tiene la probabilidad total del modelo, la distribución de Dirichlet de los tópicos sobre los términos (note el beta en el paréntesis), la distribución de los documentos sobre los tópicos (note el alpha en el paréntesis), la probabilidad de que un tópico aparezca en un documento dado y la probabilidad de que una palabra aparezca dado un tópico determinado.

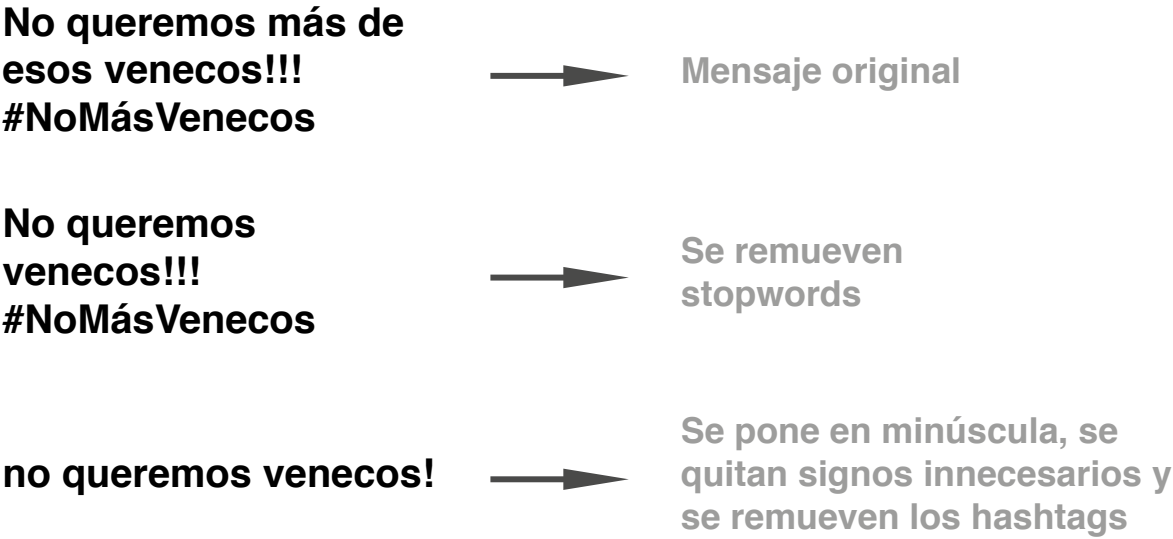
Una vez establecido el modelo se comienzan a optimizar los hiper parámetros del mismo, la idea es encontrar una combinación de beta, alpha y el número de tópicos que maximice las métricas de desempeño del modelo, para este estudio, se usarán las métricas coherence score y perplexity score que son usuales en la literatura para evaluar el desempeño de los modelos LDA y así obtener una combinación óptima orientada a una clasificación eficiente. Una vez realizada la optimización, se procede a interpretar los resultados del modelo y a establecer una marcación de tópicos por documento.

Procesamiento de Datos

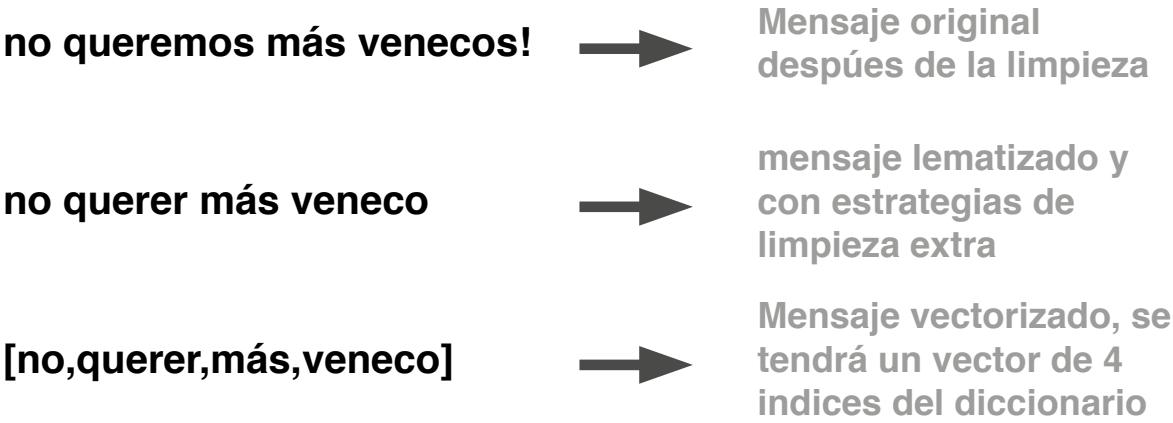
Una vez obtenida la base de datos con cada una de las publicaciones de los usuarios, el objetivo es analizar el contenido de estas. Una forma de realizar el análisis es a partir de técnicas de inteligencia artificial y modelos de clasificación que permiten entender la forma en que se relacionan los mensajes. Para poder utilizar estos modelos, es necesario procesar el contenido de las publicaciones con el fin de hacerlas comparables y eliminar el ruido. Este proceso se llama vectorización de las publicaciones, y tiene como objetivo obtener una representación única y coherente de cada mensaje.

El primer paso de la vectorización busca eliminar palabras que no interesan para definir relaciones entre publicaciones. Consiste de distintas técnicas de limpieza como la eliminación de OOV (palabras fuera del vocabulario o Out-Of-Vocabulary words), stopwords (conectores lexicográficos) y las OUW (palabras demasiado usadas o Over-Used-Words). El segundo paso consiste en eliminar elementos dentro de las publicaciones que sean diferentes a contenido de léxico, ya que esto puede ensuciar la clasificación o las relaciones que encontremos entre los documentos. Siguiendo esta idea, se eliminan links (HTTP, HTTPS), hashtags, signos de puntuación innecesarios, dobles espacios, letras repetidas y letras mayúsculas. A continuación, se presentan ejemplos de cómo se verían los mensajes luego de esta limpieza y de algunas palabras que recaen en estas categorías de eliminación.

Gráfico 4.4: Ejemplo de limpieza de un mensaje de xenofobia



Una vez realizada esta primera limpieza, se propone construir una especie de diccionario con las palabras que contiene cada texto. Para esto, debemos introducir el concepto de Bag-Of-Words (BOW o bolsa de palabras), el cual sugiere que cada palabra puede ser representada con un índice que la posiciona de forma única en la bolsa (diccionario). Previo a esto, se requiere convertir palabras en forma flexionada a su lema correspondiente, este proceso lingüístico se conoce como lematización y es ampliamente usado en el procesamiento del lenguaje natural para interpretar y representar textos. De igual manera, la técnica conocida como “Stemming” permite reducir las palabras a su raíz correspondiente, la cual nos da la oportunidad de encontrar mejores relaciones entre textos, reduciendo palabras como “bibliotecario” y “biblioteca” a una raíz conocida equivalente “biblioteca”.



Una vez construido el diccionario, se procede a construir los documentos que contienen las palabras modificadas. Un documento es un conjunto de vectores que comprenden la estructura numérica de las palabras contenidas en el mismo. Un uso importante de estos documentos es que permiten evidenciar n-gramas (subsecuencias de palabras) de las publicaciones por cada uno de los autores y así obtener diferentes representaciones de un mensaje. En la siguiente sección veremos cómo vamos a usar estos documentos para clasificar los mensajes basados en sus relaciones, para este informe, lo llamaremos modelamiento de tópicos.