

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE ESTUDIOS ESTADÍSTICOS



TESIS DOCTORAL

Medidas de polarización en redes sociales

(Polarization measures on online social networks)

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Juan Antonio Guevara Gil

Directores

Daniel Gómez González
José Manuel Robles Morales

Madrid

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE ESTUDIOS ESTADÍSTICOS



TESIS DOCTORAL

MEDIDAS DE POLARIZACIÓN EN REDES SOCIALES
(Polarization Measures on Online Social Networks)

MEMORIA PARA OPTAR AL GRADO DE DOCTOR
PRESENTADA POR

JUAN ANTONIO GUEVARA GIL

DIRECTORES

DANIEL GÓMEZ GONZÁLEZ
Y
JOSÉ MANUEL ROBLES MORALES

Medidas de Polarización en Redes Sociales

(Polarization Measures on Online Social Networks)

Memoria para optar al grado de doctor presentada por

JUAN ANTONIO GUEVARA GIL



Facultad de Estudios Estadísticos

UNIVERSIDAD COMPLUTENSE

Doctorado en Análisis de Datos (Data Science)

Tesis dirigida por:

DANIEL GÓMEZ GONZÁLEZ Y JOSÉ MANUEL ROBLES MORALES

MADRID, JULIO 2022

A mis padres, Juan Guevara y Lola Gil.

Agradecimientos

*No dejéis que vuestras cabezas se
vuelvan más grandes que vuestros
sombrosos.*

J.R.R. Tolkien

A todo aquel que se de por aludido. Así es como tenía pensado empezar, y acabar, los agradecimientos de este escrito. Somos quienes somos gracias al conjunto de facilidades, dificultades y experiencias que nos encontramos a lo largo de nuestra vida. Es gracias a todas las personas que nos quieren que conseguimos las fuerzas para seguir hacia delante. Pero también es gracias a aquellos que, alguna vez, nos pusieron las cosas difíciles que tenemos la oportunidad de mejorar, crecer y crear la mejor versión de nosotros mismos. Es por todos ellos que llegamos a ser quienes somos hoy en día. Es también por aquellas personas que no recordamos, y que nos empujaron a forjar la mejor versión de nosotros mismos, que evolucionamos. Es por ello que, si alguien, de una forma u otra, piensa que pudo haberme aportado algo en la vida, le doy las gracias.

Sin embargo, la frase “*a todo aquel que se de por aludido*” no es justa para aquellos que me han acompañado a lo largo de este intenso pero magnífico viaje, ofreciéndome la mejor versión de ellos mismos, para que pueda seguir hacia adelante. Por ello, les dedico unas líneas especiales.

Agradezco de todo corazón a mis directores de tesis, Daniel Gómez y José Manuel Robles, por acogerme con los brazos abiertos. No podría haber imaginado mejores mentores, quienes con toda su comprensión y entusiasmo, cálidamente me han guiado en el mundo de la investigación y la academia. Gracias por estar siempre ahí y confiar en mí. Si la vida comprende el factor suerte, esta jugó a mi favor al guiar mis pasos hasta vuestro encuentro. Verdaderamente me siento “*a hombros de gigantes*”. De nuevo, de todo corazón: gracias.

A mis padres, Lola y Juan, a quienes les dedico esta tesis. Con esfuerzos inimaginables, lucharon con toda la fuerza de su alma para ofrecerme su mejor

versión, proporcionándome todo lo que tengo en esta vida y permitiéndome llegar hasta donde me encuentro hoy en día. Son ellos quienes me enseñaron a vivir de forma honesta, mostrándome el valor de la justicia, la belleza que hay en el trabajo duro y que no existe nada fuera de tu alcance cuando ofreces tu mejor versión; cuando pones todo tu corazón en ello. Gracias, y mil veces gracias, por confiar en mí hasta el último momento. Esta tesis doctoral es fruto de vuestros esfuerzos como padres. Gracias, también, a mi hermana Cristina, que ha estado siempre guiándome, sirviéndome de apoyo y quien me ha arropado a lo largo de los años. A mi abuela, Catalina, que me ha envuelto de sabiduría y humildad con sus historias, con sus recuerdos. “*No desprecies las tradiciones que nos llegan de antaño; ocurre a menudo que los mayores guardan en la memoria cosas que los sabios de otros tiempos necesitaban saber*” (J.R.R. Tolkien, 1954). Y, cómo no, a María, que con una bondad y comprensión que no conocen límites, ha sido mi compañera inseparable a lo largo de todo este viaje. No habría llegado tan lejos sin vosotros.

En cuanto a la vida académica, no puedo sino empezar por el principio y avanzar a través del eje cronológico de mi vida. Es gracias a mi profesor del instituto, y amigo, Juan Ramón Barat, que me encuentro escribiendo estas líneas. Gracias por creer en mí y poner todo tu corazón en motivarme en tiempos difíciles. Es a ti a quien debo mi progreso académico. A Juan Ignacio Aragonés, que me acogió en mis primeras experiencias como investigador, enseñándome que el esfuerzo, y la excelencia, se luchan día a día. Gracias por tu valiosa guía, comprensión y sinceridad. También es gracias a la cálida acogida de Mirko Antino que hoy me encuentro escribiendo estas palabras. Gracias, a todos, por compartir vuestra sabiduría.

Gracias, Inma. Te has convertido en mi modelo a seguir. Gracias por representar todo aquello que se debe ser, guiarme en los momentos difíciles y servir de inspiración. Gracias, Ángela, por compartir tu esfuerzo, entusiasmo y dedicación.

Así mismo, sólo puedo sentirme profundamente agradecido por formar parte de un maravilloso grupo de investigación llamado "DATA SCIENCE AND SOFT COMPUTING FOR SOCIAL ANALYTICS AND DECISION AID", donde he encontrado personas maravillosas de las que no puedo hacer otra cosa sino aprender.

Gracias a la Facultad de Estudios Estadísticos y todo su personal por acogerme con los brazos abiertos. Realmente me siento como en casa. Gracias, Javier Castro, por compartir tu sabiduría y acogerme