

UNIVERSIDAD SAN FRANCISCO DE QUITO

**LA CENTRALIDAD CON RESPECTO A LA PRENSA ESCRITA DE LOS
PRESIDENTES DEL ECUADOR COMO CATALIZADOR DE LA OPINIÓN
PUBLICA.**

Juan Aurelio Alvarado Ortega.

Tesis de grado presentada como requisito para la obtención del título de
MASTER EN MATEMÁTICAS APLICADAS.

Octubre de 2010

Universidad San Francisco de Quito
Colegio de Graduados
HOJA DE APROBACIÓN DE TESIS

LA CENTRALIDAD CON RESPECTO A LA PRENSA ESCRITA DE LOS
PRESIDENTES DEL ECUADOR COMO CATALIZADOR DE LA OPINIÓN
PUBLICA.

Juan Aurelio Alvarado Ortega

Carlos Jiménez, Ph.D.
Director de Tesis (firma)

Pedro Romero, Ph.D.
Miembro del Comité de Tesis

Carlos Jiménez, Ph.D.
Director de la Maestría en M.A. (firma)

MSc. Fernando Romo P.
Decano del Colegio de Ciencias e Ingeniería (firma)

Victor Viteri, Ph.D.
Decano del Colegio de Graduados (firma)

Quito, Octubre de 2010

©Derechos de autor

Juan Aurelio Alvarado Ortega

2010

La centralidad de los presidentes de la República del Ecuador con respecto a la prensa escrita

Juan A. Alvarado O.

4 de octubre de 2010

Índice

1. Introducción	6
2. Grafos de palabras	7
2.1. Textos como grafos de palabras	7
2.2. Índices de Centralidad en el grafo	8
2.3. Algoritmo para el cálculo de la centralidad de intermediación	11
3. Procesamiento textual	14
3.1. Determinación de las frases significativas en las noticias	14
3.1.1. Criterios de importancia de las co-ocurrencias	15
3.1.2. Palabras comunes (Stopwords)	16
3.1.3. Stemming	16
3.1.4. Eliminación de palabras iniciales no sustantivas en la colocaciones	17
3.2. Elaboración del grafo de palabras	18
4. Herramientas de Análisis de Datos	18
4.1. Modelos ARIMA	18
4.1.1. Identificación del modelo	19
4.1.2. Ajuste del modelo	20
4.1.3. Validación del modelo	20
4.2. Criterio AIC	20
4.3. Determinación del modelo ARIMA usando el criterio ACI	21
4.4. Dinámica comunicacional y su perturbaciones	22

4.5.	Regresión lineal, bajo heterocedasticidad y autocorrelación	22
4.5.1.	Regresión lineal	22
4.5.2.	Regresión lineal, bajo heterocedasticidad	23
4.5.3.	Cuando existe heterocedasticidad y autocorrelación	24
5.	Resultados obtenidos	24
5.1.	Descripción de los datos	24
5.2.	Procesamiento de las noticias	25
5.2.1.	Extracción de las noticias del sitio www.eluniverso.com	25
5.2.2.	Extracción del texto noticioso de cada pagina Web	26
5.2.3.	Steeming y eliminación de palabras comunes	26
5.2.4.	Determinación de las co-ocurrencias significativos e importantes	26
5.2.5.	Determinación del grafo semanal de temas importantes	27
5.2.6.	Calculo de índice de centralidad	27
5.3.	Centralidad y Popularidad del presidente	27
5.3.1.	Centralidad	27
5.4.	Determinación de los valores aberrantes y la dinámica comunicacional	28
5.5.	Análisis de los residuos de los modelos de centralidad obtenidos	33
6.	Modelamiento de la relación entre centralidad y popularidad	33
6.1.	Popularidad y Componentes de la centralidad	37
6.2.	Modelo de la popularidad de los presidentes	37
6.2.1.	Selección de variables en el modelo centralidad-popularidad	42
6.2.2.	Modelo para la popularidad de Gustavo Noboa	42
6.2.3.	Modelo para la popularidad de Lucio Gutiérrez	45
6.2.4.	Modelo para la popularidad de Alfredo Palacio	48
6.2.5.	Modelo para la popularidad de Rafael Correa	49
7.	Conclusiones	54
8.	Recomendaciones	54
9.	Anexo	56
9.1.	Códigos en R	56
9.1.1.	Función selectora de modelos ARIMA	56
9.1.2.	Función selectora de modelos lineales	57
9.2.	Datos	58

9.2.1. Centralidad, ordenadas de izquierda a derecha	58
9.2.2. Popularidad en el Ecuador, ordenadas de izquierda a derecha	59
9.2.3. Popularidad en Quito, ordenadas de izquierda a derecha	60
9.2.4. Popularidad en Guayaquil, ordenadas de izquierda a derecha	61

1. Introducción

En el Ecuador desde los últimos 13 años hubo 3 derrocamientos de presidentes constitucionales y ha existido en ese periodo de tiempo 8 jefes de Estado, entre presidentes constitucionales e interinos. Esto hace que se lo perciba al Ecuador como un país políticamente inestable producto de la pugna congreso-ejecutivo, la cual ha originado la ruptura en la continuidad de los presidentes constitucionales por tres ocasiones.

La motivación principal para este trabajo fue explicar cuantitativamente la relación que existe entre la popularidad del presidente y las noticias que aparecen en la prensa escrita. Esta cuantificación es a través del cálculo de un índice de centralidad que tuvieron los presidentes de la República con respecto a los temas importantes que parecieron en periódicos, desde Gustavo Noboa en Marzo del 2002 hasta Rafael Correa en Diciembre del 2008. Se tomo como fuente de textos noticiosos, aquellas noticias de las secciones de política y economía que aparecen en el sitio web www.eluniverso.com, el cual corresponde al diario, “El Universo” de la ciudad de Guayaquil, que es el diario de mayor circulación en el Ecuador.

Para realizar esta cuantificación de centralidad se transformó las noticias de cada semana, de las secciones de política y economía, en un grafo de frases o palabras que representa temas importantes en la semana [4]. Dos nodos o dos frases importantes en la semana están relacionados si aparecen las dos frases dentro de una misma oración.

Estructurar las noticias semanales como un grafo de frases permite cuantificar el índice de centralidad de todos los nodos del grafo, con lo que cuantifica la importancia de un nodo respecto a los otros nodos. Uno de los nodos de particular interés en los temas noticiosos son las frases que representa la figura del Presidente de la República.

El calculo de la centralidad para cada semana permite forma un serie de tiempo semanal de centralidades, de la cual se obtienen componentes de un modelo de Arima de serie de tiempo. Estas componentes sirven de variables regresoras que explican, la popularidad de los presidentes.

En [1] se muestra un trabajo en la misma linea de capturar temas importantes en las noticias y que reflejen la preocupación social sobre estos temas, pero la diferencia con este trabajo es que no se llega a relacionar estos temas importantes con la formación de la opinión pública.

Esta tesis se organiza de la siguiente manera: la sección 2 se introduce conceptos de la teoría de grafos y que permiten calcular el índice de centralidad de un grafo; en la sección 3 se concentra en el procesamiento

de texto, el cual permite crear los grafos semanales de temas importantes; en la sección 4 se introduce conceptos de series de tiempo los cuales permiten extraer componentes no observables de la centralidad de los presidentes; en la sección 5 se presentan los resultados obtenidos de la centralidad de los presidentes y adicionalmente se presentan modelos que explican la popularidad de los presidentes basada en la centralidad y finalmente en las secciones 6 y 7 se presentan las conclusiones y recomendaciones de este trabajo.

2. Grafos de palabras

2.1. Textos como grafos de palabras

El paradigma de modelar un texto como un grafo de palabras comenzó con los trabajos de Sowa en 1984 [4], en el cual se modelaba el conocimiento humano usando grafos bipartitos de conceptos-relaciones conceptuales, donde los nodos son conceptos y que se relacionan por medio de otros tipos de nodos llamados relaciones conceptuales. Estos grafos reciben el nombre de “Estructuras conceptuales” ó “Redes Semánticas”.

En este trabajo se usará la transformación de un texto en un grafo, mediante la asignación de palabras o frase del texto en conceptos. Estos harán las veces de los nodos de la red, la relación de los nodos, es la relación de concurrencia en un mismo contexto gramatical de estas frases o palabras correspondiente. El contexto gramatical va ser una oración, ya que son unidades de expresión básicas dentro del texto.

En el grafo de palabras, los nodos representan entidades estáticas, es decir los nodos representaran frases sustantivas, las cuales son una secuencia de palabras que comienza con un sustantivo, Ej: “presidente de la república ”, “fondo monetario internacional”

Ejemplo Un ejemplo de este tipo de representación, es el texto:

“Los niños juegan en el parque, y los niños comieron unos helados”

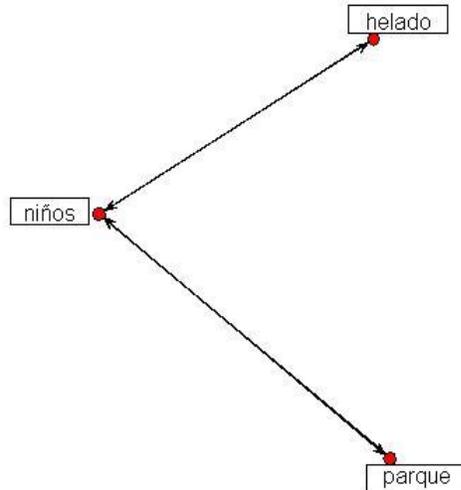
Se transforma en el grafo de la figura 1

Los nodos representan entidades que aparecen en el texto y los arcos representa relaciones entre los nodos, estas relaciones por lo general están definidas mediante verbos y preposiciones

En el análisis de texto bajo grafos, a diferencia de la lingüística computacional [10], no se trata que el grafo de palabras capture todo el sentido del texto, sino mas bien capturar la dependencia entre las entidades o conceptos a las cuales el texto referencia. Esto hace que el texto se lo pueda ver como un grafo es decir como una Red Social [12] de palabras donde los puntos de encuentro son los contextos gramaticales donde parecen estas palabras. Y estos contextos pueden ser: Frases, Oraciones.

El modelar al texto como una Red Social de palabras, permite aplicar técnicas de análisis de las Redes Sociales [12], y específicamente el índice de centralidad intermedia [13] (betweenness centrality) de los nodos de un grafo de palabras.

Figura 1: Ejemplo de grafo de palabras



2.2. Índices de Centralidad en el grafo

Sea $G=(V,E)$ un grafo, donde V es un conjunto de vértices y E es un conjunto de enlaces entre vértices. Se supondrá que $|V|=n$ y $|E|=m$

Para efecto de notación se representan los nodos como números enteros del 1 a n , donde n es la cantidad de nodos del grafo y los arcos serán los pares (i,j) donde i y j son números enteros que representan nodos y siempre y cuando exista un arco que los une de i a j

Camino sobre un grafo Un camino entre nodos u y v es una sucesión de enlaces, tal que siempre exista un nodo común entre dos enlaces consecutivos no terminales

$$Camino = \{e_i \in E / e_l = (i, k), e_{l+1} = (k, j)\}$$

Un grafo es conexo si para cada par de nodos u y v siempre existe un camino que une el par de nodos

Distancia geodésica Sea u y v dos nodos que pertenecen a un grafo conexo $G=(V,E)$. La distancia geodésica $d_G(u,v)$ es la distancia del camino mas corto que une los nodos u y v . Adicionalmente se tiene que σ_{uv} es la cantidad de caminos geodésicos desde el nodo u al nodo v . Y $\sigma_{uv}(s)$ es la cantidad de caminos geodésicos desde el nodo u al nodo v , pero que pasen por el nodo s

Tipos de índices de centralidad Sobre un grafo $G=(V,E)$ se puede calcular para cada nodo índices de centralidad que indique la importancia del nodo dentro de su contexto relacional (social).

El cálculo del índice de centralidad de cada nodo depende del criterio con que uno aplique el concepto de centralidad del grafo.

Existen tres criterios populares:

a) Centralidad basado en el grado del nodo

$$C_G(v) = \frac{\text{deg}(v)}{n-1} \quad (1)$$

Donde $\text{deg}(v)$: es la cantidad de vecinos que tiene el nodo v

n : es la cantidad de nodos del grafo

De acuerdo con esta definición, cuando más vecinos tienen un nodo, este nodo es más central es.

b) Centralidad basado en la accesibilidad del nodo desde resto del grafo

$$C_C(v) = \frac{1}{\sum_{t \in V \setminus v} d_G(v, t)} \quad (2)$$

Donde $d_G(v, t)$: es la distancia mínima entre el nodo v y el nodo t

La expresión $\sum_{t \in V \setminus v} d_G(v, t)$ es el total de la distancia entre los nodos del grafo hacia el nodo v , mientras mas pequeña sea esta distancia más fácilmente accesible será el nodo v para el resto de los nodos. Así que cuando C_c es más grande mayor será la accesibilidad de este nodo desde el resto de nodos del grafo

c) Centralidad basado en la importancia de intermediación o de conexión del nodo sobre el resto de nodos del grafo

$$C'_B(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (3)$$

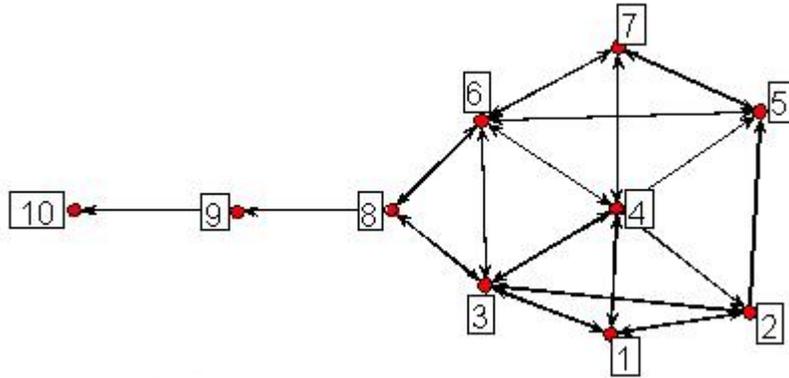
Es la suma sobre todas las combinaciones de caminos posibles entre cualquier par de nodos s y t , siempre y cuando pase por nodo v . La razón $\delta_{st}(v) = \frac{\sigma_{st}(v)}{\sigma_{st}}$ se conoce como dependencia del par (s, t) bajo la intermediación de v y mide el porcentaje de caminos más cortos que salen del nodo s y llegan hacia el nodo t y que pasan por el nodo v . Mientras más central es el nodo v , la expresión 3 será mayor, ya que el nodo v acapara más caminos geodésicos entre s y t . Esta suma tiene $(n-1)(n-2)$ sumandos, entonces se puede definir un índice de centralidad normalizado entre 0 y 1 como:

$$C_B(v) = \frac{\sum_{s \neq t \neq v \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}}{(n-1)(n-2)} \quad (4)$$

Un ejemplo que ilustra las diferencias entre estos tres índices de centralidad, y que se basa en grafo de la figura 2 :

Desde el punto de vista de la popularidad el nodo 4, es el que tiene más vecinos, pero desde el punto de la accesibilidad, los nodos 3 y 6 son más accesibles ya que se necesitan máximo 3 enlaces para conectar estos nodos desde el resto de nodo del grafo. El nodo 8 es el nodo mas influyente en el grafo ya que conecta conjuntos de nodos $\{9,10\}$ y $\{1,2,3,4,5,6,7\}$ que solo se conectan a través del nodo 8, es decir el nodo 8 hace

Figura 2: Ejemplo



Nodo	Centralidad Grado	Centralidad Cercanía	Centralidad Intermediación
1	0.334	0.500	0.000
2	0.444	0.529	0.023
3	0.556	0.642	0.232
4	0.667	0.600	0.102
5	0.334	0.500	0.023
6	0.556	0.642	0.232
7	0.334	0.500	0.000
8	0.334	0.600	0.389
9	0.222	0.428	0.222
10	0.111	0.310	0.000

Cuadro 1: Comparativo de centralidades

de “cuello de botella” en el grafo y su centralidad se basa en la intermediación.

Este trabajo usará como índice de centralidad el criterio de intermediación, bajo un grafo de frases sustantivas. Se va a usar este criterio debido al éxito en el uso de este índice como forma de extraer palabras relevantes de texto no estructurados [3], en esta referencia también se demuestra la superioridad del índice de centralidad sobre criterio de frecuencia de palabras o la frecuencia inversa de palabras [14], en la tarea de extraer palabras relevantes (Keywords) de un conjunto de textos.

De ahora en adelante nos referiremos como índice de centralidad al índice de centralidad por intermediación de un grafo

2.3. Algoritmo para el cálculo de la centralidad de intermediación

Usando el algoritmo de Dijkstra [20] se puede calcular las cantidades σ_{uv} y $\sigma_{uv}(s)$ de 4 pero tiene un tiempo de ejecución proporcional a $o(n^3)$, esto hace que el cálculo del índice de centralidad usando este algoritmo sea muy costosos para grafos de centenas de nodos. Afortunadamente se ha desarrollado un algoritmo [7] que calcula el índice de centralidad en un tiempo $o(nm)$ donde n y m son la cantidad de nodos y enlaces en el grafo.

El fundamento de este algoritmo son dos identidades:

1. La descomposición de la cantidad σ_{sv} en términos de la cantidad de caminos mínimos entre s y los nodos precedentes de v que pertenecen al camino mínimo

$$\sigma_{sv} = \sum_{u \in P_s(v)} \sigma_{su} \quad (5)$$

Donde $P_s(v)$ es el conjunto de los nodos predecesores de v , y que pertenecen al camino mas corto entre s y v

$$P_s(v) = \{u \in V / (u, v) \in E, d_G(s, v) = d_G(s, u) + 1\} \quad (6)$$

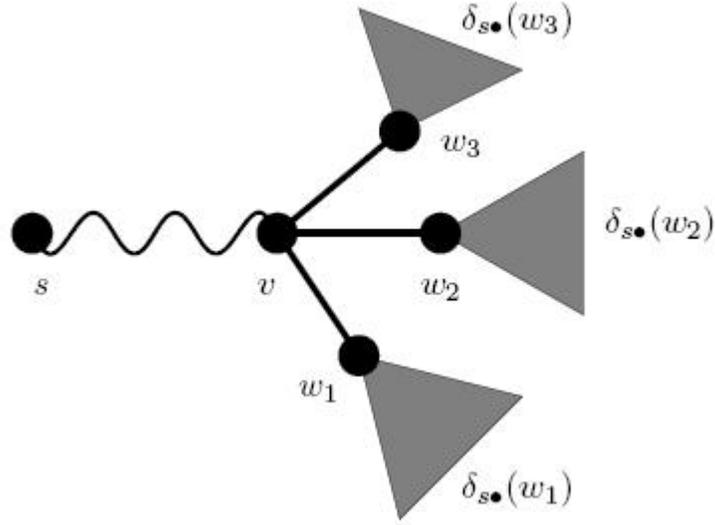
2. La descomposición del índice de centralidad en pares de dependencia acumulado $\delta_{s*}(v)$, que es definida como

$$\delta_{s*}(v) = \sum_{t \in V} \delta_{st}(v) \quad (7)$$

$$C_B(v) = \frac{\sum_{s, t \in V} \delta_{s*}(v)}{(n-2)(n-1)} \quad (8)$$

Y el par dependencia acumulado tiene una descomposición recursiva usando el conjunto de los nodos sucesores del nodo v , es decir aquellos nodos w tal que v sea un nodo precedente en la ruta mas corta de s a w , tal como se ve en la figura 3.

Figura 3: Descomposición de la dependencia acumulada



$$\delta_{s*}(v) = \sum_{w:v \in P_s(w)} \frac{\sigma_{sv}}{\sigma_{sw}} (1 + \delta_{s*}(w)) \quad (9)$$

La idea de este algoritmo es recorrer el grafo usando el procedimiento de búsqueda en amplitud en un grafo (breadth first search), en la cual se van encontrando para cada nodo v los nodos predecesores que están en la ruta mas corta desde el nodo s , es decir el conjunto $P_s(v)$

Con los nodos precedentes $P_s(v)$ y las ecuaciones 5,7,8 y 9 se calcula el índice de centralidad 4

El pseudo código de este algoritmo es el siguiente:

<p>Entrada: Un arreglo aNodo[] de listas de nodos, que representa a el grafo usando listas de adyacencias. La distancia entre nodo del grafo es 1</p> <p>Salida: Un arreglo CB[] de valores de centralidad para cada nodo</p> <p>Variables Intermedias:</p> <p>Cola: Cola con las funciones usuales de Enqueue(), Dequeue(), Empty(). Esta cola servirá para el recorrido del grafo en amplitud</p> <p>Pila: Pila con las funciones usuales de Push(), Pop(). En la pila se almacena de forma ordenada los nodos que tienen una distancia no decreciente con respecto al nodo s de cada iteración. Esto sirve para realizar el calculo recursivo 9</p> <p>P[]: Arreglo de lista de nodos. En este arreglo se almacenaran los nodos precedentes que están en la ruta mas corta $P_s(v)$</p> <p>Sigma, Delta : Arreglo para almacenar σ_{st} y $\delta_{st}(v)$</p> <p>D[]: Arreglo que almacena la distancia de un nodo al nodo s</p>
--

```

CB[v] ← 0 , para todo v en V
For s en V do
  Pila ← Vacio
  P[v] ← Vacio ;
  D[v] ← -1 ;
  Sigma[v] ← 0 para todo v en V
  Sigma[s] ← 1 ;
  D[s] ← 0
  Cola ← Vacio
  Cola.Enqueue(s)
  While !Cola.Empty()
    v=Cola.Dequeue()
    Pila.Push(v)
    For w es vecino de v do
      // Si w Es encontrada La primera vez
      If D[w] <0 then
        Cola.Enqueue(w)
        D[w] ← D[v] + 1
      End
      // Si v esta en la ruta mas corta entre s y w
      If D[w] = D[v] + 1
        Sigma[w] = Sigma[w] + Sigma[v]
        P[w].Append( v )
      End
    End
  End
End
Delta[v] ← 0 , para todo v en V
// Pila retorna vértices con distancias al nodo s de mayor a menor
While !Pila.Empty()
  W ← Pila.Pop()
  For v en P[w] do
    Delta[v] = Delta[v] + Sigma[v]*(1+Delta[w])/Sigma[w]
    If w s then CB[w] ← CB[w] + Delta[w] End
  End
End
End

```

Co-ocurrencia	Palabra1	Palabra2	Palabra3	Frecuencia
C1	Presidente	Lucio	Gutiérrez	15
C2	Coronel	Lucio	Gutiérrez	5
C3	Lucio	Gutiérrez		21
C4	Presidente	Lucio		15

Cuadro 2: Frecuencias de sub-secuencias

3. Procesamiento textual

3.1. Determinación de las frases significativas en las noticias

Co-ocurrencias Son secuencias de palabras que juntas son estadísticamente significativa es decir estas secuencias de palabras siempre tiende a aparecer juntas antes que separadas. Existen muchos criterios [5] para determinar si una secuencia de palabras es una co-ocurrencia, pero el criterio para determinar si un conjunto de palabras adyacentes forman una co-ocurrencia será el criterio de la información mutua, debido a su sencillez de cálculo y a que según en [5] no existe un criterio claramente dominante entre los criterios que determinan co-ocurrencia:

Información Mutua Sea $w = w_1w_2w_3 \dots w_n$ una secuencia de n palabras adyacentes w_i , la información mutua $I(w)$ se define como

$$I(w) = \log \left(\frac{p(w_1, w_2, \dots, w_n)}{p(w_1)p(w_2) \dots p(w_n)} \right) \quad (10)$$

Donde $p(w)$ es la probabilidad empírica de aparición de la secuencia y $f(w)$ es la frecuencia de aparición

$$p(w) = \frac{f(w)}{N} \quad (11)$$

Según [5], una co-ocurrencia será significativa cuando $I(w) > 3$.

El criterio anterior adolece de una falla: Si la co-ocurrencia w aparece 1 vez en el texto, casi automáticamente la co-ocurrencia se acepta como significativa ya que el termino $p(w_1)p(w_2) \dots p(w_n)$ es extremadamente pequeño y la información mutua por lo general es mayor a 3, lo cual hace que se tenga muchas co-ocurrencias accidentales, no significativas. Así que se considerará como co-ocurrencia aquellas secuencias de palabras que se repitan más de 2 veces, es decir la frecuencia $f(\cdot)$ sea mayor a 2: $f(w) > 2$, que es un criterio adicional que se propone en [5].

Ajustes de la frecuencia de las co-ocurrencia Al aplicar automáticamente el criterio de $f(w) > 2$ para seleccionar co-ocurrencias significativas, resulta que aparecen co-ocurrencias que son sub-secuencias de palabras contenidas en otras co-ocurrencias de mayor longitud por ejemplo:

Co-ocurrencia	Palabra1	Palabra2	Palabra3	Frecuencia
C1	Presidente	Lucio	Gutiérrez	15
C2	Coronel	Lucio	Gutiérrez	5
C3	Lucio	Gutiérrez		1
C4	Presidente	Lucio		0

Cuadro 3: Frecuencias ajustadas de sub-secuencias

En este caso la co-ocurrencia C3 pertenece a C1 y C2, así que la frecuencia de C1 y C2 se debe a las frecuencias de las co-ocurrencias C3, ya que la frase “Lucio Gutiérrez” esta contenida en las frases “Presidente Lucio Gutiérrez” y “Coronel Lucio Gutiérrez”, así que el conteo de C3 incluye cuando C3 aparezca en C1 y en C2. Entonces se tiene que hacer un ajuste de la frecuencia de las co-ocurrencias disminuyendo la frecuencia de las co-ocurrencias que están ya contenidas en otras co-ocurrencias de mayor longitud, en una cantidad igual a la frecuencia de la co-ocurrencia de la mayor longitud. En el caso anterior la tabla de frecuencia de las co-ocurrencia se ajusta como:

El ajuste de la frecuencia para la co-ocurrencia w es definido como

$$f'(w) = f(w) - \sum_{len(x)=len(w)+1; w \sqsubset x} f(x) \quad (12)$$

Donde $f'(w)$ es la frecuencia ajustada de la co-ocurrencia w , x es una co-ocurrencia de longitud superior en 1 a la longitud de la co-ocurrencia w y que contiene a la co-ocurrencia w .

3.1.1. Criterios de importancia de las co-ocurrencias

No todas las co-ocurrencias que son significativas en las noticias de una semana, son importantes, es decir existe co-ocurrencias que siempre aparecen en las noticias Ej: “El Universo el diario de mayor venta nacional”, pero que no son importantes ya que siempre aparece todos los días y por lo tanto no aportan ninguna información. El criterio para eliminar estas co-ocurrencias es dividir la probabilidad empírica de aparición de la co-ocurrencia en la semana contra la probabilidad empírica de aparición de la co-ocurrencia en todas las semanas. Entonces el índice de importancia de la co-ocurrencia w es:

$$Im(w) = \frac{\frac{f_s(w)}{N_s}}{\frac{f_T(w)}{N_T}} \quad (13)$$

Y seleccionará aquellas co-ocurrencias, como frases importantes, cuya importancia sea mayor a 5.

Resumiendo las tres condiciones que selecciona a una co-ocurrencia como nodo del grafo de palabras semanales:

Palabra	Palabra Estandarizada (Stem)
Chico	Chic
Chica	Chic
Chicos	Chic
Chicas	Chic
Mecanica	Mecan

Cuadro 4: Ejemplo de steeming

▪

$$I(w) > 3 \tag{14}$$

▪

$$f(w) - \sum_{len(x)=len(w)+1; w \sqsubset x} f(x) > 2 \tag{15}$$

▪

$$Im(w) > 5 \tag{16}$$

Se denominaran frases o colocaciones aquellas co-ocurrencias que satisfacen los tres criterios anteriores.

3.1.2. Palabras comunes (Stopwords)

Previo a la extracción de colocaciones de un texto, se debe extraer las palabras que no aportan información al análisis de texto. Estas palabras son: las preposiciones, artículos y pronombres. Bajo la tradición inglesa del procesamiento de textos, se conocen como stopwords, que un equivalente en español sería “palabras comunes”. Estas palabras al ser muy comunes no aportan mayor información al contenido relacional de las palabras. Estas palabras son útiles cuando se pretende realizar un análisis sintáctico profundo del texto [10].

La librería `tm` del lenguaje R [18] tiene un conjunto de palabras comunes por defecto para el español. Se usará este conjunto de palabras como stopwords, estas palabras serán eliminadas de las noticias de economía y política.

3.1.3. Stemming

El stemming [11] es una técnica de preprocesamiento de los textos, la cual consiste en estandarizar las palabras quitándole accidentes gramaticales como el género (masculino ó femenino), el número (plural ó singular) y las inflexiones verbales de la palabra (conjugaciones), y unificando todas las palabras diferentes a una palabra estándar (stem) que corresponde a un mismo significado.

El propósito del stemming en el preprocesamiento es estandarizar las palabras a una palabra representante (stem) con la esperanza que esta palabra represente un concepto.

Existen algunos procedimientos de stemming para el español [11, 16, 17]. Pero en este trabajo se usará el stemming de Porter para la lengua española [17] y que esta implementada en la librería Snowball de R.

3.1.4. Eliminación de palabras iniciales no sustantivas en la colocaciones

Una vez determinada las colocaciones, de acuerdo a los criterios 14 - 16 se procede a realizar una poda de la primera palabra de estas colocaciones, siempre y cuando no sea un sustantivo. Esto se hace para obtener colocaciones que tenga la forma: <Sustantivo>+ <Otras palabras>. Este tipo de colocación forma una frase sustantiva. Las frases sustantivas, determinan entidades a los cuales los textos referencias y este tipo de frases es el que se pretende representar en los nodos del grafo de palabras.

La frase sustantiva del presidente de la República Dentro del grafo de palabras, existe una frase sustantiva de particular importancia: la cual representa la figura del presidente de la República. Suele presentarse al Presidente de la República de múltiples maneras, por ejemplo para el presidente Gustavo Noboa, el diario El Universo se refería a él como: “presidente Noboa”, “presidente Gustavo Noboa” ó “mandatario Gustavo Noboa”.

Previo a crear el grafo de palabras para las noticias de la semana se unifican las colocaciones que correspondan al presidente de la República en una sola frase. La lógica que permite esta unificación es: Se clasifica a una co-ocurrencia como la frase “Presidente” si en la co-ocurrencia, aparece el nombre y el apellido del presidente de la República en funciones o si en las co-ocurrencias aparece la palabra “presidente” o “mandatario”, seguido por el apellido del presidente actual.

La frase del presidente de la República aparece todas las semanas en las noticias y llega a ser una frase no importante ya su índice de importancia $Im(w)$ no llega a ser mayor a 5, por tal motivo se obliga a que esta frase sea una frase importante y que aparezca en el grafo de la palabras de cada semana. Esto hace necesariamente que la frase del presidente pueda tener una alta centralidad siempre y cuando la frase presidente este relacionada textualmente con otras frases importantes de la semana y que estas frases importantes no estén previamente relacionadas.

Cuando sucede esto ?

Cuando el presidente esta en el centro de los temas importantes de la semana, ya sea que fue mencionado como un agente pasivo que recibe la acción de la noticia o como agente activo cuando es protagonista de la noticia

3.2. Elaboración del grafo de palabras

En el grafo de palabras se seleccionan como nodos las frases más importantes de la semana, siguiendo los criterios 14,15 y 16 más la frase que represente al presidente de la república .

Dos nodos del grafo de palabras están relacionados si sus frases sustantivas correspondientes están en una misma oración, la cual esta definida como:

Inicio Oración::= Inicio de texto ó Palabra con primera letra en mayúscula y que precede a un punto

Fin de Oración::=Punto que antecede a un fin de texto o que antecede una palabra en cuya primera letra esta en mayúscula que corresponde al inicio de otro párrafo.

Para propósito de capturar las co-ocurrencias dentro de las noticias, se divide las oraciones en segmentos, que están definidos como el conjunto de palabras dentro de una oración y que están delimitadas por los caracteres “,” , “;,” “:,” “)” y “(“ .

Dentro del grafo de palabras, debe estar identificada como un solo nodo las colocaciones que identifiquen al presidente de la República. Debido a la frecuencia con que aparece el presidente de la República en las noticias de manera mas o menos uniforme en la semana, así que el criterio 16 no se va satisfacer en la mayoría de las semanas, entonces se define el grafo de palabras como

$$G = (V, A)$$

$$V = \{Co - ocurrencias\ que\ satisfacen\ las\ condciones\ (14)\ (15)\ (16)\} \cup \{Frase\ que\ identifica\ al\ presidente\}$$

$$A = \{(i, j)/i, j \in V\ y\ están\ en\ la\ misma\ oración\}$$

4. Herramientas de Análisis de Datos

4.1. Modelos ARIMA

Los modelos Arimas surgen de la metodología Box-Jenkinhs [28] de series de tiempo. Estos modelos en su forma mas general es:

Sea y_t una serie de tiempo que corresponde a un modelo o a un proceso ARIMA, entonces se cumple lo siguiente

$$\phi(L) \Phi(L^s) (1-L)^d (1-L^s)^D y_t = \theta(L) \Theta(L^s) \epsilon_t \quad (17)$$

Donde

- s : es la estacionalidad de la serie de tiempo

- L : es el operador de desplazamiento de la serie de tiempo
- $\phi ()$: es un polinomio de grado p , que corresponde al proceso AR(p) de la serie de tiempo
- $Phi ()$: es un polinomio de grado P , que corresponde al proceso AR(P) de la estacionalidad
- $(1 - L)^d$: es la operación diferencia de la serie de tiempo
- $(1 - L^s)^D$: es la operación diferencia estacional de la serie de tiempo
- y_t : serie de tiempo que corresponde a un proceso ARIMA
- $\theta ()$: es un polinomio de grado q , que corresponde al proceso MA(q) de la serie de tiempo
- $\Theta ()$: es un polinomio de gado Q , que corresponde al proceso MA(Q) de la estacionalidad
- ϵ_t : es un ruido blanco

La notacion $(p,d,q) \times ((P,D,Q) \times (s))$ es usada para representar el orden del modelo ARIMA estacional, también llamado SARIMA.

Los modelos ARIMA son específicos de la metodología Box-Jenkins, la cual tiene tres etapas:

1. Identificación del modelo
2. Ajuste del modelo
3. Validación del modelo

4.1.1. Identificación del modelo

En esta etapa se estima el orden del modelo ARIMA que mejor a va ajustar los datos de la serie. Para realizar esta estimación se utiliza las gráficas propiedades de autocorrelación y autocorrelación parcial de la series de tiempos. Y teniendo en cuenta estas gráficas determinar que orden (p,d,q) va a tener la serie de tiempo. Para que funcione esta determinación usando las gráficas de autocorrelación y autocorrelación parcial, se necesita que la serie de tiempo no tenga estacionalidad y que la serie de tiempo sea estacionaria débil, es decir que se cumpla dos condiciones:

- $E [y_t] = \mu$
- $Cov [y_t y_{t-r}] = \gamma_r$

Una alternativa para seleccionar el modelo ARIMA mas adecuado para una serie de tiempo, es utilizar el criterio AIC [30] que se detalla en la siguiente sección

4.1.2. Ajuste del modelo

El ajuste consiste en determinar los coeficientes de los polinomios de la ecuación 17, dado que en la etapa anterior se determino el orden del modelo $(p,d,q) \times (P,D,Q) \times (s)$.

Para determinar los coeficientes de estos polinomios se usa el criterio de máxima verosimilitud del modelo ARIMA [28].

La función `arima(stat)` del lenguaje R, tiene ya implementado los procedimientos necesarios para ajustar un modelo ARIMA, dado un determinado orden $= (p,d,q) \times (P,D,Q) \times (s)$.

4.1.3. Validación del modelo

Una vez ajustado el modelo, se procede a realizar las verificaciones necesarias, si el modelo ajustado cumple con supuestos necesarios que permitieron su adecuada estimación. Estos supuestos son :

1. Normalidad de los residuos del modelo
2. Independencia de los residuos
3. Homocedasticidad de los residuos
4. No existen observaciones aberrantes
5. El modelo no tiene sobreajuste

Para los puntos desde 1 al 3, la verificación consiste en correr prueba de hipótesis para la normalidad, independencia y homocedasticidad de los residuos

Existen muchas pruebas de hipótesis para la normalidad de los residuos, pero se usará la prueba de Shapiro-Wilk Wilk , que es la mas popular

Para verificar la independencia de los residuos se usará dos criterios: el gráfico de las autocorrelaciones y la prueba de hipotesis de Durbin Watson

Para verificar la homocedaticidad de los residuos, se usará la prueba de Bartlett

En la verificación de datos aberrantes, se usará el criterio $|e_t| \geq z_{0,005}$, es decir se seleccionará como datos aberrantes si caen en la region de rechazo que una confianza de 0,01.

LA verificación si el modelo esta sobreajustado, consiste en verificar si existe algún parámetro del modelo que esta demás. Esto se los hace usando el criterio AIC o usando prueba de hipótesis sobre la significancia de cada coeficiente asociado.

4.2. Criterio AIC

El criterio AIC (Akaike Information Criterion) es un indicador que califica el ajuste del modelo con los datos. Este indicador sirve para comparar diferentes modelos que han sido ajustado con el mismo conjunto

de datos, seleccionado el modelo que tenga el mejor ajuste este decir menor valor de AIC. El indicador tiene la siguiente forma

$$AIC = 2k - \ln(L)$$

Donde

- k: es la cantidad de parámetros que tiene el modelo, que en el caso del modelo ARIMA tenemos k es igual a $p+q+P+Q$
- L: Es la verosimilitud del modelo

El criterio AIC no solo incorpora la verosimilitud del modelo sino que también incorpora la cantidad de parámetros involucrados con signo opuesto. El numero de parámetros representa la complejidad del modelo ya que si solo se incorporara la verosimilitud se estaría escogiendo los modelos con mas cantidad de parámetros, ya que al aumentar la cantidad de parámetros de un modelo su ajuste mejora, pero la complejidad del modelo aumenta lo cual puede hacer caer el modelo en un sobreajuste. Un modelo ajusta es aquel modelo que tiene un excelente ajuste con los datos de prueba, pero tiene un pobre desempeño sobre los datos nuevos, en este caso se dice que el modelo memoriza los datos, pero no los generaliza. Así que el criterio AIC trata de evitar los modelos sobreajustados y los modelos con pobre ajuste.

4.3. Determinación del modelo ARIMA usando el criterio ACI

En este trabajo se usará el criterio AIC para identificar un conjunto de los 10 mejores modelos ARIMA que tengan el mejor ajuste equilibrado sobre los datos. De este conjunto de modelos se escoge, de acuerdo al orden de AIC, el modelo que satisfaga todos los requerimientos de validación de un modelo ARIMA, es decir el modelo finalmente elegido debe tener residuos normales, independientes y homocedástico y el modelo no debe ser sobreajustado, es decir todos sus parámetros deben ser significativos.

Para seleccionar 10 los mejores modelos ARIMA se escogen del siguiente conjunto de parámetros .

$$EspacioParametros = \{(p, d, q, P, D, Q, s) / p, q, P, Q \in \{0, 1, 2, 3\} \ d, D \in \{0, 1, 2\} \ s \in \{3, \dots, 17\}\} \quad (18)$$

Como el criterio AIC depende de la verosimilitud o del ajuste que tenga el modelo y este ajuste depende de valores aberrantes que los datos tenga, entonces al seleccionar el modelo con el mejor AIC se debe incorporar posibles valores aberrantes que el modelo va a tener. Para evitar una búsqueda combinatoria se usará la siguiente estrategia.

1. Se ajusta un modelo AR(1) sobre los datos
2. Se seleccionara como valores aberrantes aquellas observaciones cuyos residuos caigan en la región de rechazo con 0.02 de confianza

Se escoge una región de rechazo mayor que la establecida anteriormente para los valores aberrantes, debido a que se quiere escoger las posibles observaciones aberrantes y a estas observaciones aplicarles variables dummy de la centralidad de cada presidente.

Estos valores aberrantes se incorpora en el ajuste de los modelos de 18 y se escoge los 10 mejores modelos con el menor AIC. Se obtiene una tabla de los 10 mejores modelos con el menor AIC

A cada uno de los 10 mejores modelos obtenidos, se vuelve a determinar si existe algún valor aberrantes pero considerando una región de rechazo de 0,01. Si existirá algún valor aberrante en estos modelos se incorpora estos nuevos valores como variables dummy, obteniéndose 10 modelos ARIMAS sin valores aberrantes.

Del conjunto de los 10 modelos se escoge el modelo con el menor AIC, pero al cual se pueda eliminar parámetros no significativos sin eliminar las buenas características que deben tener los residuos como normalidad, no aberrante, independiente y homocedástico.

4.4. Dinámica comunicacional y su perturbaciones

El modelo ARIMA finalmente ajustado, sin considerar los valores aberrantes, representa una dinámica comunicacional que tiene cada mandatario, es decir la dinámica de cuando y como aparecer en los medios de prensa escrita.

Los valores aberrantes mas los residuos del modelo se consideran perturbaciones exógenas a la estrategia comunicacional de cada presidente. La volatilidad de cada observación es la varianza estimada de cada perturbación asociada, es decir el cuadrado de la perturbación asociada a cada semana. Se considera que estas perturbaciones afectan la imagen del presidente frente a la opinión pública.

Así que el modelo ARIMA provee dos variables para medir la gestión comunicacional de cada presidente, semana a semana: Dinámica comunicacional y Volatilidad.

4.5. Regresión lineal, bajo heterocedasticidad y autocorrelación

Para encontrar si existe relación entre centralidad y la popularidad, se planteará un modelo lineal. Estimar modelos lineales están sujetos a problemas como heterocedasticidad y autocorrelación en los residuos. Existen algunas técnicas para lidiar con estos problemas [32], que ayudan pero aseguran la eliminación de estos problemas. Afortunadamente existe un método que asegura corregir este problema, siempre y cuando se tenga una muestra grande.

4.5.1. Regresión lineal

La regresión lineal consiste en estimar el parámetro β , del modelo lineal

$$y_i = x_i^T \beta + u_i \quad (i = 1, 2, \dots, n) \quad (19)$$

Donde y_i es una variable que depende del vector columna x_i (px1), β es un vector de coeficientes (px1) y un termino de error. En términos matriciales el modelo de la ecuación 19 es $y = X\beta + \epsilon$ donde X es una matriz (npx).

Bajo los supuestos de los residuos normalidad, independencia y homocedasticidad, los coeficientes β del modelo de la ecuación 19 pueden ser estimados usando el procedimiento de mínimos cuadrados

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (20)$$

$$\hat{u} = (I_n - H) u = (I - X (X^T X)^{-1} X^T) u \quad (21)$$

Donde I_n es una matriz identidad de tamaño n y H es la matriz sombrero. El estimador $\hat{\beta}$ es insesgado y asintoticamente normal [33].

La matriz de covarianza del estimador $\hat{\beta}$ es

$$\Psi = VAR[\hat{\beta}] = (X^T X)^{-1} X^T \Omega X (X^T X)^{-1} \quad (22)$$

Donde Ω es matriz de covarianza de los errores residuales u_i

La ecuación 22 también se puede expresarse como

$$\Psi = VAR[\hat{\beta}] = (X^T X)^{-1} \frac{1}{n} \Phi (X^T X)^{-1} \quad (23)$$

Donde $\Phi = nX^T \Omega X$ es la matriz de covarianza de los score definidos como: $V_i(\beta) = x_i (y_i - x_i^T \beta)$

Si se cumple las condiciones de normalidad, independencia y homocedasticidad de los residuos, $\Omega = \sigma^2 I_n$, entonces matriz de covarianza $VAR[\hat{\beta}]$ se reduce a $\Psi = \sigma^2 (X^T X)^{-1}$. Y finalmente la matriz puede ser estimada usando el estimador de σ^2 $\hat{\Psi} = \hat{\sigma}^2 (X^T X)^{-1}$

Pero si no se cumple las condiciones de independencia y homocedasticidad, el estimador de $\hat{\beta}$ sigue siendo insesgado, no así el estimador de la matriz de covarianza de los coeficientes $VAR[\hat{\beta}]$. Lo cual hace que la prueba de hipótesis sobre los coeficientes comiencen a fallar.

Para estimar adecuadamente la matriz de covarianza, se tiene que usar la ecuación 22, pero el problema es como estimar la matriz Ω . Este un problema complicado ya esta matriz tiene $\frac{n(-1)}{2}$ parámetros, que necesitan ser estimados.

4.5.2. Regresión lineal, bajo heterocedasticidad

Si solo existiera heterocedasticidad, entonces la matriz Ω tiene una forma diagonal, n parámetros a estimar en este caso el problema de estimar esta matriz se simplifica. Existen algunos estimadores para esta matriz.

1. $w_i = \text{constante}$
2. $w_i = \hat{u}_i^2$
3. $w_i = \frac{n}{n-p} \hat{u}_i^2$
4. $w_i = \frac{\hat{u}_i^2}{1-h_i}$
5. $w_i = \frac{\hat{u}_i^2}{(1-h_i)^2}$
6. $w_i = \frac{\hat{u}_i^2}{(1-h_i)^{\delta_i}}$

Donde h_i es la diagonal de la matriz H , \bar{h} es su media, $\delta_i = \min \{4, h_i/\bar{h}\}$.

El primer estimador produce la matriz de covarianza $VAR[\hat{\beta}]$ estándar, los otros estimadores producen mejores estimaciones de la matriz Ω . Para calcular la matriz de covarianza $VAR[\hat{\beta}]$, se tiene que reemplazar estos estimadores en la matriz Ω de la ecuación 22

La estimación de la matriz 22, sujeto a estas mejoras están implementadas en función `vcovHC()` en la librería, `sandwich` del lenguaje R.

4.5.3. Cuando existe heterocedasticidad y autocorrelación

En el caso general cuando exista heterocedasticidad y autocorrelación, se usa el ajuste de latriz de covarianza $VAR[\hat{\beta}]$ de acuerdo al trabajo de Newey y West [33]. En este trabajo se propone un estimador para la matriz Ψ de la ecuación 23.

$$\hat{\Phi} = \frac{1}{n} \sum_{i,j=1}^n w_{|i-j|} \hat{V}_i \hat{V}_j^T \quad (24)$$

Donde $w = (w_0, w_1, \dots, w_{n-1})$ es un vector de pesos y \hat{V}_i es el vector de los score del i -esimo coeficiente. Existen diferentes recomendaciones para estimar estos pesos, pero en este trabajo se usará la recomendación de Newey West

$$w_i = 1 - \frac{i}{L+1} \quad (25)$$

Donde L es máximo lag en que los pesos son positivo, pesos con lag superiores son ceros.

5. Resultados obtenidos

5.1. Descripción de los datos

Para este trabajo se creo un corpus (conjunto de textos) que consiste en todas las noticias de la secciones de política y economía del diario El Universo de Ecuador (www.eluniverso.com) desde el 23 de marzo 2002

hasta el 21 de Diciembre 2008.

Las noticias fueron extraídas de las páginas web del diario a través de un programa de licencia libre llamado wget [9] . Este programa baja automáticamente las páginas de sitios web.

Se considera como unidad de texto el conjunto consolidado de noticias de economía y política del diario “El Universo” de la ciudad de Guayaquil, ocurrido en una semana, de lunes a domingo. A su vez a cada noticia se descompone en oraciones. Se reconoce como oración aquellas secuencias de palabras que comience con un punto y una letra en mayúscula o que comience en el inicio de un párrafo y termine con un punto final, el fin de un párrafo o finalice con un punto seguido y el inicio de una letra mayúscula.

Finalmente a cada oración se la descompone en segmentos. Se reconocerá como segmentos de oración aquella secuencia de palabras que están dentro de una oración y están delimitados por los símbolos “,” , “;” , “:” y “.” o que están dentro paréntesis. Entonces el corpus tiene la siguiente descomposición :

Corpus →Semana (Unidad de texto) → Noticia → Oración → Segmento

Se reserva el termino “frase” a aquellas secuencias de palabras que tenga significado.

5.2. Procesamiento de las noticias

Se procesaron 76330 noticias de economía y política aparecidas en la pagina web www.eluniverso.com desde el 23 de Marzo 2002 hasta el 21 de Diciembre 2008, Este periodo de tiempo comprenden 352 semanas. Se procesaron las noticias de cada semana aplicando los siguientes pasos

1. Extracción de las noticias del sitio www.eluniverso.com
2. Extracción del texto noticioso de cada pagina Web
3. Steeming y eliminación de palabras comunes
4. Determinación de las co-ocurrencias significativos e importantes
5. Determinación del grafo semanal de temas importantes
6. Calculo de índice de centralidad

5.2.1. Extracción de las noticias del sitio www.eluniverso.com

Se extrajeron 76330 paginas Web, cada pagina correspondiente a una noticia, entre los 23 de Marzo 2002 hasta 21 de Diciembre 2008. periodo de tiempo que corresponde a 252 semanas. La extracción de estas paginas se lo hizo con la herramienta wget [9], la cual es una herramienta que permite bajar se forma automática paginas HTML de sitios Web de forma automática. En el sitio www.eluniverso.com las noticias están estructuradas en directorios que corresponde a la fecha donde salio la noticia y dentro de cada directorio

Columna	Descripción
Fecha	Fecha cuando apareció la noticia
Noticia	Identificador de la noticia en la tabla de noticias
Sección	Número secuencial que enumera la oración dentro de la noticia
Oración	Número secuencial de segmento de oración
Segmento	Número de secuencia de la palabra dentro de la oración
Orden	Número de secuencia de la palabra dentro de la oración
A4	La palabra transformada usando el procedimiento de steeming

Cuadro 5: Columnas de la tabla que almacena las noticias

de fechas aparecen los subdirectorios que corresponden las diferentes secciones. Esta estructuración permitió la configuración sencilla de la herramienta wget.

5.2.2. Extracción del texto noticioso de cada pagina Web

Se programó un procedimiento ah-doc en base a los tags HTML encontrados al interior de las paginas web. Este procedimiento consiste en reconocer los tags de inició del texto noticioso y comenzar a almacenar el texto encontrado, hasta llegar a los tags que determinan el fin del texto noticioso. El almacenamiento del texto es en base a los caracteres en minúscula del alfabeto latino es decir se eliminan acentos en los caracteres á,é,í,ó,ú,ñ,ü. El propósito de esta eliminación es para utilizar las función Steeming de la librería tm del lenguaje R

5.2.3. Steeming y eliminación de palabras comunes

Una vez extraído el texto noticioso se eliminan las palabras comunes y las palabras que quedan se las transforma usando el procedimiento de Steeming el cual permite tener palabras estandarizadas. Para ambos procedimiento se usa las funciones de la librería tm del lenguaje R.

5.2.4. Determinación de las co-ocurrencias significativos e importantes

Con el paso anterior se obtiene textos noticioso sin palabras comunes y expresado en palabras estandarizadas. Con el propósito de facilitar la determinación de co-ocurrencias de palabras, se transforma estos textos noticiosos en tablas relacionales. El cuadro 5 muestra las columnas de esta tabla.

Luego se programó un procedimiento en SQL el cual contaba rápidamente las frecuencias de ocurrencias de palabras o grupos de palabras que parecen dentro de un segmento de oración. Estas frecuencias de conteo son básicas para la determinación si una co-ocurrencia es significativa e importante dentro de la semana, de acuerdo con los criterios 14, 15 y 16

Presidente	Fecha Inicio	Fecha Fin	Palabras ocurrentes	Semana Inicio	Semana Fin
Gustavo Noboa	23/03/2002	15/01/2003	“gustavo noboa” ,“presidente noboa”	1	43
Lucio Gutiérrez	16/01/2003	20/04/2005	“lucio gutiérrez ”,“presidente gutiérrez ”	44	161
Alfredo Palacio	20/04/2005	15/01/2007	“alfredo palacio”, “presidente palacio”	162	251
Rafael Correa	16/01/2007	21/12/2008	“rafael correa”, “presidente correa”	252	352

Cuadro 6: Rangos de semanas y palabras ocurrentes para cada mandato

5.2.5. Determinación del grafo semanal de temas importantes

Una vez determinadas las co-ocurrencias significativas e importantes que van formar los nodos del grafo semanal de noticias, se agrega un nodo adicional que representa al presidente de la República para esa semana, por tal motivo se unifican las diferentes co-ocurrencias significativas que referencian al presidente y esta identificación es dependiente de la semana cuando aparezca esta co-ocurrencia, de acuerdo con la tabla 6.

Al agregar el nodo que representa al presidente se completa todos los nodos que va a estar en el grafo de palabras, luego se determinan los arcos de este grafo.

Usando una simple operación Join del lenguaje SQL sobre la tabla, se determina la ocurrencia simultanea entre cada par de frases dentro de alguna oración aparecida en alguna noticia de la semana. Esta ocurrencia simultanea determina el arco entre dos frases.

5.2.6. Calculo de índice de centralidad

Las frases mas significativas de la semana, mas la frase del presidente forman los nodos del grafo semanal de los temas noticiosos de la semana. A este grafo se aplica el algoritmo de la sección 2.3 y se determina la centralidad de intermediación del nodo, del presidente de la República para la semana correspondiente

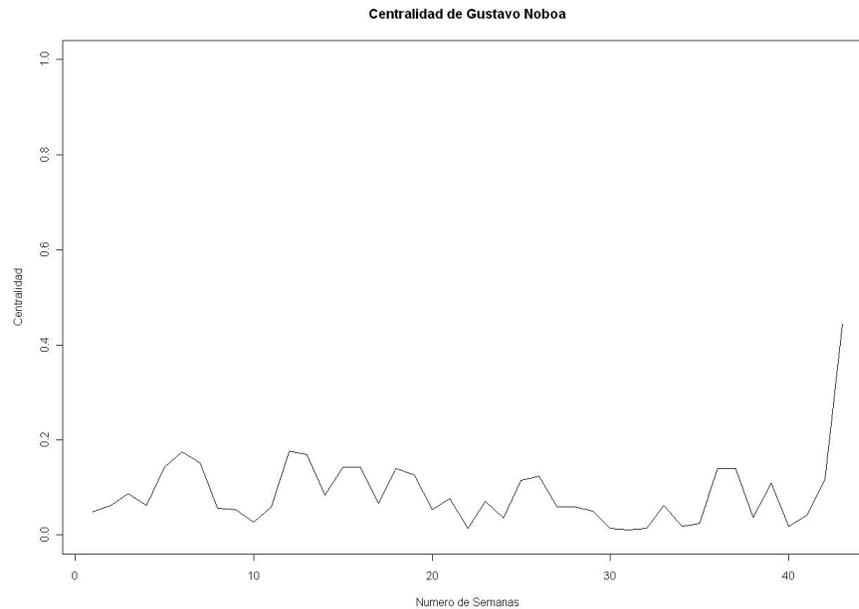
5.3. Centralidad y Popularidad del presidente

5.3.1. Centralidad

Para cada una de las 352 semanas se determina el grafo semanal de palabras y se calcula el índice de centralidad del nodo “Presidente de la República” Con esto se obtiene una serie de tiempo semanal de 352 valores. Esta serie de tiempo se ha dividido de acuerdo con los periodos de cada mandatario.

En los gráficos 4, 5,6 y 7 se muestran las centralidades de cada presidente. De la observación de estos gráficos, se nota que los picos mas altos de centralidad lo tiene Rafael Correa con 0.6 en la semana 26 de su mandato. Pero Lucio Gutierrez y Alfredo Palacios tienen un patrón mas irregular en la centralidad que

Figura 4: Centralidad de Gustavo Noboa



los otros presidentes. La diferencia entre Lucio Gutiérrez y Alfredo Palacio, es que la centralidad de Alfredo Palacio tendió a reducirse, desde la mitad de su mandato hasta el final, debido al proceso electoral que paso el país en esa época.

Gustavo Noboa tiene la menor centralidad que todos presidentes, este es debido a la misma razón del proceso electoral, ya que la serie de la centralidad de Gustavo Noboa apenas tiene 43 observaciones, y en este periodo se vivio el proceso electoral donde fue elegido Lucio Gutiérrez.

Ni Lucio Gutiérrez ni Rafael Correa vivieron procesos electorales, ya que el primero fue derrocado por el congreso y el segundo la serie de tiempo de Rafael Correa llega hasta diciembre del 2008.

5.4. Determinación de los valores aberrantes y la dinámica comunicacional

Con el propósito de elegir el modelo ARIMA mas adecuado para la centralidad de cada presidente se determina los valores aberrantes iniciales, los cuales son determinado por un modelo AR(1) que se ajusta para periodo presidencial. Estos valores aberrantes permitirán generar variables dummy para los modelos ARIMA de la ecuación 18, de este conjunto de modelos ARIMA se escoge 10 modelos con el mejor AIC. Estos modelos escogidos ya tendrá incorporado las mayoría de valores aberrantes para la centralidad. De los 10 modelos ARIMA seleccionado se escoge un modelo que tenga buenos residuos.

Al ajustar un modelo AR(1) para la centralidad de cada presidente y seleccionar valores que caigan en la zona de rechazo de 0.01, es decir valores que sean 2.57 veces mas que la desviación estándar de los residuos.

Figura 5: Centralidad de Lucio Gutiérrez

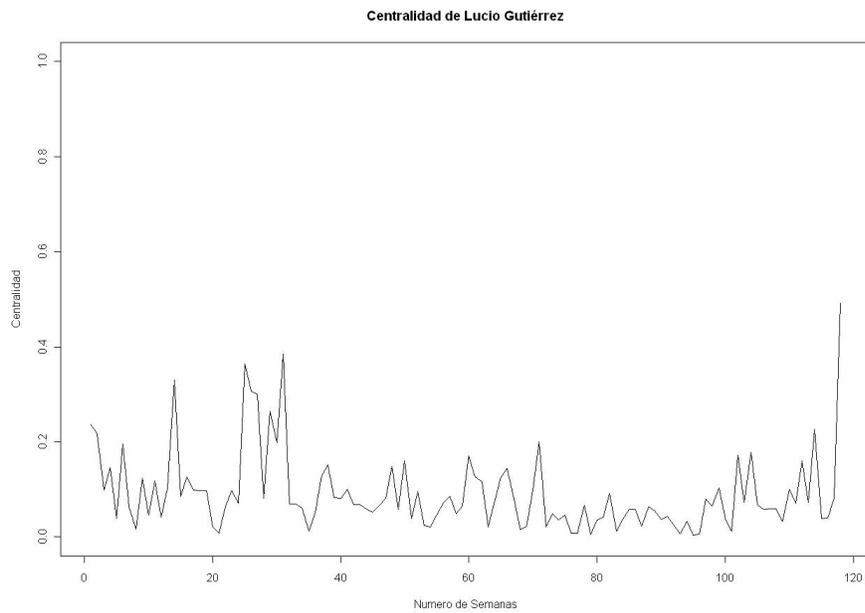
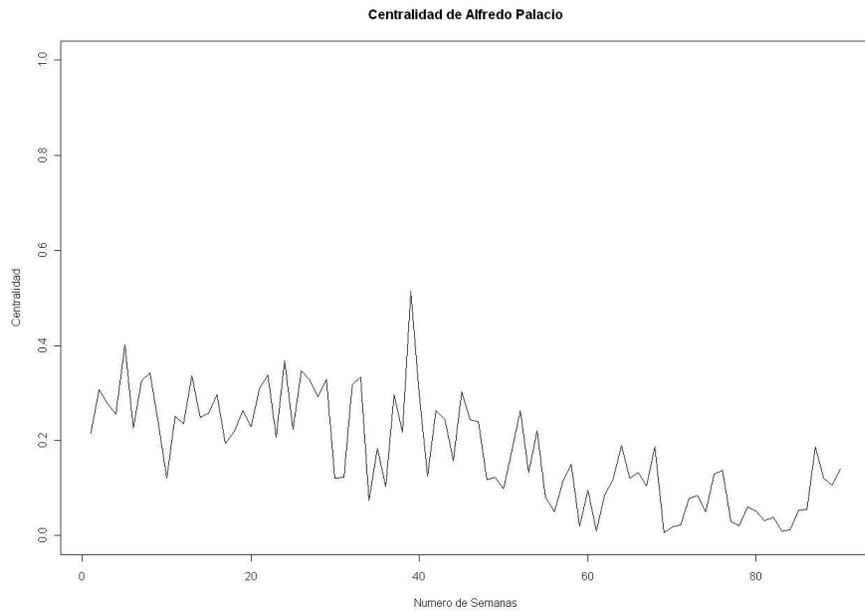


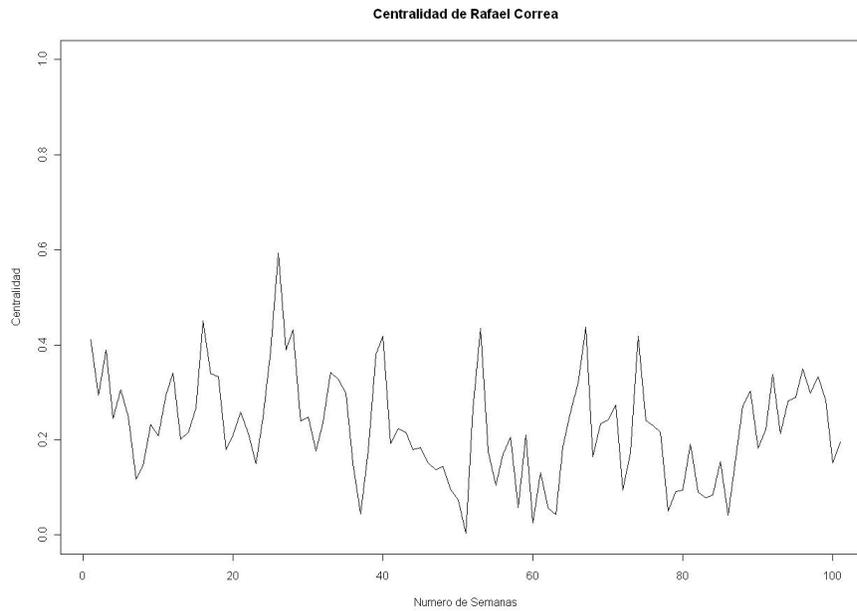
Figura 6: Centralidad de Alfredo Palacio



Presidente	Semanas
Gustavo Noboa	43
Lucio Gutiérrez	44, 49, 57, 68, 69, 70, 72, 74, 75, 115, 157, 161
Alfredo Palacio	195, 200
Rafael Correa	277, 325

Cuadro 7: Observaciones aberrantes para la centralidad según el modelo AR(1)

Figura 7: Centralidad de Rafael Correa



Presidente	Parametros						
	p	d	q	P	D	Q	s
Gustavo Noboa	0	0	1	0	0	0	0
Lucio Gutiérrez	1	0	1	0	0	1	3
Alfredo Palacio	0	1	1	0	0	0	0
Rafael Correa	1	0	0	2	0	2	15

Cuadro 8: Parámetros de los modelos ARIMA

Los siguientes modelos ARIMA son los mejores modelos que se encontraron para cada presidente considerando AIC, calidad de los residuos y eliminación del sobreajuste.

De estos modelos se obtuvieron nuevos valores aberrantes, con una zona de rechazo de 0,01 que junto con los valores aberrantes iniciales forman un conjunto total de valores aberrantes.

Estos valores aberrantes, están resumidos en el cuadro 9

Al incorporar estos valores aberrantes como variables dummy en los modelos ARIMA, se obtiene modelos que no poseen mas valores aberrantes. Y no se obtuvieron modelos sobreajustados.

En la tabla 10 se muestran los parametros de los modelos ARIMA de cade presidente.

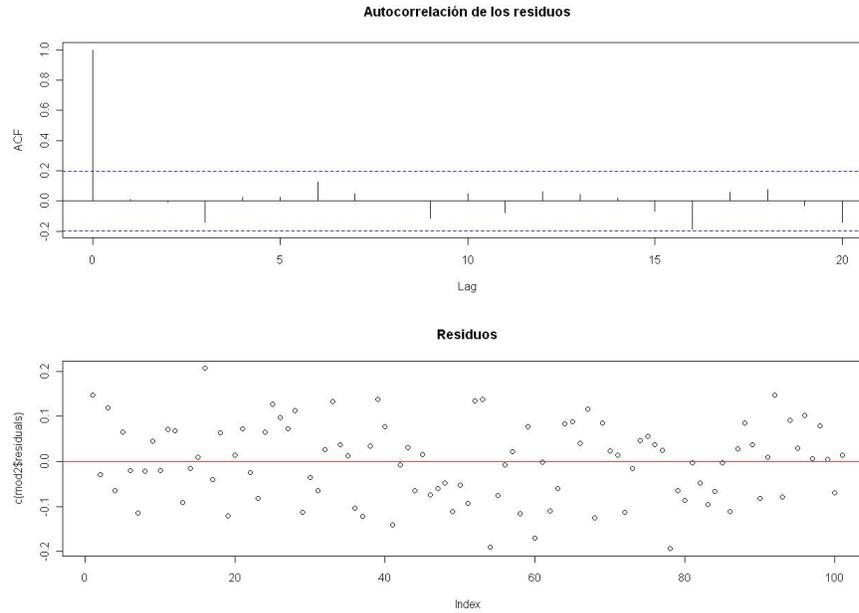
Presidentee	Semana	Fecha Inicio	Fecha Fin	Eventos
Gustavo Noboa	43	13/01/2003	19/01/2003	Transición al nuevo gobierno
Lucio Gutiérrez	44	20/01/2003	26/01/2003	Inicio del mandato
Lucio Gutiérrez	45	27/01/2003	02/02/2003	Criticas por reparto de cargos
Lucio Gutiérrez	49	24/02/2003	02/03/2003	Postergación de la consulta popular
Lucio Gutiérrez	57	21/04/2003	27/04/2003	100 primeros días de gobierno
Lucio Gutiérrez	68	07/07/2003	13/07/2003	Rompimiento con el MPD
Lucio Gutiérrez	69	14/07/2003	20/07/2003	Criticas al gobierno por no cumplimiento de promesas de campaña
Lucio Gutiérrez	70	21/07/2003	27/07/2003	Fricciones con Pachakutik
Lucio Gutiérrez	72	04/08/2003	10/08/2003	Rompimiento con Pachakutik
Lucio Gutiérrez	73	11/08/2003	17/08/2003	Viaje a Paraguay
Lucio Gutiérrez	74	18/08/2003	24/08/2003	Marchas contra el régimen
Lucio Gutiérrez	75	25/08/2003	31/08/2003	Visita del presidente a China
Lucio Gutiérrez	93	29/12/2003	04/01/2004	Ultima semana del año
Lucio Gutiérrez	103	08/03/2004	14/03/2004	Proyecto del voto militar
Lucio Gutiérrez	114	24/05/2004	30/05/2004	Pugna con LFC
Lucio Gutiérrez	115	31/05/2004	06/06/2004	Denuncias de conspiración contra el gobierno
Lucio Gutiérrez	145	27/12/2004	02/01/2005	Ultima semana del año
Lucio Gutiérrez	147	10/01/2005	16/01/2005	Presentacion informe a la nación y pugnias
Lucio Gutiérrez	157	21/03/2005	27/03/2005	Maniobras del gobierno para calmar crisis politica
Lucio Gutiérrez	161	18/04/2005	24/04/2005	Caída del gobierno
Alfredo Palacio	171	27/06/2005	03/07/2005	Bajo nivel mediático
Alfredo Palacio	191	14/11/2005	20/11/2005	Bajo nivel mediático
Alfredo Palacio	192	21/11/2005	27/11/2005	Bajo nivel mediático
Alfredo Palacio	195	12/12/2005	18/12/2005	Bajo nivel mediático
Alfredo Palacio	197	26/12/2005	01/01/2006	Ultima semana del año
Alfredo Palacio	200	16/01/2006	22/01/2006	Informe polémico a la nación
Alfredo Palacio	202	30/01/2006	05/02/2006	Bajo nivel mediático
Rafael Correa	277	09/07/2007	15/07/2007	Separación de 18 diputados del bloque de la dignidad
Rafael Correa	325	09/06/2008	15/06/2008	Criticas a Alberto Acosta sobre el manejo de la asamblea

Cuadro 9: Observaciones aberrantes para la centralidad de cada presidente

Presidente	Parámetros	Semana	Coficiente	Desviación Estandar	Estadistico t
Gustavo Noboa	ma1		0.45557079	0.15406721	2.956961
Gustavo Noboa	intercepto		0.08048046	0.01008828	7.977619
Gustavo Noboa	outliers	43	0.34621015	0.04580129	7.558960
Lucio Gutiérrez	ar1		0.91174133	0.063031254	14.464908
Lucio Gutiérrez	ma1		-0.65281596	0.114581000	-5.697419
Lucio Gutiérrez	sma1		-0.40221459	0.112226970	-3.583939
Lucio Gutiérrez	intercepto		0.06574627	0.006895801	9.534247
Lucio Gutiérrez	outliers1	44	0.18409401	0.034441849	5.345068
Lucio Gutiérrez	outliers2	45	0.11769949	0.033782047	3.484084
Lucio Gutiérrez	outliers3	49	0.14601368	0.030036858	4.861150
Lucio Gutiérrez	outliers4	57	0.25236205	0.029384409	8.588298
Lucio Gutiérrez	outliers5	68	0.30640300	0.032056936	9.558088
Lucio Gutiérrez	outliers6	69	0.24410353	0.032673222	7.471058
Lucio Gutiérrez	outliers7	70	0.22231527	0.033167610	6.702782
Lucio Gutiérrez	outliers8	72	0.18943510	0.032915644	5.755169
Lucio Gutiérrez	outliers9	73	0.12515915	0.032824198	3.813015
Lucio Gutiérrez	outliers10	74	0.32434333	0.031772274	10.208376
Lucio Gutiérrez	outliers11	93	0.07975282	0.028860457	2.763394
Lucio Gutiérrez	outliers12	103	0.07169866	0.030152384	2.377877
Lucio Gutiérrez	outliers13	114	0.12571493	0.029238697	4.299608
Lucio Gutiérrez	outliers14	145	0.13176547	0.029413358	4.479783
Lucio Gutiérrez	outliers15	147	0.10290892	0.029842827	3.448364
Lucio Gutiérrez	outliers16	157	0.15435733	0.029693353	5.198380
Lucio Gutiérrez	outliers17	161	0.42294255	0.032737502	12.919206
Alfredo Palacio	ma1		-0.7074289	0.07729583	-9.152225
Alfredo Palacio	outliers1	171	-0.1536261	0.05519963	-2.783101
Alfredo Palacio	outliers2	191	-0.1783634	0.05636725	-3.164309
Alfredo Palacio	outliers3	192	-0.1738187	0.05645893	-3.078675
Alfredo Palacio	outliers4	195	-0.2028707	0.05586894	-3.631189
Alfredo Palacio	outliers5	197	-0.1608436	0.05581712	-2.881618
Alfredo Palacio	outliers6	200	0.2572041	0.05578997	4.610220
Alfredo Palacio	outliers7	202	-0.1240638	0.05576740	-2.224666
Rafael Correa	ar1		0.4853925	0.094408279	5.141419
Rafael Correa	sar1		1.2451552	0.172751032	7.207802
Rafael Correa	sar2		-0.6749624	0.139647550	-4.833328
Rafael Correa	sma1		-1.9136446	0.498125743	-3.841690
Rafael Correa	sma2		0.9992718	0.494912973	2.019086
Rafael Correa	intercepto		0.2215544	0.007597501	29.161476
Rafael Correa	outliers1	277	0.1991933	0.068198879	2.920772
Rafael Correa	outliers2	325	0.2256338	0.071903030	3.138030

Cuadro 10: Parámetros del modelo de centralidad para los presidentes

Figura 8: Residuos y su correlograma, centralidad de Gustavo Noboa



5.5. Análisis de los residuos de los modelos de centralidad obtenidos

De los gráficos 8, 9,10,11 y las pruebas estadísticas sobre los residuos las cuales están resumidas en el cuadro 11 se puede decir que los modelos de centralidad encontrados para cada presidente son adecuados para estos datos. Ya que según el gráfico de los residuos de la centralidad de cada presidente, no se observa presencia de heterocedasticidad y en las pruebas estadísticas no se observa autocorrelación de los residuos y se observa normalidad de los residuos.

6. Modelamiento de la relación entre centralidad y popularidad

Se desea conocer si existe una relación entre la centralidad de cada presidente con su popularidad. Esta hipótesis es plausible ya que los Estados democráticos, la formación de la opinión de publica de un mandatario depende de como este aparezca en los medios de comunicación.

La formación de la opinión pública es un proceso dinámico, en la cual los ciudadanos se informan sobre la gestión del mandatario. La información que recibe el ciudadano viene de diferentes fuentes: uno son las informaciones de los medios de comunicación y otra son las opiniones que vierten las personas allegadas al ciudadano.

Figura 9: Residuos y su correlograma, centralidad de Lucio Gutierrez

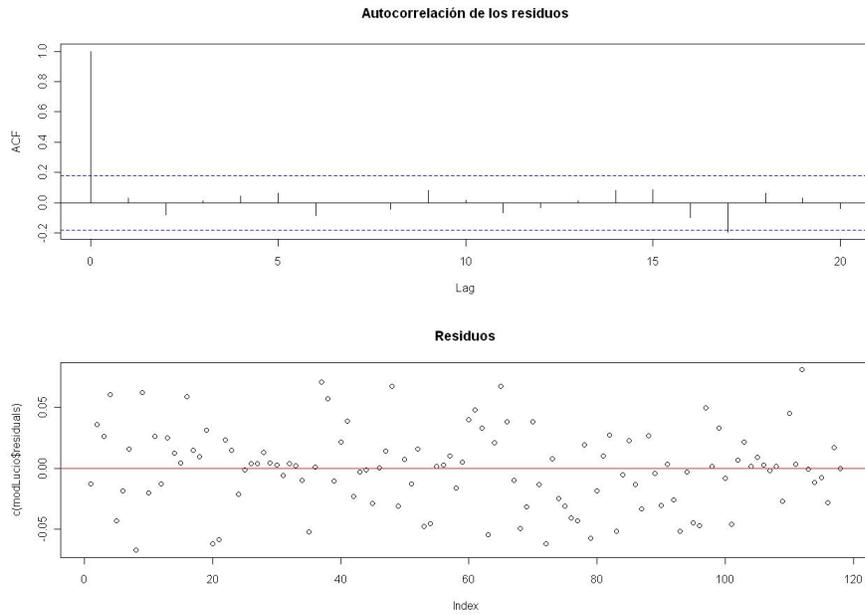


Figura 10: Residuos y su correlograma, centralidad de Alfredo Palacio

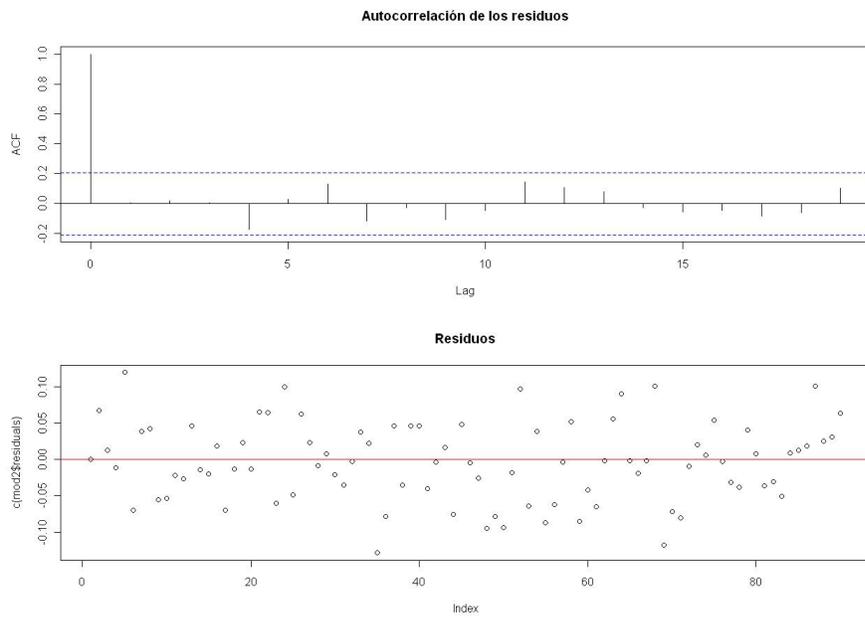
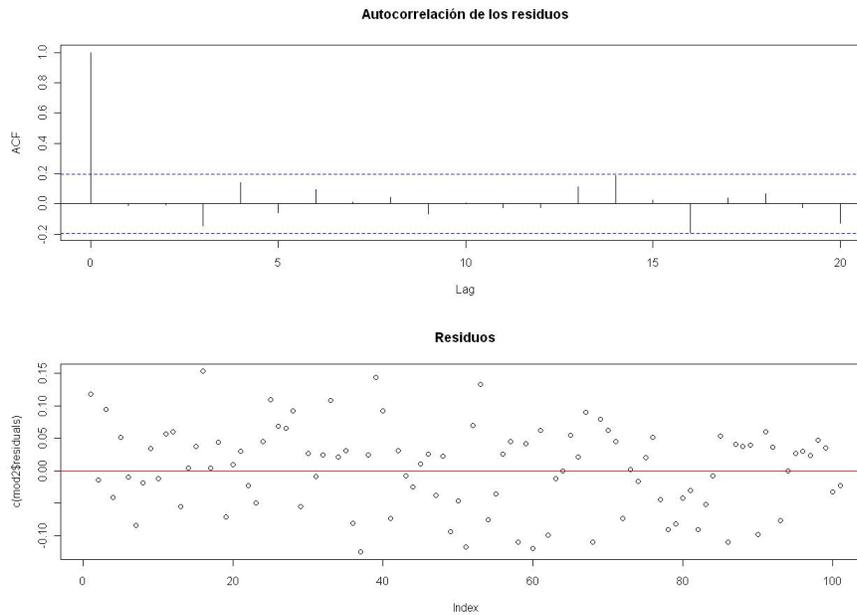


Figura 11: Residuos y su correlograma, centralidad de Rafael Correa



Pruebas Estadísticas	Gustavo Noboa		Lucio Gutiérrez		Alfredo Palacio		Rafael Correa	
	Estadístico	Valor p	Estadístico	Valor p	Estadístico	Valor p	Estadístico	Valor p
Durbin Watson	2.026339	p>0.1	2.005938	p>0.1	2.005938	p>0.1	2.005938	p>0.1
shapiro-Wilk	0.9719	0.3674	0.9852	0.2245	0.9852	0.2245	0.9795	0.1171
Box-Lung(1)	0.01	0.90	0.13	0.72	0.10	0.75	0.01	0.91
Box-Lung(2)	0.04	0.98	0.88	0.64	0.39	0.82	0.02	0.99
Box-Lung(3)	0.96	0.81	0.91	0.82	0.60	0.90	2.21	0.53
Box-Lung(4)	2.95	0.57	1.15	0.89	3.96	0.41	4.30	0.37
Box-Lung(5)	3.38	0.64	1.68	0.89	3.96	0.55	4.69	0.45
Box-Lung(6)	4.85	0.56	2.56	0.86	4.57	0.60	5.66	0.46
Box-Lung(7)	5.30	0.62	2.56	0.92	6.71	0.46	5.67	0.58
Box-Lung(8)	5.38	0.72	2.80	0.95	7.62	0.47	5.88	0.66
Box-Lung(9)	5.46	0.79	3.68	0.93	8.16	0.52	6.41	0.70
Box-Lung(10)	6.79	0.75	3.72	0.96	8.64	0.57	6.41	0.78
Box-Lung(11)	6.82	0.81	4.26	0.96	11.53	0.40	6.50	0.84
Box-Lung(12)	6.89	0.86	4.41	0.97	12.21	0.43	6.57	0.88
Box-Lung(13)	7.44	0.88	4.42	0.99	12.63	0.48	8.06	0.84
Box-Lung(14)	7.50	0.91	5.37	0.98	12.90	0.53	12.20	0.59
Box-Lung(15)	13.70	0.55	6.37	0.97	13.47	0.57	12.27	0.66
Box-Lung(16)	13.76	0.62	7.71	0.96	14.23	0.58	16.86	0.39

Cuadro 11: Pruebas Estadísticas para los residuos de los modelos ARIMA

Existen algunos modelos [26] que explican el proceso de formación de la opinión pública, hay modelos que describen este proceso en base a la estructura de vecindad que tiene el ciudadano, otros modelos en cambio incluyen la mecánica de convencimiento entre un ciudadano a otro o de un grupo de ciudadanos a otro grupo de ciudadanos.

En este trabajo se explica el proceso de formación de la opinión, en base a componentes de la centralidad del presidente como son el nivel de la dinámica comunicacional y la volatilidad en la comunicación, los cuales funcionan como catalizadores del proceso de formación, es decir estos componentes podrían acelerar o desacelerar el proceso de formación de la opinión pública.

El proceso de formación se modela como un proceso dinámico en la cual, se asume como hipótesis, que los ciudadanos a favor y los ciudadanos en contra de un mandatario esta perfectamente distribuidos en el país. Con esta hipótesis hace posible que se modele el porcentaje de opinión favorable $P(t)$ como una ecuación diferencial logística.

$$\frac{dP}{dt} = a(1 - P)P \quad (26)$$

Donde a es un factor de proporcionalidad, que regula la velocidad de homogenización de la opinión pública.

La expresión $a(1 - P)P$ es proporcional a la cantidad de enlaces o de encuentros entre los opinan a favor $P(t)$ y los que opinan en contra $1 - P(t)$. La ecuación 26 nos dice la velocidad en formación de la opinión favorable depende de la cantidad de encuentros que existan entre los que están a favor y los que están en contra.

La solución de la ecuación 26 tiene la siguiente forma

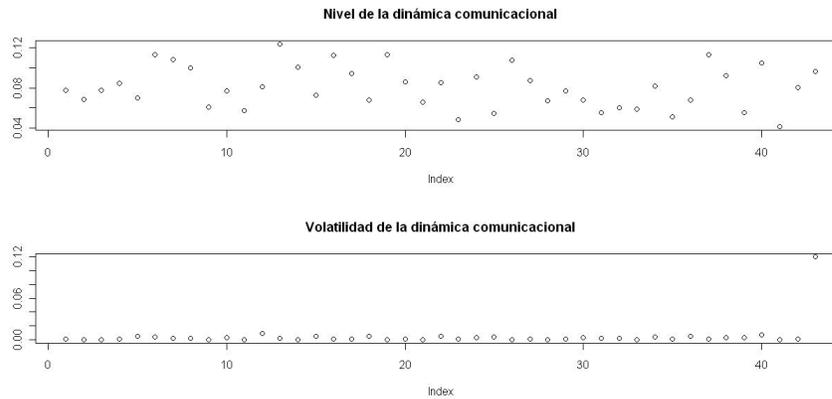
$$P(t) = \frac{1}{1 + \exp(\beta(t))} \quad (27)$$

Esta función no es otra que el inverso de la función logit, donde $\beta(t)$ es nivel de activación de la función, que despejando este factor en la ecuación 27 se obtiene la función logit.

$$\beta(t) = \log \left(\frac{1 - P(t)}{P(t)} \right) \quad (28)$$

Y en este el punto es donde se une la parte dinámica en la formación de opinión pública con la centralidad del presidente, modelando no la popularidad sino que se modela el factor de activación $\beta(t)$ de la ecuación 27 en base a los componentes de la centralidad. Es decir factor de activación, que es el logit de la popularidad de cada mandatario va estar en función de los componentes de la centralidad.

Figura 12: Componentes de la centralidad de Gustavo Noboa



6.1. Popularidad y Componentes de la centralidad

La popularidad de los presidentes de la República se la obtuvo de la empresa Informe Confidencial. Esta empresa no realiza estudios de opinión de pública cada semana, sino que por lo general lo realiza cada 15 días sobre las ciudades de Quito y Guayaquil. Para comparar la popularidad con la centralidad, que es semanal, se tuvo que interpolar linealmente valores semanales de la popularidad, a partir de los valores registrados de esta popularidad, tanto para Quito y Guayaquil.

Con los valores interpolados de Quito y Guayaquil se obtuvo una popularidad promedio del país, ponderando las popularidades de Quito y Guayaquil, en base a la cantidad de habitantes en cada ciudad. En las figuras 13, 15, 17 y 19 se muestran las popularidades de cada mandatario.

Los componentes de la centralidad se lo obtuvo, para cada presidente, en base a los modelos ARIMA de la tabla 8. La parte determinista del modelo ARIMA es el componente, que le damos el nombre de: Dinámica Comunicacional y el estimado de la varianza de las observaciones del modelo ARIMA es el componente que le damos el nombre de: Volatilidad. En las figuras 12, 14, 16 y 18 se muestran los componentes de la centralidad de cada mandatario.

En el gráfico 17 de la popularidad de Alfredo Palacio se observa una meseta entre las semanas 61 a la 90, esta meseta es debido a que la empresa Informe Confidencial no realizó observaciones de la popularidad del presidente, en ese periodo. Por lo tanto se completo la popularidad de Alfredo Palacio, en esas semanas, con la última popularidad observada.

De los gráficos anteriores se observa que en orden de popularidad de los presidentes, de mayor a menor es: Correa, Noboa, Palacio y Gutiérrez

6.2. Modelo de la popularidad de los presidentes

Se plantean dos modelos que relacionan la popularidad con los componentes de la centralidad. Estos modelos se basan en las siguientes hipótesis

Figura 13: Popularidad de Gustavo Noboa

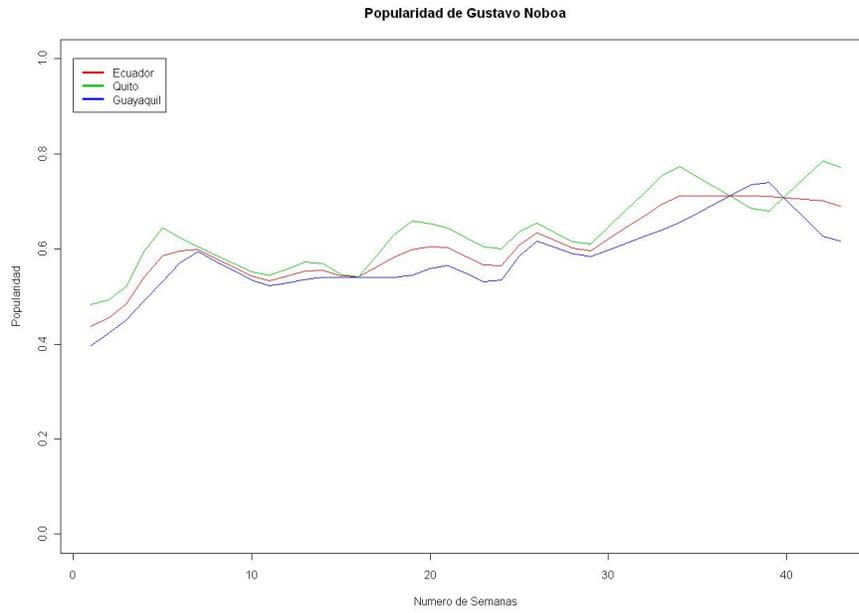


Figura 14: Componentes de la centralidad de Lucio Gutiérrez

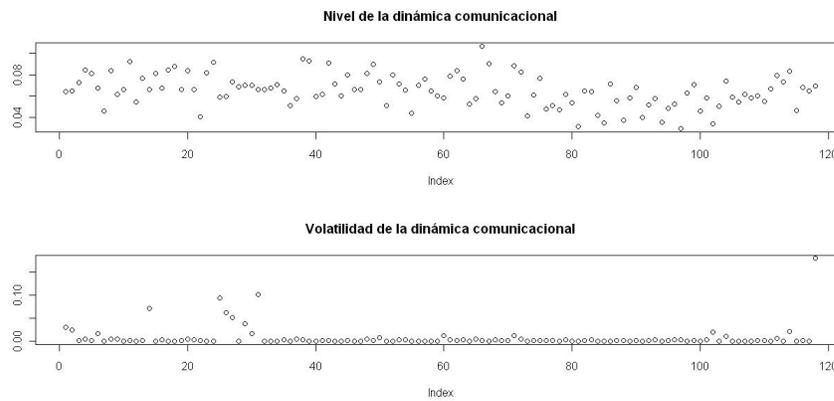


Figura 15: Popularidad de Lucio Gutiérrez

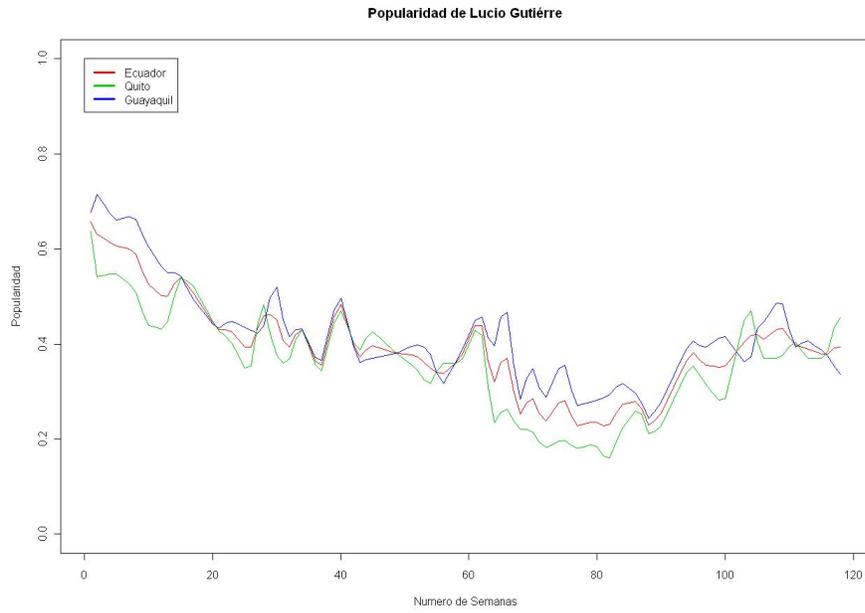


Figura 16: Componentes de la centralidad de Alfredo Palacio

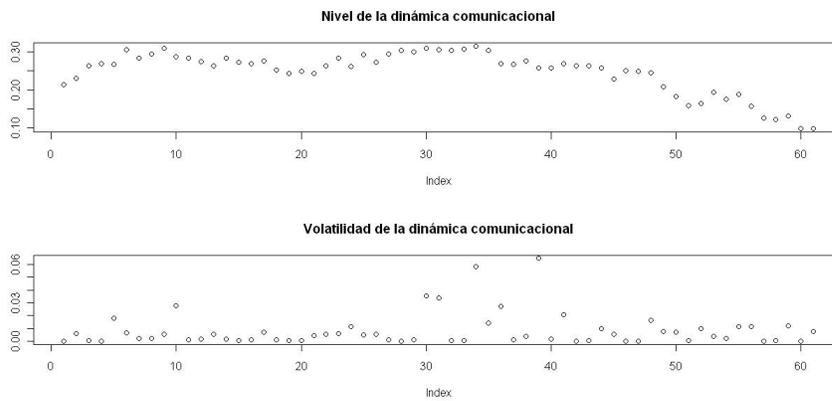


Figura 17: Popularidad de Alfredo Palacio

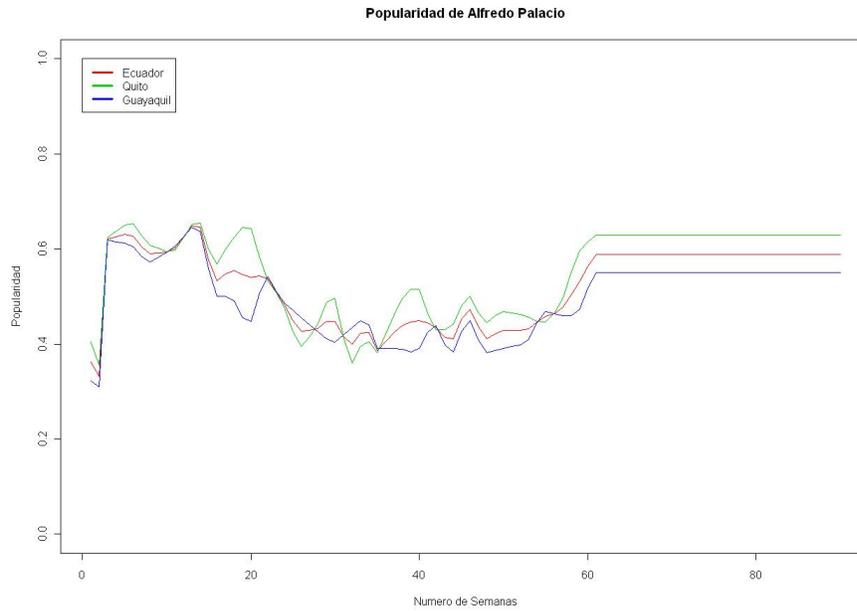


Figura 18: Componentes de la centralidad de Rafael Correa

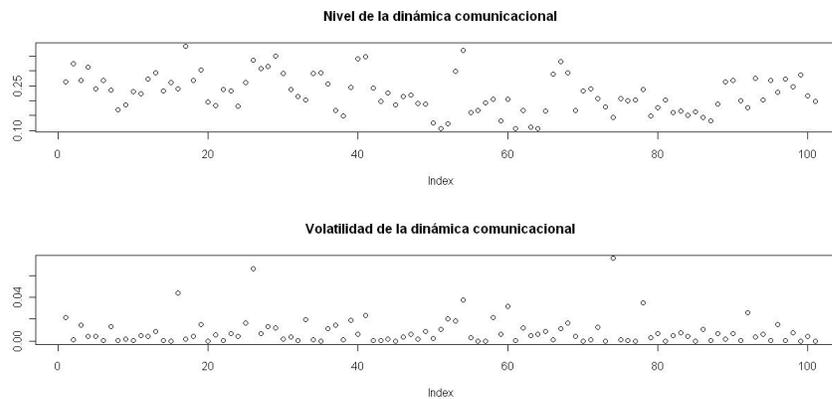
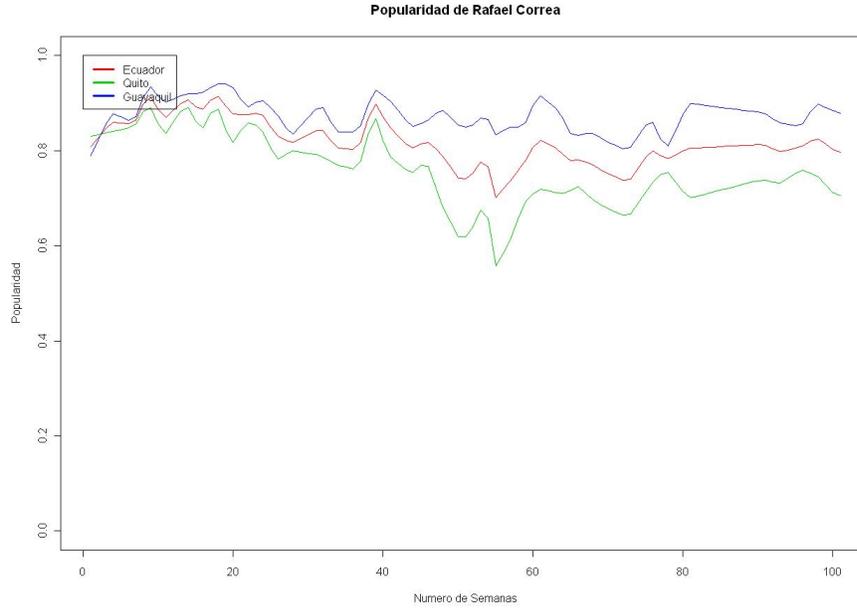


Figura 19: Popularidad de Rafael Correa



- La relación entre centralidad y popularidad es dependiente a cada mandatario
- El nivel de la dinámica comunicacional $Esti_t$ dada por el estimado del modelo ARIMA de la centralidad y la volatilidad semanal de la centralidad $Vola_t$, son criterios independientes y lineales para la formulación de la opinión acerca del presidente.
- Existe un tiempo finito entre que aparece la noticia hasta que se forme una opinión con respecto al presidente. Este tiempo varía desde una semana hasta cuatro semanas
- La información que provee los medios de comunicación afectan directamente el factor de activación $\beta(t)$ de la ecuación 27. Es decir los medios de comunicación aporta insumos en la formación de la publica, haciendo que la activación tenga valores negativos o positivos. Así que la variable dependiente será el factor de activación $\beta(t) = \text{logit}(Pop_t)$

El primer modelo plantea un modelo dinamico lineal, el cual tiene como variables dependiente el logit de la popularidad y como variables regresoras los valores rezagados de de 1 a 4 semanas del logit de la popularidad y los componentes de la centralidad $Esti_t, Vola_t$.

$$\text{logit}(Pop_t) = \text{Intercepto} + \sum_{i=1}^4 \alpha_i Esti_{t-i} + \sum_{i=1}^4 \beta_i Vola_{t-i} + \sum_{i=1}^4 \gamma_i \text{logit}(Pop_{t-i}) + \epsilon \quad (29)$$

El segundo modelo plantea que la diferencia de los logit de la popularidad son explicados en base a los componentes de la centralidad

$$\text{logit}(Pop_t) - \text{logit}(Pop_{t-1}) = \text{Intercepto} + \sum_{i=1}^4 \alpha_i \text{Esti}_{t-i} + \sum_{i=1}^4 \beta_i \text{Vola}_{t-i} + \epsilon \quad (30)$$

Se ajustará cada modelo de acuerdo a cada presidente.

6.2.1. Selección de variables en el modelo centralidad-popularidad

Para cada tipo de modelo y presidente se plantea tres modelos para la popularidad: en todo el Ecuador, en la ciudad de Quito y en la ciudad de Guayaquil. Y en cada modelo se eliminan las variables no significativas del modelo general, usando la prueba de significancia sobre los coeficientes dejan las variables cuyo coeficientes tenga un p menor 0.1.

Para la eliminación de variables, se usará, matriz de covarianzas corregida de Newey West, mas de 60, y esto se aplicará en el caso de Lucio Gutiérrez, Alfredo Palacio y Rafael Correa. Para el caso de Gustavo Noboa, que tiene 43 observaciones, se usará la matriz de varianza de los mínimos cuadrados ordinarios.

Posteriormente se determinan la normalidad de los residuos, si los residuos no son normales, se mejora la normalidad de estos, creando variables dummy en el modelo, para los valores aberrantes. La creación de estas variables dummy esta condicionada a la mejora de la normalidad de los residuos, si la normalidad no mejora, se eliminan estas variables dummy y se deja el modelo original.

6.2.2. Modelo para la popularidad de Gustavo Noboa

En los cuadros 12 y 13 se muestra los modelos para la popularidad de Gustavo Noboa tanto en Ecuador, Quito y Guayaquil, basado en el modelo tipo 1 29 y en el modelo tipo 2 30 y una vez eliminadas las variables no significativas. Y para este no se agregaron variables dummy que absorban valores aberrantes de la popularidad para el modelo tipo 1.

En el modelo tipo 1 para la popularidad en Ecuador y en Quito se observa que ningún de los componentes de la centralidad no intervienen en la formación de la opinión pública. Esto sugiere que la formación de la opinión pública en Ecuador y en Quito depende de otros factores.

En cambio en el modelo tipo 1 para la popularidad en Guayaquil, tiene los componentes de nivel comunicacional y volatilidad rezagadas 2 y 1 semana respectivamente.

En los modelos tipo 2, se refleja la situación inversa de los demás presidentes: el componente de la dinámica comunicacional tiene un impacto negativo y el componente de la volatilidad tiene un impacto positivo. Una forma de interpretar este resultado es que el periodo de tiempo de las observaciones de Gustavo Noboa corresponde a la campaña electoral donde gano Lucio Gutierrez, en este periodo de tiempo la figura del presidente en funciones pasa a un segundo plano, y parece que las intervenciones excepcionales de Gustavo Noboa mejoraban su popularidad, antes que las intervenciones periódicas.

Variables	Modelo Ecuador		Modelo Quito		Modelo Guayaquil	
	Coefficiente	Valor p	Coefficiente	Valor p	Coefficiente	Valor p
<i>Intercepto</i>					0.126432	0.002619
<i>Esti</i> ₁						
<i>Esti</i> ₂					-1.690164	0.000342
<i>Esti</i> ₃						
<i>Esti</i> ₄						
<i>Vola</i> ₁					8.095856	0.035637
<i>Vola</i> ₂						
<i>Vola</i> ₃						
<i>Vola</i> ₄						
<i>logit</i> _{E1}	1.964669	0	1.892353	0	1.986336	0
<i>logit</i> _{E2}	-1.517925	1e-06	-1.458055	1e-06	-1.740111	4e-06
<i>logit</i> _{E3}	0.524196	0.000618	0.532777	0.000248	1.021209	0.003237
<i>logit</i> _{E4}					-0.333394	0.041892

Cuadro 12: Modelos tipo 1 para la popularidad de Gustavo Noboa

Variables	Modelo Ecuador		Modelo Quito		Modelo Guayaquil	
	Coefficiente	Valor p	Coefficiente	Valor p	Coefficiente	Valor p
<i>Intercepto</i>						
<i>Esti</i> ₁						
<i>Esti</i> ₂					-0.5571	0.000406
<i>Esti</i> ₃	-0.3715	0.0173				
<i>Esti</i> ₄			-0.3523	0.120		
<i>Vola</i> ₁						
<i>Vola</i> ₂						
<i>Vola</i> ₃						
<i>Vola</i> ₄					1.7898	0.005989

Cuadro 13: Modelos tipo 2 para la popularidad de Gustavo Noboa

La centralidad de Gustavo Noboa solo tuvo solo un valor aberrante, lo cual indica que este presidente tuvo un comportamiento adecuado con el modelo ARIMA que se le ajusto. Este comportamiento también se refleja en la popularidad en la cual no se tuvo que agregar variables dummy para absorber algún valor aberrante.

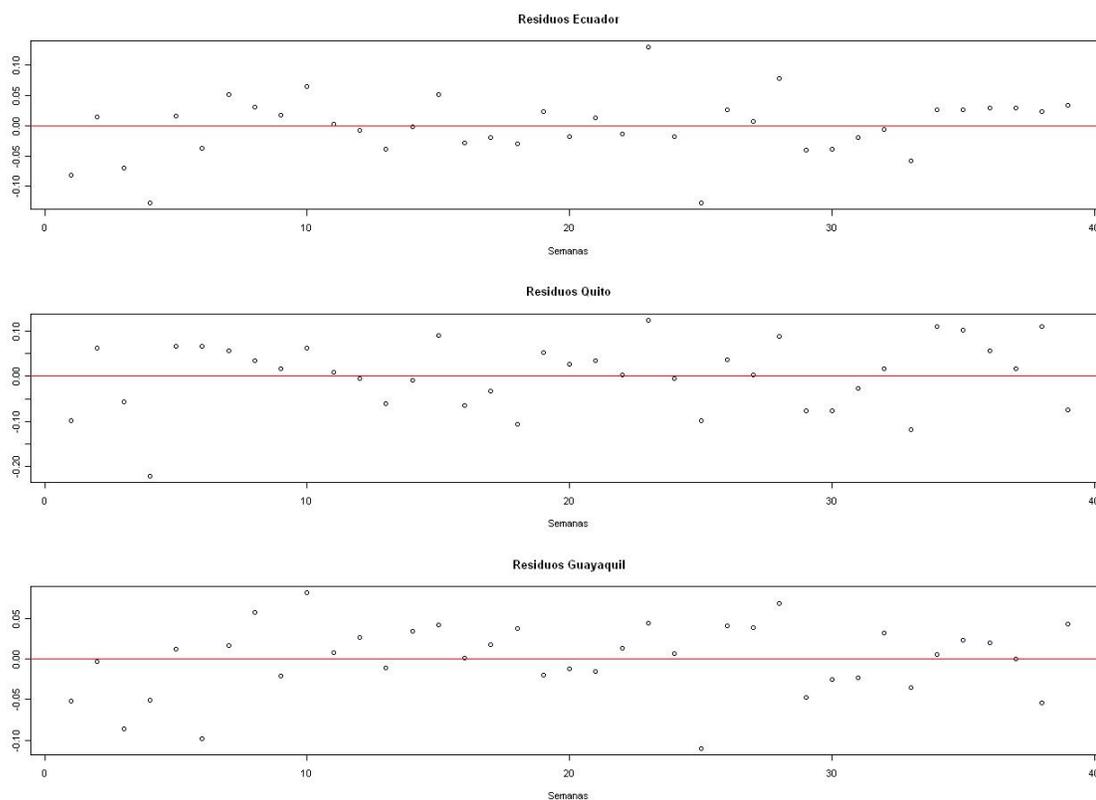
Residuos del modelo de popularidad para Gustavo Noboa

A los residuos de los modelos de las tablas 14 y 15 se aplicaron las pruebas estadísticas para verificar la idoneidad sus residuos.

De las pruebas realizadas, se puede concluir que no se puede rechazar la hipótesis de normalidad (Shapiro-Wilk), de la homocedasticidad (Breusch-Pagan) y de la autocorrelación (Durbni-Watson).

Al verificarse la idoneidad de los residuos, se puede afirmar que los parámetros encontrados en los modelos de las tablas 12 y 13 son correctos.

Figura 20: Residuos del modelo de popularidad para Gustavo Noboa



Pruebas Estadísticas	Modelo Ecuador		Modelo Quito		Modelo Guayaquil	
	Estadístico	Valor p	Estadístico	Valor p	Estadístico	Valor p
Breusch-Pagan	2.41	0.299625	6.07	0.639881	4.33	0.631744
Durbin-Watson	2.15	0.72	2.53	0.204	1.57	0.104
Shapiro-Wilk	0.97	0.527478	0.96	0.187892	0.96	0.274266
Box-Lung)(1)	2.52	0.64	11.11	0.03	3.54	0.47
Box-Lung)(2)	2.55	0.77	12.36	0.03	4.73	0.45
Box-Lung)(3)	3.78	0.71	13.96	0.03	6.76	0.34
Box-Lung)(4)	4.52	0.72	15.45	0.03	6.82	0.45
Box-Lung)(5)	5.12	0.74	15.46	0.05	6.88	0.55
Box-Lung)(6)	5.28	0.81	15.53	0.08	7.26	0.61
Box-Lung)(7)	5.32	0.87	15.56	0.11	9.47	0.49
Box-Lung)(8)	6.53	0.84	15.92	0.14	9.56	0.57
Box-Lung)(9)	7.85	0.8	16.07	0.19	10.13	0.6
Box-Lung)(10)	9.47	0.74	16.17	0.24	10.38	0.66
Box-Lung)(11)	10.43	0.73	16.19	0.3	13.08	0.52
Box-Lung)(12)	12.86	0.61	16.26	0.37	14.62	0.48
Box-Lung)(13)	13	0.67	16.86	0.39	15.64	0.48
Box-Lung)(14)	13.11	0.73	17.08	0.45	18.27	0.37
Box-Lung)(15)	13.16	0.78	19.48	0.36	18.33	0.43
Box-Lung)(16)	13.83	0.79	19.49	0.43	20.17	0.38

Cuadro 14: Pruebas Estadísticas de los residuos del modelo tipo 1 para Gustavo Noboa

Pruebas Estadísticas	Modelo Ecuador		Modelo Quito		Modelo Guayaquil	
	Estadístico	Valor p	Estadístico	Valor p	Estadístico	Valor p
Breusch-Pagan	2.41	0.299625	6.07	0.639881	4.33	0.631744
Breusch-Pagan	0.1	0	0.02	0	1.22	0.269052
Durbin-Watson	0.77	0	0.83	0	0.72	0
Shapiro-Wilk	0.95	0.075316	0.9	0.002276	0.97	0.26122
Box-Lung)(1)	21.76	0	23.87	0	17.72	0
Box-Lung)(2)	22.77	0	28.41	0	17.92	0
Box-Lung)(3)	23.17	0	29.3	0	17.92	0.01
Box-Lung)(4)	23.17	0	29.42	0	18.23	0.01
Box-Lung)(5)	23.17	0	29.91	0	18.8	0.02
Box-Lung)(6)	24.83	0	30.33	0	19.36	0.02
Box-Lung)(7)	31.24	0	34.84	0	19.56	0.03
Box-Lung)(8)	33.18	0	38.65	0	19.62	0.05
Box-Lung)(9)	33.93	0	38.72	0	19.67	0.07
Box-Lung)(10)	38.9	0	45.08	0	19.67	0.1
Box-Lung)(11)	41.55	0	55.84	0	20.32	0.12
Box-Lung)(12)	41.63	0	61.19	0	22.15	0.1
Box-Lung)(13)	41.73	0	61.28	0	24.98	0.07
Box-Lung)(14)	42.9	0	62.83	0	27.48	0.05
Box-Lung)(15)	44.49	0	67.81	0	27.8	0.07
Box-Lung)(16)	45.44	0	71.91	0	27.81	0.09

Cuadro 15: Pruebas Estadísticas de los residuos del modelo tipo 2 para Gustavo Noboa

6.2.3. Modelo para la popularidad de Lucio Gutiérrez

En las tablas 16 y 17 se muestran los modelos para la popularidad de Lucio Gutierrez tanto en Ecuador, Quito y Guayaquil, basado en los modelos de las ecuaciones 29 y 30 y una vez eliminadas las variables no significativas de acuerdo con la matriz ajustada Newey-West.

En los modelos tipo 1 para la popularidad en Ecuador y en Quito se agrego variables dummy para mejorar la normalidad de los residuos. Estas variables dummy aparecen en la tabla 16, con la semana correspondiente asociada.

También en los modelos tipo 1 para la popularidad en Ecuador y en Quito aparece la variable del nivel comunicacional, desfasado 3 semanas. En el modelo para Guayaquil aparece el nivel comunicacional desfasada 4 semanas.

En los modelos tipo 2 se refleja la situación donde la dinámica comunicacional tiene un efecto positivo sobre la popularidad.

En el modelo de centralidad de Lucio Gutiérrez tuvo 17 valores aberrantes, es decir 17 semanas donde su centralidad fue realmente inesperada, y las fluctuaciones de estos valores aberrantes forman parte de la volatilidad. Sin embargo estos componentes no fueron significativos en los modelos de la popularidad tipo 1 y tipo 2. Esto sugiere que la popularidad de Lucio Gutiérrez no dependía tanto de sus errores comunicacionales, como se creía sino que la popularidad se basó principalmente en la dinámica comunicacional que tuvo.

A pesar que la popularidad de Lucio Gutiérrez no se vio afectada mayormente por sus errores de comuni-

Variables	Modelo Ecuador			Modelo Quito			Modelo Guayaquil		
	Sem.	Coefficiente	Valor p	Sem.	Coefficiente	Valor p	Sem.	Coefficiente	Valor p
<i>Intercepto</i>		-0.025295	0.422154		-0.046797	0.251431		-0.084151	0.045196
<i>Esti</i> ₁									
<i>Esti</i> ₂									
<i>Esti</i> ₃		0.576365	0.174813		0.728414	0.175233			
<i>Esti</i> ₄								1.212625	0.038014
<i>Vola</i> ₁									
<i>Vola</i> ₂									
<i>Vola</i> ₃									
<i>Vola</i> ₄									
<i>logit</i> _{E1}		1.771944	0		1.834769	0		1.743519	0
<i>logit</i> _{E2}		-1.357192	0		-1.367757	0		-1.48849	0
<i>logit</i> _{E3}		0.608208	0		0.530017	0		1.018411	0
<i>logit</i> _{E4}								-0.305867	0.001798
<i>Outlier</i> ₁	106	0.306605	3e-06	44	0.385447	3e-06			
<i>Outlier</i> ₂	110	0.28431	2.9e-05	70	0.405841	3e-06			
<i>Outlier</i> ₃	130	0.201795	0.00162	106	0.21582	0.006355			
<i>Outlier</i> ₄				130	-0.278058	0.000564			
<i>Outlier</i> ₅				146	-0.210449	0.008636			

Cuadro 16: Modelos tipo 1 para la popularidad de Lucio Gutiérrez

Variables	Modelo Ecuador		Modelo Quito		Modelo Guayaquil	
	Coefficiente	Valor p	Coefficiente	Valor p	Coefficiente	Valor p
<i>Intercepto</i>	-0.098508	8.5e-05	-0.080512	0.03181	-0.092821	0.00763
<i>Esti</i> ₁						
<i>Esti</i> ₂						
<i>Esti</i> ₃			1.316558	0.01132		
<i>Esti</i> ₄	1.672811	4e-06			1.671924	0.000358
<i>Vola</i> ₁						
<i>Vola</i> ₂						
<i>Vola</i> ₃			0.787428	0.000279	-0.910566	0.006159
<i>Vola</i> ₄						

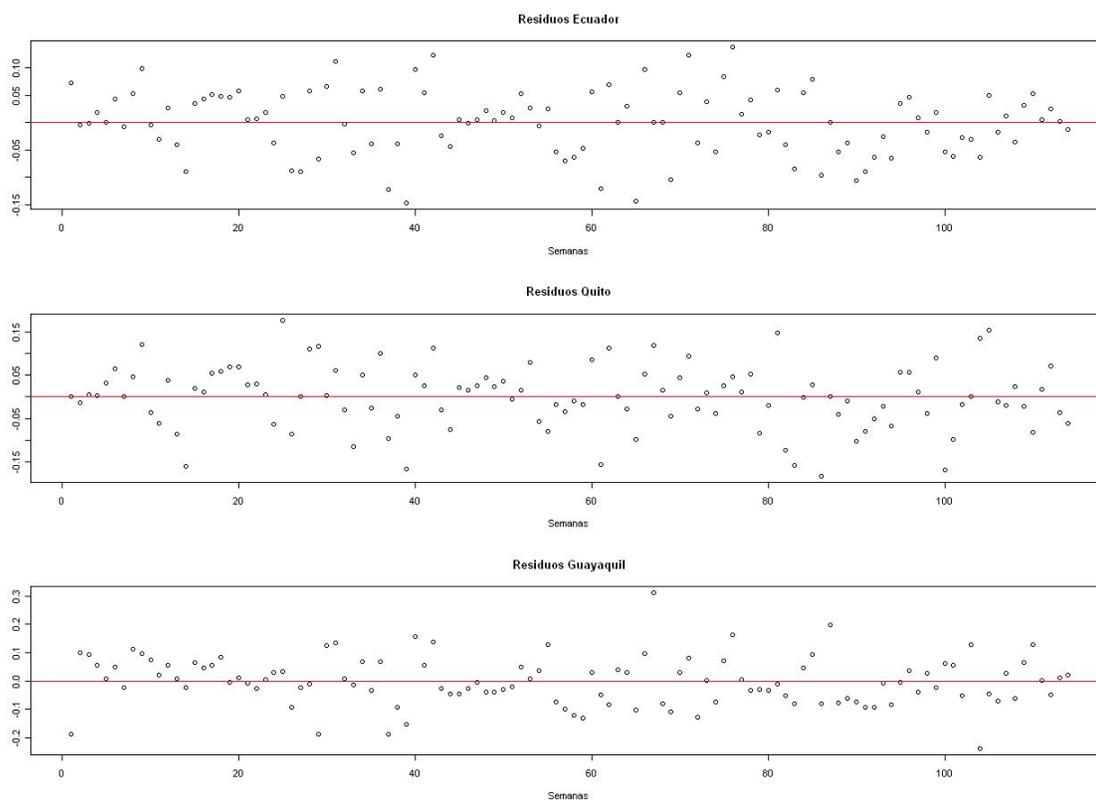
Cuadro 17: Modelos tipo 2 para la popularidad de Lucio Gutiérrez

cación, este presidente fue finalmente derrocado. Esto sugiere que los errores de comunicación no actúan en la popularidad de corto plazo pero podría pensarse que si actúa sobre la gobernabilidad del mandatario en el largo plazo.

Residuos para el modelo de popularidad de Lucio Gutiérrez

Del gráfico 21 se puede apreciar, heterocedasticidad en los residuos y esto se confirma en las pruebas estadísticas del tabla 18. La heterocedasticidad de los residuos no es problemática, ya que se validaron los coeficientes del modelo de popularidad con la matriz de covarianza de Newey-Watson

Figura 21: Residuos del modelo de popularidad para Lucio Gutiérrez



Pruebas Estadísticas	Modelo Ecuador		Modelo Quito		Modelo Guayaquil	
	Estadístico	Valor p	Estadístico	Valor p	Estadístico	Valor p
Breusch-Pagan	10.22	0.176286	3.65	0.933012	3.2	0.669098
Durbin-Watson	2	0.84	1.97	0.724	1.91	0.488
Shapiro-Wilk	0.99	0.593706	0.99	0.420861	0.99	0.396117
Box-Lung)(1)	1.95	0.74	1.4	0.84	0.52	0.97
Box-Lung)(2)	2.97	0.71	5.75	0.33	4.31	0.51
Box-Lung)(3)	3.08	0.8	6.03	0.42	4.32	0.63
Box-Lung)(4)	4.44	0.73	6.46	0.49	5.06	0.65
Box-Lung)(5)	4.58	0.8	6.46	0.6	5.09	0.75
Box-Lung)(6)	4.59	0.87	6.54	0.69	5.91	0.75
Box-Lung)(7)	5.01	0.89	6.59	0.76	6.98	0.73
Box-Lung)(8)	7.28	0.78	7.01	0.8	7.37	0.77
Box-Lung)(9)	7.57	0.82	7.85	0.8	9.34	0.67
Box-Lung)(10)	7.62	0.87	7.9	0.85	9.55	0.73
Box-Lung)(11)	7.64	0.91	7.94	0.89	9.62	0.79
Box-Lung)(12)	9.27	0.86	9.47	0.85	9.74	0.84
Box-Lung)(13)	12.25	0.73	10.99	0.81	9.74	0.88
Box-Lung)(14)	14.12	0.66	11.18	0.85	14.42	0.64
Box-Lung)(15)	14.12	0.72	12.03	0.85	14.57	0.69
Box-Lung)(16)	15.02	0.72	12.24	0.88	16.55	0.62

Cuadro 18: Pruebas Estadísticas para los residuos para modelo tipo 1 de Lucio Gutiérrez

Pruebas Estadísticas	Modelo Ecuador		Modelo Quito		Modelo Guayaquil	
	Estadístico	Valor p	Estadístico	Valor p	Estadístico	Valor p
Breusch-Pagan	2.39	0.12176	0.9	0.638732	0.65	0.722007
Durbin-Watson	1.23	0	1.06	0	1.28	0
Shapiro-Wilk	0.98	0.142765	0.97	0.018003	0.97	0.015592
Box-Lung)(1)	37.27	0	47.99	0	33.44	0
Box-Lung)(2)	37.29	0	48.13	0	33.46	0
Box-Lung)(3)	37.46	0	48.41	0	36.13	0
Box-Lung)(4)	38.91	0	48.51	0	39.56	0
Box-Lung)(5)	39.63	0	48.52	0	39.57	0
Box-Lung)(6)	39.88	0	48.53	0	41.52	0
Box-Lung)(7)	41.34	0	49.55	0	42.38	0
Box-Lung)(8)	44.14	0	51.81	0	42.92	0
Box-Lung)(9)	46.35	0	52.21	0	44.28	0
Box-Lung)(10)	46.53	0	52.36	0	44.58	0
Box-Lung)(11)	46.54	0	54.27	0	44.71	0
Box-Lung)(12)	47.38	0	56.57	0	44.73	0
Box-Lung)(13)	50.22	0	56.57	0	45.82	0
Box-Lung)(14)	51	0	60.57	0	49.88	0
Box-Lung)(15)	51.06	0	64.32	0	51.76	0
Box-Lung)(16)	51.73	0	64.39	0	52.22	0

Cuadro 19: Pruebas Estadísticas para los residuos para modelo tipo 2 de Lucio Gutiérrez

6.2.4. Modelo para la popularidad de Alfredo Palacio

En los cuadros 20 y 21 se muestran los modelos para la popularidad de Alfredo Palacio tanto en Ecuador, Quito y Guayaquil, basado en los modelos de las ecuaciones 29 y 30 y una vez eliminadas las variables no significativas

Para el modelo tipo 2 de popularidad de este presidente se agrego variables dummy para mejorar la normalidad de los residuos. Estos variables tienen asociadas a la semana correspondiente en la tabla 20.

En el modelo de popularidad tipo 1 del presidente Palacio para Ecuador solo aparece la volatilidad comunicacional desfasada 3 semanas.

En el modelo tipo 1 de popularidad para Quito aparece la variable del nivel comunicacional desfasada 3 y 4 semana, pero también aparece la volatilidad desfasada 2 y 3 semanas.

En el modelo tipo 1 de popularidad para Guayaquil solo aparece la volatilidad desfasada 3 y 4 semanas. en los modelos tipo 2 se refleja lo mismo que en los otros presidentes: la dinámica comunicacional tiene un efecto positivo sobre la popularidad y la volatilidad tiene un efecto negativo.

Es de notar que las muchas volatilidades en la centralidad que aparecen en los modelos popularidad se deben que tuvo el presidente Palacio, tuvo un bajo nivel mediático en los medios de comunicación, 7 de los 8 valores aberrantes en su centralidad, se debió a un bajón inesperado de su centralidad, y este tipo de volatilidades es calificado negativamente por la población. Así que la estrategia de “evitar hacer olas” no fue conveniente a la popularidad de este presidente.

Variables	Modelo Ecuador			Modelo Quito			Modelo Guayaquil		
	Sem.	Coefficiente	Valor p	Sem.	Coefficiente	Valor p	Sem.	Coefficiente	Valor p
<i>Intercepto</i>									
<i>Esti₁</i>									
<i>Esti₂</i>									
<i>Esti₃</i>					0.387126	0.659528			
<i>Esti₄</i>					-0.111222	0.899337			
<i>Vola₁</i>									
<i>Vola₂</i>					-3.40752	0.006087			
<i>Vola₃</i>		-0.581317	0.399131		-2.119247	0.088266		-0.296451	0.704496
<i>Vola₄</i>								-0.644932	0.407401
<i>logitE₁</i>		0.951804	0		1.042587	0		0.992859	0
<i>logitE₂</i>		-0.015441	0.800242		-0.185541	0.029322		-0.066909	0.278012
<i>logitE₃</i>									
<i>logitE₄</i>									
<i>Outlier₁</i>	163	0.308187	0.00026	163	-1.04735	2.24e-11	163	0.337817	0.000215
<i>Outlier₂</i>	175	-1.154254	0				175	-1.243393	0
<i>Outlier₃</i>							176	0.230136	0.008177
<i>Outlier₄</i>							181	-0.220339	0.012124

Cuadro 20: Modelos tipo 1 para la popularidad de Alfredo Palacio

Variables	Modelo Ecuador		Modelo Quito		Modelo Guayaquil	
	Coefficiente	Valor p	Coefficiente	Valor p	Coefficiente	Valor p
<i>Intercepto</i>	-0.117485	0.006523			-0.113277	0.028343
<i>Esti₁</i>	0.417013	0.064984			0.398537	0.105343
<i>Esti₂</i>						
<i>Esti₃</i>						
<i>Esti₄</i>						
<i>Vola₁</i>						
<i>Vola₂</i>			-1.307957	0.057637		
<i>Vola₃</i>						
<i>Vola₄</i>						

Cuadro 21: Modelos tipo 2 para la popularidad de Alfredo Palacio

Residuos del modelo para la popularidad de Alfredo Palacio

A los residuos de los modelos de las tablas 22 y 23 se aplicaron las pruebas estadísticas para verificar la idoneidad sus residuos.

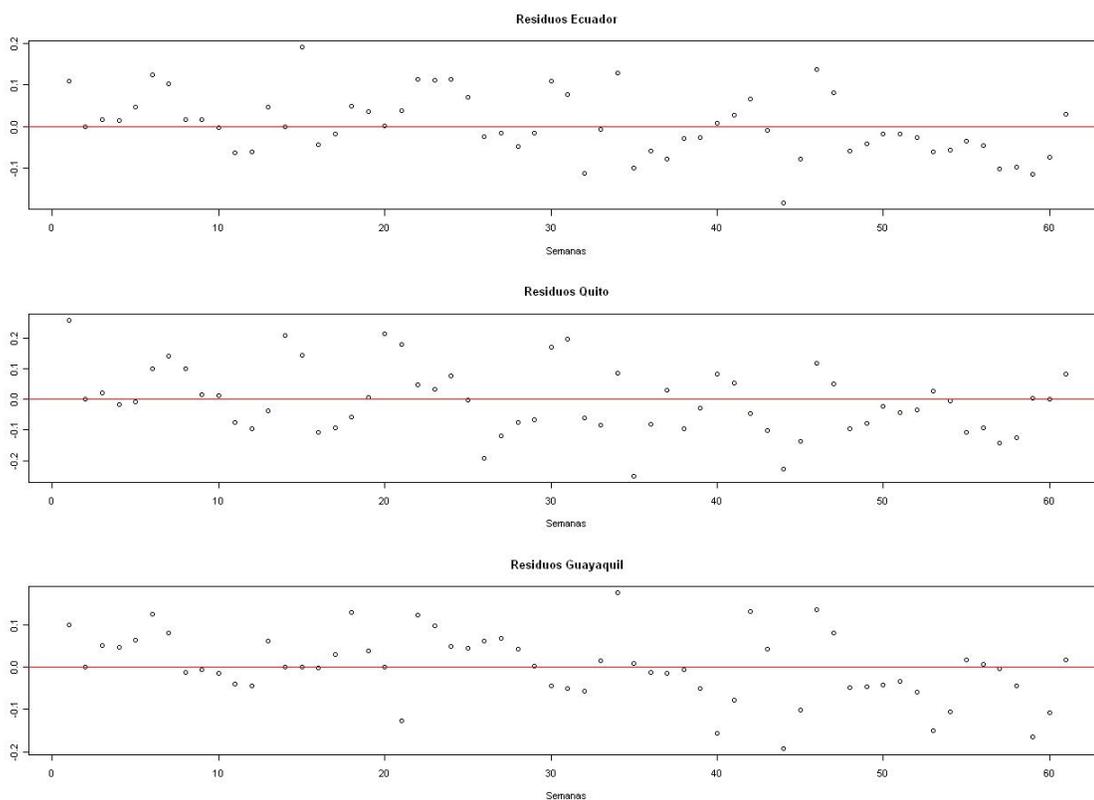
De las pruebas realizadas, se puede observar que los residuos son normales y homocedasticos, pero son autocorrelacionados. La autocorrelacion de los residuos no es problema ya que las variable del modelo fueron seleccionada con la matriz de varianza corregida de Newey Watson .

6.2.5. Modelo para la popularidad de Rafael Correa

En los cuadros 24 y 25 se muestran los modelos para la popularidad de Rafael Correa tanto en Ecuador, Quito y Guayaquil, basado en los modelos de la ecuaciones 29 y 30 y una vez eliminadas las variables no significativas.

Estos modelos no se agregaron variables dummy ya que estas no mejoraban la normalidad de los residuos.

Figura 22: Residuos del modelo de popularidad para Alfredo Palacio



Pruebas Estadísticas	Modelo Ecuador		Modelo Quito		Modelo Guayaquil	
	Estadístico	Valor p	Estadístico	Valor p	Estadístico	Valor p
Breusch-Pagan	3.38	0.49575	17.43	0.007814	3.99	0.78109
Durbin-Watson	1.31	0.004	1.2	0	1.29	0.008
Shapiro-Wilk	0.98	0.518477	0.98	0.424783	0.98	0.582373
Box-Lung)(1)	8.44	0.08	12.93	0.01	14.37	0.01
Box-Lung)(2)	9.1	0.11	14.16	0.01	17.08	0
Box-Lung)(3)	9.81	0.13	17.41	0.01	17.23	0.01
Box-Lung)(4)	10.58	0.16	20.79	0	17.49	0.01
Box-Lung)(5)	13.14	0.11	20.88	0.01	19.34	0.01
Box-Lung)(6)	16.93	0.05	23.69	0	25.21	0
Box-Lung)(7)	17.16	0.07	26.02	0	25.37	0
Box-Lung)(8)	17.92	0.08	26.3	0.01	25.4	0.01
Box-Lung)(9)	19.07	0.09	26.56	0.01	26.51	0.01
Box-Lung)(10)	19.67	0.1	27.14	0.01	26.53	0.01
Box-Lung)(11)	21	0.1	27.18	0.02	26.56	0.02
Box-Lung)(12)	22.42	0.1	27.2	0.03	27.12	0.03
Box-Lung)(13)	29.35	0.02	30.71	0.01	27.3	0.04
Box-Lung)(14)	29.99	0.03	31.43	0.02	27.9	0.05
Box-Lung)(15)	30.37	0.03	31.98	0.02	27.95	0.06
Box-Lung)(16)	30.68	0.04	32.26	0.03	30.38	0.05

Cuadro 22: Pruebas Estadísticas para los residuos para modelos tipo 1 de Alfredo Palacio

Pruebas Estadísticas	Modelo Ecuador		Modelo Quito		Modelo Guayaquil	
	Estadístico	Valor p	Estadístico	Valor p	Estadístico	Valor p
Breusch-Pagan	0.15	0.695092	0.48	0	0.13	0.713994
Durbin-Watson	2.02	0.874	1.74	0.294	1.92	0.478
Shapiro-Wilk	0.52	0	0.74	0	0.57	0
Box-Lung)(1)	0.43	0.98	2.76	0.6	0.94	0.92
Box-Lung)(2)	0.47	0.99	3.26	0.66	1.03	0.96
Box-Lung)(3)	0.48	1	3.29	0.77	1.31	0.97
Box-Lung)(4)	0.51	1	3.39	0.85	1.35	0.99
Box-Lung)(5)	0.98	1	3.43	0.9	1.99	0.98
Box-Lung)(6)	1.87	0.99	4.82	0.85	3.25	0.95
Box-Lung)(7)	2.06	1	6.61	0.76	3.72	0.96
Box-Lung)(8)	2.13	1	6.61	0.83	3.72	0.98
Box-Lung)(9)	4.86	0.96	9.01	0.7	5.71	0.93
Box-Lung)(10)	6.21	0.94	10	0.69	6.86	0.91
Box-Lung)(11)	6.21	0.96	10	0.76	6.86	0.94
Box-Lung)(12)	6.24	0.98	10.22	0.81	6.95	0.96
Box-Lung)(13)	6.3	0.98	11.77	0.76	8.08	0.95
Box-Lung)(14)	6.3	0.99	11.81	0.81	8.49	0.95
Box-Lung)(15)	6.33	0.99	14.37	0.7	9.71	0.94
Box-Lung)(16)	6.35	1	15	0.72	11.06	0.92

Cuadro 23: Pruebas Estadísticas para los residuos para modelos tipo 2 de Alfredo Palacio

En todos los modelos tipo 1, tanto para Ecuador, Quito y Guayaquil, aparecen como variables regresoras el nivel comunicacional del presidente, desfasado 4 semanas y la volatilidad en la comunicación desfasada 3 semanas, con valores positivo y negativo respectivamente.

En los modelos tipo 2, so se llega explicar la popularidad en Quito y no se obtiene variables significativas en ni en Ecuador ni en Guayaquil, esto se interpreta que la popularidad del presidente Correa es mas compleja como para ser explicada con estos dos componentes.

Residuos del modelo para la popularidad de Rafael Correa

A los residuos de los modelos de las tablas 26 y 27 se aplicaron las pruebas estadísticas para verificar la idoneidad de sus residuos.

De las pruebas realizadas, se puede observar que los residuos son independientes, pero no normales ni homocedásticos. Este problema no se pudo corregir agregando variables dummy, sin embargo las variables de este modelo fueron seleccionadas de acuerdo con la matriz de varianza corregida de Newey Watson, la cual es inmune a la heterocedasticidad de los residuos.

De la figura 23 se observa que los residuos de la popularidad, son bastante grande al inicio de l mandato de Correa, sin embargo estos residuos tienden a ser pequeños, a partir de la semana 52 del mandato de Correa, especialmente en Ecuador y en Quito. En Guayaquil continua con un patrón errático. Esto sugiere que a partir de la semana 52, la relación entre centralidad y popularidad entra en una fase mas estable.

Variables	Modelo Ecuador		Modelo Quito		Modelo Guayaquil	
	Coefficiente	Valor p	Coefficiente	Valor p	Coefficiente	Valor p
<i>Intercepto</i>					0.19267	0.045248
<i>Esti</i> ₁						
<i>Esti</i> ₂						
<i>Esti</i> ₃						
<i>Esti</i> ₄	0.351629	0.02827	0.26887	0.075837	0.339705	0.085556
<i>Vola</i> ₁						
<i>Vola</i> ₂						
<i>Vola</i> ₃	-1.198907	0.124092	-1.381039	0.111756	-1.201354	0.213914
<i>Vola</i> ₄						
<i>logitE</i> ₁	1.544235	0	1.487604	0	1.658437	0
<i>logitE</i> ₂	-1.060551	0	-1.039004	0	-1.283494	0
<i>logitE</i> ₃	0.470963	1e-06	0.514862	0	0.617192	0.001718
<i>logitE</i> ₄					-0.125934	0.232757

Cuadro 24: Modelos tipo 1 para la popularidad de Rafael Correa

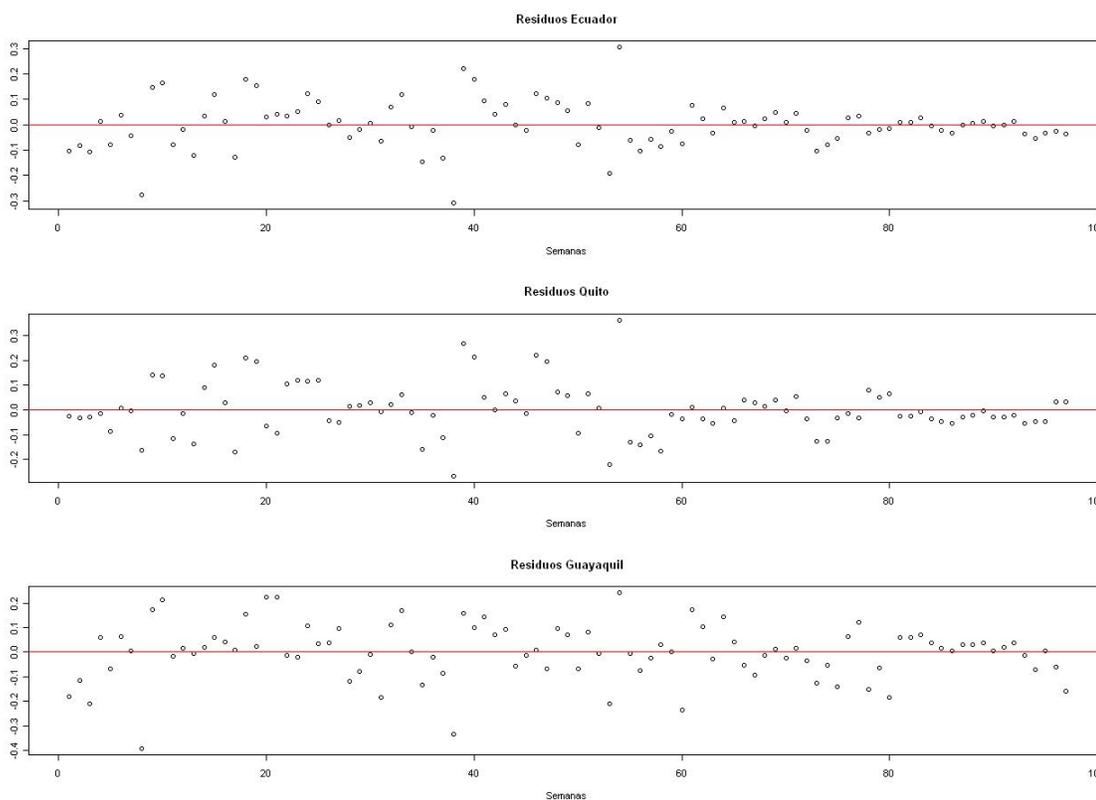
Variables	Modelo Ecuador		Modelo Quito		Modelo Guayaquil	
	Coefficiente	Valor p	Coefficiente	Valor p	Coefficiente	Valor p
<i>Intercepto</i>						
<i>Esti</i> ₁			0.420619	0.093904		
<i>Esti</i> ₂			-0.392951	0.085102		
<i>Esti</i> ₃						
<i>Esti</i> ₄						
<i>Vola</i> ₁						
<i>Vola</i> ₂						
<i>Vola</i> ₃						
<i>Vola</i> ₄						

Cuadro 25: Modelos tipo 2 para la popularidad de Rafael Correa

Pruebas Estadísticas	Modelo Ecuador		Modelo Quito		Modelo Guayaquil	
	Estadístico	Valor p	Estadístico	Valor p	Estadístico	Valor p
Breusch-Pagan	12.39	0.014673	9.18	0.056806	15.47	0.016907
Durbin-Watson	1.87	0.462	1.86	0.46	1.91	0.514
Shapiro-Wilk	0.96	0.01081	0.96	0.004488	0.97	0.01856
Box-Lung)(1)	5.21	0.27	3.93	0.42	4.49	0.34
Box-Lung)(2)	5.22	0.39	4.13	0.53	4.59	0.47
Box-Lung)(3)	5.68	0.46	4.14	0.66	5.85	0.44
Box-Lung)(4)	6.48	0.48	4.85	0.68	9.02	0.25
Box-Lung)(5)	6.49	0.59	5.23	0.73	9.41	0.31
Box-Lung)(6)	6.53	0.69	5.26	0.81	9.42	0.4
Box-Lung)(7)	6.92	0.73	5.64	0.84	9.74	0.46
Box-Lung)(8)	11.16	0.43	9.71	0.56	11.48	0.4
Box-Lung)(9)	11.78	0.46	10.58	0.57	11.78	0.46
Box-Lung)(10)	14.64	0.33	12.59	0.48	12.7	0.47
Box-Lung)(11)	14.64	0.4	12.6	0.56	13.37	0.5
Box-Lung)(12)	20.96	0.14	17.06	0.32	15.12	0.44
Box-Lung)(13)	25.95	0.05	19.7	0.23	16.22	0.44
Box-Lung)(14)	28.11	0.04	19.73	0.29	21.01	0.23
Box-Lung)(15)	28.51	0.05	20.39	0.31	21.38	0.26
Box-Lung)(16)	31.34	0.04	23.63	0.21	22.4	0.26

Cuadro 26: Pruebas Estadísticas para los residuos para los modelos tipo 1 de Rafael Correa

Figura 23: Residuos del modelo de popularidad para Rafael Correa



Pruebas Estadísticas	Modelo Ecuador		Modelo Quito		Modelo Guayaquil	
	Estadístico	Valor p	Estadístico	Valor p	Estadístico	Valor p
Breusch-Pagan	0	0	0.12	0.726961	0	0
Durbin-Watson	1.23	0	1.4	0.002	1.04	0
Shapiro-Wilk	0.97	0.052786	0.95	0.001309	0.99	0.341998
Box-Lung)(1)	29.78	0	31.57	0	40.35	0
Box-Lung)(2)	30.29	0	33.86	0	40.59	0
Box-Lung)(3)	30.58	0	33.87	0	41.57	0
Box-Lung)(4)	30.71	0	34.08	0	43.78	0
Box-Lung)(5)	31.57	0	34.3	0	44.82	0
Box-Lung)(6)	33.25	0	35.65	0	45.05	0
Box-Lung)(7)	33.53	0	35.66	0	45.76	0
Box-Lung)(8)	42.47	0	42.31	0	49.4	0
Box-Lung)(9)	51.62	0	48.92	0	52.61	0
Box-Lung)(10)	52.82	0	51.19	0	53.21	0
Box-Lung)(11)	58.04	0	53.01	0	54.72	0
Box-Lung)(12)	67.89	0	61.37	0	59.63	0
Box-Lung)(13)	68.29	0	61.41	0	59.75	0
Box-Lung)(14)	72.61	0	62.01	0	66.78	0
Box-Lung)(15)	74.57	0	62.8	0	72.61	0
Box-Lung)(16)	77.85	0	64.62	0	74.85	0

Cuadro 27: Pruebas Estadísticas para los residuos para los modelos tipo 2 de Rafael Correa

7. Conclusiones

El índice de centralidad del presidente, el cual está basado en la centralidad de los grafos de temas principales de la semana, sí tiene incidencia en la popularidad del presidente.

Este proceso de formación de la opinión pública tiene una duración de 3 a 4 semanas. Aunque hubieron presidentes como Gustavo Noboa que tuvo un componente de volatilidad temprano de una semana de rezago.

El nivel comunicacional generalmente tiene un efecto positivo y la volatilidad generalmente tiene un efecto negativo, Gustavo Noboa es la excepción del resto de presidentes, el nivel comunicacional tiene en cambio un efecto negativo y la volatilidad tiene un efecto positivo, para el caso de popularidad en Guayaquil.

Una posible explicación a esta anomalía es que el número de observaciones de Gustavo Noboa, son apenas de 43 y todas corresponden a un periodo de elecciones, donde la figura del presidente pasa a un segundo plano. Bajo este periodo específicamente, la aparición continua que aumenta el nivel comunicacional genera un efecto negativo en la opinión pública, en cambio apariciones excepcionales generan un efecto positivo. Esta última afirmación no se pudo corroborar con los datos de Palacio, ya que no se cuenta con valores de la popularidad de este presidente en los últimos meses de su gestión.

La afirmación anterior sugiere que el proceso de opinión pública, está sujeta a cambios estructurales al interior de cada mandato presidencial. Esto es algo plausible ya que el proceso de formación de la opinión pública no es un proceso estático, sino que es un proceso que la población aprende como valorar la gestión de cada mandatario. Y esta valoración, que este trabajo se representa como modelo lineal de popularidad-lineal, no tiene por qué ser fija en el tiempo.

En los modelos tipo 2 se omiten las variables de desfasadas del logit de la popularidad, sin embargo se observa que en muchos casos, excepto para el presidente Correa, los componentes de la centralidad sí llagan a explicar la variación del logit de la popularidad de los presidentes.

8. Recomendaciones

En este trabajo se considera a las noticias de la prensa escrita como un grafo de palabras, las cuales están desprovistas de alguna connotación emocional, política o sentimental y lo que importa es la conectividad entre una palabra u otra dentro de una oración.

Se propone dos mejoras en la extracción de la significación de los contenidos noticiosos con el propósito de mejorar la estimación de la relación que existe entre las noticias y la opinión pública a favor del presidente.

Connotación emocional de las palabras sobre la popularidad Apriori existen términos como "corrupción" o "desempleo" o "inflación" que son muy sentidas por la opinión pública. Pero como capturar el peso que tienen estas palabras en la popularidad del presidente?. Cuando estas palabras están relacionadas cer-

canamente con la figura del presidente en la prensa escrita. Una forma de hacerlo es recurrir al análisis de semántica latente LSI [29], el cual mapea en dos dimensiones las co-ocurrencias entre una palabra y otra dentro de un conjunto de texto, de tal forma que las palabras con una mayor correlación están mas próximas entre si.

Esta proximidad entre palabras permiten relacionar numéricamente las palabras cercanas al presidente y su popularidad.

El presidente como agente pasivo o activo de las noticias Otra mejorar, seria diferenciar los contextos gramaticales donde aparece la figura del presidente: Aquellos contextos donde el presidente aparece como el agente activo de la información y aquellos contextos donde aparece como agente pasivo de la información.

La diferenciación de ambos contextos es importante para la opinión pública, ya que el presidente en un contexto activo es un presidente que esta siendo protagonista la acción, esta generando la noticia. En cambio un presidente en un contexto pasivo es el receptor la acciones que expone la noticia, y la mayoría de los casos se considera que son criticas o cosas que debería hacer el presidente.

Teniendo en cuenta ambos contextos se puede asignar un nodo en el grafo semanal de palabras para cada contexto donde se desenvuelva la figura del presidente: sea sujeto activo o pasivo. Entonces se puede calcular dos tipos de centralidades para el presidente: centralidad activa o centralidad pasiva.

9. Anexo

9.1. Códigos en R

9.1.1. Función selectora de modelos ARIMA

```
seleArima = function(serie, sFile, outliers) {  
  imin1=0; imin2=0; imin3=0; imin4=0; imin5=0; imin6=0; imin7=0;  
  minaic = 0;  
  esta = data.frame(0);  
  i=0;  
  for (i1 in 0:3) {  
    for (i2 in 0:2) {  
      for (i3 in 0:3) {  
        for (i4 in 0:2)  
          for (i5 in 0:1)  
            for (i6 in 0:2)  
              for (i7 in 3:17) {  
                if (nAic < minaic) {  
                  minaic = nAic ;  
                  imin1=i1; imin2=i2; imin3=i3; imin4=i4;imin5=i5; imin6=i6; imin7=i7;  
                }  
              }  
            }  
          }  
        }  
      }  
    }  
  }  
  show(paste(i1,i2,i3,i4,i5,i6,i7,imin1,imin2,imin3,imin4,imin5,imin6,imin7,minaic)) ;  
  }  
  write.table(esta, sFile, sep="," )  
  }  
}  
}  
return (list(imin1,imin2,imin3,imin4,imin5,minaic))  
}
```

9.1.2. Función selectora de modelos lineales

Esta función selecciona variables de un modelo lineal de acuerdo con la matriz de covarianza (fcova) ya sea la matriz de mínimos cuadrados ó la de Newey-West.

```
library(lmtest)
library(sandwich)
step2 = function(modini, fcova=vcov, elim )
n = length(attr(modini$terms, "variables"))
sy = attr(modini$terms, "variables")[[2]]
aCol = matrix(0, nrow=n-1)
aCol[1] = "1"
for (i in 3:n)
  aCol[i-1] = as.character( attr(modini$terms, "variables")[[i]])
aSel = rep(TRUE,n-1); abe = NULL
sx = paste(aCol[aSel], collapse="-")
mod = lm(as.formula(paste(sy, - ", sx )), data = modini$model)
sp = coeftest(mod, df = mod$df.residuals, vcov = fcova(mod))
i = which(sp[,4]==max(sp[,4]))
while ( sp[i,4]>0.1)
  if (names(i)[1]==("Intercept"))
    i2 = 1; aCol[1] = "1"

  else
    i2 = find( aCol, names(i)[1] ); aSel[i2] = FALSE

  sx = paste(aCol[aSel], collapse="-")
  mod = lm(as.formula(paste(sy, - ", sx )), data = modini$model)
  sp = coeftest(mod, df = mod$df.residuals, vcov = fcova(mod))
  i = which(sp[,4]==max(sp[,4]))
return (list(mod=mod,abe=as.numeric(names(abe)) ))
```

9.2. Datos

9.2.1. Centralidad, ordenadas de izquierda a derecha

0.0487	0.0619	0.0868	0.0624	0.1430	0.1753	0.1516	0.0567	0.0533	0.0272	0.0592	0.1763	0.1686	0.0841	0.1432
0.1434	0.0666	0.1404	0.1259	0.0534	0.0762	0.0143	0.0712	0.0351	0.1145	0.1236	0.0584	0.0594	0.0496	0.0131
0.0109	0.0135	0.0625	0.0182	0.0240	0.1401	0.1393	0.0373	0.1084	0.0184	0.0409	0.1157	0.4428	0.2361	0.2181
0.0987	0.1450	0.0382	0.1952	0.0617	0.0168	0.1235	0.0462	0.1184	0.0417	0.1012	0.3309	0.0856	0.1262	0.0987
0.0974	0.0971	0.0216	0.0075	0.0639	0.0969	0.0704	0.3641	0.3072	0.2997	0.0817	0.2639	0.1981	0.3847	0.0699
0.0692	0.0607	0.0125	0.0525	0.1280	0.1519	0.0822	0.0814	0.0999	0.0682	0.0685	0.0589	0.0512	0.0666	0.0805
0.1485	0.0582	0.1605	0.0385	0.0957	0.0236	0.0205	0.0457	0.0724	0.0857	0.0489	0.0650	0.1702	0.1264	0.1170
0.0216	0.0731	0.1251	0.1446	0.0803	0.0147	0.0221	0.0988	0.2003	0.0204	0.0492	0.0363	0.0455	0.0074	0.0080
0.0662	0.0045	0.0356	0.0414	0.0921	0.0126	0.0369	0.0575	0.0581	0.0223	0.0640	0.0540	0.0375	0.0436	0.0261
0.0061	0.0325	0.0041	0.0057	0.0792	0.0644	0.1033	0.0381	0.0120	0.1725	0.0723	0.1784	0.0676	0.0575	0.0594
0.0595	0.0331	0.1001	0.0703	0.1599	0.0729	0.2258	0.0391	0.0401	0.0818	0.4925	0.2142	0.3065	0.2781	0.2557
0.4007	0.2265	0.3259	0.3427	0.2353	0.1209	0.2510	0.2350	0.3360	0.2480	0.2577	0.2974	0.1942	0.2195	0.2625
0.2299	0.3101	0.3371	0.2077	0.3691	0.2232	0.3472	0.3279	0.2930	0.3285	0.1210	0.1227	0.3177	0.3337	0.0738
0.1841	0.1031	0.2974	0.2169	0.5132	0.2978	0.1251	0.2624	0.2438	0.1576	0.3021	0.2438	0.2389	0.1172	0.1219
0.0985	0.1792	0.2635	0.1330	0.2198	0.0815	0.0503	0.1139	0.1494	0.0201	0.0955	0.0089	0.0852	0.1185	0.1893
0.1207	0.1328	0.1053	0.1860	0.0061	0.0177	0.0226	0.0785	0.0838	0.0505	0.1293	0.1365	0.0303	0.0210	0.0599
0.0520	0.0316	0.0385	0.0095	0.0117	0.0530	0.0543	0.1860	0.1209	0.1062	0.1400	0.4106	0.2942	0.3891	0.2460
0.3050	0.2480	0.1178	0.1456	0.2317	0.2092	0.2950	0.3401	0.2013	0.2158	0.2671	0.4502	0.3397	0.3337	0.1793
0.2101	0.2589	0.2131	0.1495	0.2469	0.3874	0.5925	0.3894	0.4312	0.2389	0.2483	0.1766	0.2362	0.3426	0.3274
0.2981	0.1503	0.0451	0.1785	0.3809	0.4187	0.1925	0.2228	0.2160	0.1800	0.1833	0.1512	0.1375	0.1447	0.0962
0.0745	0.0040	0.2641	0.4342	0.1756	0.1048	0.1685	0.2064	0.0580	0.2104	0.0255	0.1306	0.0562	0.0425	0.1852
0.2579	0.3204	0.4377	0.1653	0.2342	0.2426	0.2726	0.0948	0.1737	0.4191	0.2404	0.2301	0.2157	0.0510	0.0913
0.0940	0.1917	0.0905	0.0782	0.0842	0.1546	0.0410	0.1612	0.2719	0.3029	0.1826	0.2220	0.3383	0.2127	0.2815
0.2893	0.3498	0.2976	0.3335	0.2830	0.1519	0.1952								

9.2.2. Popularidad en el Ecuador, ordenadas de izquierda a derecha

0.4378	0.4557	0.4847	0.5413	0.5850	0.5965	0.5993	0.5805	0.5617	0.5429	0.5324	0.5427	0.5531	0.5542	0.5431
0.5411	0.5617	0.5823	0.5986	0.6040	0.6028	0.5845	0.5663	0.5655	0.6083	0.6342	0.6177	0.6012	0.5964	0.6209
0.6454	0.6698	0.6943	0.7118	0.7118	0.7118	0.7118	0.7119	0.7110	0.7080	0.7050	0.7019	0.6904	0.6577	0.6318
0.6229	0.6140	0.6068	0.6038	0.6008	0.5887	0.5539	0.5257	0.5136	0.5015	0.5004	0.5265	0.5408	0.5253	0.5082
0.4871	0.4661	0.4450	0.4301	0.4306	0.4263	0.4101	0.3940	0.3932	0.4310	0.4588	0.4618	0.4512	0.4068	0.3931
0.4204	0.4305	0.3973	0.3641	0.3552	0.4070	0.4587	0.4837	0.4416	0.3995	0.3736	0.3881	0.3972	0.3925	0.3879
0.3833	0.3796	0.3784	0.3771	0.3727	0.3606	0.3493	0.3402	0.3372	0.3495	0.3617	0.3790	0.4090	0.4390	0.4388
0.3629	0.3204	0.3613	0.3700	0.2978	0.2530	0.2766	0.2847	0.2542	0.2378	0.2567	0.2756	0.2803	0.2493	0.2281
0.2315	0.2348	0.2352	0.2283	0.2304	0.2549	0.2732	0.2759	0.2786	0.2648	0.2286	0.2389	0.2542	0.2822	0.3102
0.3382	0.3662	0.3820	0.3672	0.3559	0.3532	0.3506	0.3535	0.3705	0.3874	0.4044	0.4190	0.4199	0.4097	0.4202
0.4307	0.4326	0.4133	0.3982	0.3940	0.3894	0.3842	0.3790	0.3790	0.3923	0.3930	0.3622	0.3314	0.6212	0.6258
0.6304	0.6273	0.6049	0.5896	0.5916	0.5937	0.6018	0.6250	0.6481	0.6452	0.5771	0.5326	0.5468	0.5545	0.5456
0.5401	0.5431	0.5373	0.5097	0.4809	0.4493	0.4271	0.4288	0.4338	0.4474	0.4477	0.4152	0.3989	0.4234	0.4237
0.3859	0.4049	0.4239	0.4395	0.4464	0.4497	0.4443	0.4344	0.4135	0.4107	0.4533	0.4730	0.4354	0.4113	0.4210
0.4280	0.4282	0.4284	0.4324	0.4460	0.4577	0.4645	0.4769	0.5031	0.5312	0.5643	0.5880	0.5880	0.5880	0.5880
0.5880	0.5880	0.5880	0.5880	0.5880	0.5880	0.5880	0.5880	0.5880	0.5880	0.5880	0.5880	0.5880	0.5880	0.5880
0.5880	0.5880	0.5880	0.5880	0.5880	0.5880	0.5880	0.5880	0.5880	0.5880	0.5880	0.5880	0.8089	0.8281	0.8472
0.8604	0.8584	0.8565	0.8646	0.8974	0.9131	0.8860	0.8708	0.8854	0.9000	0.9063	0.8921	0.8874	0.9070	0.9148
0.8934	0.8778	0.8765	0.8765	0.8793	0.8744	0.8499	0.8300	0.8216	0.8179	0.8259	0.8339	0.8419	0.8418	0.8213
0.8062	0.8046	0.8030	0.8168	0.8689	0.8985	0.8719	0.8480	0.8312	0.8144	0.8052	0.8149	0.8180	0.8045	0.7883
0.7653	0.7424	0.7406	0.7523	0.7768	0.7665	0.7017	0.7188	0.7372	0.7586	0.7811	0.8066	0.8224	0.8138	0.8052
0.7931	0.7793	0.7811	0.7766	0.7705	0.7605	0.7515	0.7448	0.7381	0.7400	0.7634	0.7867	0.8002	0.7889	0.7832
0.7912	0.7991	0.8051	0.8059	0.8068	0.8076	0.8085	0.8093	0.8102	0.8110	0.8119	0.8128	0.8112	0.8036	0.7990
0.8020	0.8050	0.8102	0.8209	0.8253	0.8139	0.8025	0.7966							

9.2.3. Popularidad en Quito, ordenadas de izquierda a derecha

0.4829	0.4929	0.5214	0.5964	0.6443	0.6243	0.6050	0.5875	0.5700	0.5525	0.5440	0.5580	0.5720	0.5700	0.5467
0.5424	0.5857	0.6290	0.6586	0.6536	0.6443	0.6243	0.6043	0.6000	0.6350	0.6543	0.6343	0.6143	0.6103	0.6463
0.6823	0.7183	0.7543	0.7737	0.7517	0.7297	0.7077	0.6857	0.6800	0.7150	0.7500	0.7850	0.7714	0.6364	0.5410
0.5443	0.5476	0.5471	0.5371	0.5271	0.5086	0.4686	0.4390	0.4357	0.4324	0.4457	0.5007	0.5386	0.5336	0.5229
0.4979	0.4729	0.4479	0.4271	0.4171	0.4024	0.3757	0.3491	0.3543	0.4393	0.4829	0.4229	0.3754	0.3592	0.3680
0.4100	0.4295	0.3929	0.3562	0.3443	0.3943	0.4443	0.4705	0.4371	0.4038	0.3871	0.4121	0.4264	0.4139	0.4014
0.3889	0.3771	0.3671	0.3571	0.3443	0.3243	0.3171	0.3421	0.3600	0.3600	0.3600	0.3686	0.3986	0.4286	0.4186
0.3086	0.2357	0.2557	0.2629	0.2379	0.2200	0.2200	0.2143	0.1943	0.1819	0.1886	0.1952	0.1971	0.1871	0.1810
0.1843	0.1876	0.1843	0.1643	0.1600	0.1950	0.2248	0.2414	0.2581	0.2529	0.2114	0.2164	0.2280	0.2560	0.2840
0.3120	0.3400	0.3543	0.3343	0.3152	0.2986	0.2819	0.2857	0.3407	0.3957	0.4507	0.4700	0.4057	0.3700	0.3700
0.3700	0.3757	0.3957	0.4043	0.3843	0.3700	0.3700	0.3700	0.3843	0.4343	0.4557	0.4057	0.3557	0.6238	0.6371
0.6505	0.6529	0.6279	0.6081	0.6014	0.5948	0.5976	0.6243	0.6510	0.6543	0.5993	0.5686	0.5986	0.6257	0.6457
0.6429	0.5829	0.5329	0.5079	0.4757	0.4257	0.3957	0.4157	0.4429	0.4879	0.4957	0.4107	0.3600	0.3950	0.4057
0.3814	0.4214	0.4614	0.4957	0.5157	0.5157	0.4657	0.4300	0.4300	0.4414	0.4814	0.5000	0.4650	0.4443	0.4593
0.4690	0.4657	0.4624	0.4571	0.4471	0.4457	0.4657	0.4957	0.5507	0.5957	0.6157	0.6300	0.6300	0.6300	0.6300
0.6300	0.6300	0.6300	0.6300	0.6300	0.6300	0.6300	0.6300	0.6300	0.6300	0.6300	0.6300	0.6300	0.6300	0.6300
0.6300	0.6300	0.6300	0.6300	0.6300	0.6300	0.6300	0.6300	0.6300	0.6300	0.6300	0.6300	0.8300	0.8337	0.8374
0.8410	0.8443	0.8476	0.8571	0.8821	0.8900	0.8550	0.8367	0.8600	0.8833	0.8914	0.8614	0.8486	0.8786	0.8871
0.8421	0.8171	0.8421	0.8586	0.8536	0.8400	0.8050	0.7829	0.7929	0.7993	0.7968	0.7943	0.7918	0.7871	0.7771
0.7690	0.7657	0.7624	0.7771	0.8371	0.8671	0.8221	0.7862	0.7729	0.7595	0.7543	0.7693	0.7671	0.7221	0.6811
0.6500	0.6189	0.6200	0.6400	0.6750	0.6571	0.5571	0.5821	0.6129	0.6579	0.6943	0.7093	0.7190	0.7157	0.7124
0.7100	0.7157	0.7257	0.7107	0.6971	0.6871	0.6781	0.6714	0.6648	0.6667	0.6900	0.7133	0.7343	0.7493	0.7543
0.7343	0.7143	0.7011	0.7051	0.7091	0.7131	0.7171	0.7211	0.7251	0.7291	0.7331	0.7371	0.7386	0.7336	0.7329
0.7429	0.7529	0.7586	0.7536	0.7452	0.7286	0.7119	0.7064							

9.2.4. Popularidad en Guayaquil, ordenadas de izquierda a derecha

0.3971	0.4221	0.4514	0.4914	0.5314	0.5714	0.5943	0.5743	0.5543	0.5343	0.5220	0.5290	0.5360	0.5400	0.5400
0.5400	0.5400	0.5400	0.5443	0.5593	0.5652	0.5486	0.5319	0.5343	0.5843	0.6162	0.6029	0.5895	0.5840	0.5980
0.6120	0.6260	0.6400	0.6557	0.6757	0.6957	0.7157	0.7357	0.7393	0.7018	0.6643	0.6268	0.6171	0.6771	0.7143
0.6943	0.6743	0.6610	0.6643	0.6676	0.6614	0.6314	0.6043	0.5843	0.5643	0.5500	0.5500	0.5429	0.5179	0.4950
0.4775	0.4600	0.4425	0.4329	0.4429	0.4481	0.4414	0.4348	0.4286	0.4236	0.4371	0.4971	0.5200	0.4500	0.4160
0.4300	0.4314	0.4014	0.3714	0.3652	0.4186	0.4719	0.4957	0.4457	0.3957	0.3614	0.3664	0.3707	0.3732	0.3757
0.3782	0.3819	0.3886	0.3952	0.3986	0.3936	0.3786	0.3386	0.3167	0.3400	0.3633	0.3886	0.4186	0.4486	0.4571
0.4121	0.3971	0.4571	0.4671	0.3521	0.2829	0.3279	0.3486	0.3086	0.2886	0.3186	0.3486	0.3557	0.3057	0.2710
0.2743	0.2776	0.2814	0.2864	0.2943	0.3093	0.3171	0.3071	0.2971	0.2757	0.2443	0.2593	0.2780	0.3060	0.3340
0.3620	0.3900	0.4071	0.3971	0.3929	0.4029	0.4129	0.4150	0.3975	0.3800	0.3625	0.3729	0.4329	0.4457	0.4657
0.4857	0.4843	0.4293	0.3929	0.4029	0.4071	0.3971	0.3871	0.3743	0.3543	0.3362	0.3229	0.3095	0.6190	0.6157
0.6124	0.6043	0.5843	0.5729	0.5829	0.5929	0.6057	0.6257	0.6457	0.6371	0.5571	0.5000	0.5000	0.4900	0.4550
0.4471	0.5071	0.5414	0.5114	0.4857	0.4707	0.4557	0.4407	0.4257	0.4107	0.4043	0.4193	0.4343	0.4493	0.4400
0.3900	0.3900	0.3900	0.3886	0.3836	0.3900	0.4250	0.4386	0.3986	0.3829	0.4279	0.4486	0.4086	0.3814	0.3864
0.3910	0.3943	0.3976	0.4100	0.4450	0.4686	0.4636	0.4600	0.4600	0.4729	0.5179	0.5500	0.5500	0.5500	0.5500
0.5500	0.5500	0.5500	0.5500	0.5500	0.5500	0.5500	0.5500	0.5500	0.5500	0.5500	0.5500	0.5500	0.5500	0.5500
0.5500	0.5500	0.5500	0.5500	0.5500	0.5500	0.5500	0.5500	0.5500	0.5500	0.5500	0.5500	0.7900	0.8232	0.8563
0.8714	0.8648	0.8714	0.9114	0.9343	0.9143	0.9019	0.9086	0.9152	0.9200	0.9200	0.9229	0.9329	0.9400	0.9400
0.9329	0.9079	0.8929	0.9029	0.9057	0.8907	0.8729	0.8479	0.8350	0.8525	0.8700	0.8875	0.8914	0.8614	0.8400
0.8400	0.8400	0.8529	0.8979	0.9271	0.9171	0.9043	0.8843	0.8643	0.8514	0.8564	0.8643	0.8793	0.8856	0.8700
0.8544	0.8500	0.8543	0.8693	0.8657	0.8329	0.8429	0.8500	0.8500	0.8600	0.8950	0.9162	0.9029	0.8895	0.8686
0.8371	0.8314	0.8364	0.8371	0.8271	0.8181	0.8114	0.8048	0.8067	0.8300	0.8533	0.8600	0.8250	0.8095	0.8429
0.8762	0.8994	0.8974	0.8954	0.8934	0.8914	0.8894	0.8874	0.8854	0.8834	0.8814	0.8771	0.8671	0.8590	0.8557
0.8524	0.8571	0.8821	0.8981	0.8914	0.8848	0.8786								

Referencias

- [1] Manuel Montes-y-Gómez, Alexander F. Gelbukh Aurelio López-López Text Mining as a Social Thermometer
- [2] K. M. Carley. Network text analysis: The network position of concepts. In C. W. Roberts, editor, Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts, pages 79–100. Lawrence Erlbaum Associates, 1997.
- [3] S. R. Corman, T. Kuhn, R. D. McPhee, and K. J. Dooley. Studying complex discursive systems: Centering resonance analysis of communication. *Human Communication Research*, 28(2):157–206, 2002
- [4] John Sowa Conceptual Structure Addison-Wesley Pub., Reading, MA 1983
- [5] A Comparative Evaluation of Collocation Extraction Techniques Darren Pearce
- [6] Y. Matsuo, M. Ishizuka Keyword extraction from a single document using word co-occurrence statistical informatiton
- [7] Ulrik Brandes A Faster Algorithm for Betweenness Centrality
- [8] Maria Begoña Villada Moirón Data-driven identification of fixed expressions and their modifiability
- [9] <http://www.gnu.org/software/wget/> (17-Nov-2008)
- [10] Sofía N. Galicia Haro y Alexander Gelbukh Investigaciones en Analisis sintactico para el español Instituto Politécnico Nacional México 2007
- [11] Asunción Honrado, Ruben Leon, Ruairi O'Donnell, Duncan Sinclair A Word Stemming Algorithm for the Spanish Language
- [12] Robert A. Hanneman and Mark Riddle; Introduction to social network methods <http://www.faculty.ucr.edu/~hanneman/nettext/index.html>
- [13] Freeman, L. C. A set of measures of centrality based on betweenness. *Sociometry* (1977)
- [14] Andreas Hotho, Andreas Nurnberger, Gerhard Paaß, Sankt Augustin. A Brief Survey of Text Mining. May 13, 2005
- [15] Wikipedia; http://en.wikipedia.org/wiki/Mutual_information (2.Dic-2008)
- [16] Carlos G. Figuerola, Raquel Gómez, Angel F. Zazo Rodríguez, José Luis Alonso Berrocal. Stemming in Spanish: A First Approach to its Impact on Information Retrieval.
- [17] <http://snowball.tartarus.org/algorithms/spanish/stemmer.html> (4-Ene-2009)

- [18] Ingo Feinerer, Wirtschaftsuniversität, Kurt Hornik, David Meyer. Text Mining Infrastructure in R. Journal of Statistical Software March 2008, Volume 25, Issue 5.
- [19] Landauer, T. K., Foltz, P. W., & Laham, D. Introduction to Latent Semantic Analysis. Discourse Processes, 25, 259-284 (1998)
- [20] http://en.wikipedia.org/wiki/Algoritmo_de_Dijkstra (10-Ene-2009)
- [21] http://en.wikipedia.org/wiki/Kalman_filter (15-Jun-2009)
- [22] Andrew C. Harvey and Neil Shephard. Structural Time Series Models. G. S. Maddala, C. R. Rao and H. D. Vinod, eds., Handbook of Statistics, Vol. 11
- [23] Jacques J. F. Commandeur, Siem Jan Koopman. An Introduction to State Space Time Series Analysis. Oxford University Press 2007
- [24] Ana Pérez Espartero. Modelo con memoria larga para la media proceso arfina. Dpto. Economía Aplicada (Estadística y Econometría) Universidad de Valladolid
- [25] Anna Mikusheva. Time Series Analysis. course materials for 14.384 Fall 2007 MIT OpenCourseWare (<http://ocw.mit.edu>), Massachusetts Institute of Technology
- [26] Gérard Weisbuch. Social Opinion Dynamics. Econophysics and Sociophysics Trends and Perspectives Edited by Bikas K. Chakrabarti, Anirban Chakraborti, and Arnab Chatterjee WILEY-VCH
- [27] Trevor Hastie, Robert Tibshirani, Jerome Friedman. The elements of Statistics Learning. Springer Verlag 2001
- [28] KW Hipel, AI McLeod, WC Lennox; Advances in Box-Jenkins Modeling 1. Model Construction ;Water Resources Research, 1977
- [29] James Douglas Hamilton ; Time Series Analysis ; Princeton University Press; 1 edition (January 11, 1994)
- [30] http://en.wikipedia.org/wiki/Akaike_information_criterion
- [31] http://en.wikipedia.org/wiki/Normality_test
- [32] Damodar Gujarati ; Econometría cuarta edición; Mc Graw Hill
- [33] A simple positive semi-definite heteroskedasticity and autocorrelation consistent WK Newey, KD West; Econometrica: Journal of the Econometric Society, 1987
- [34] Econometric Computing with HC and HAC Covariance Matrix Estimators Achim Zeileis ; Wirtschaftsuniversität Wien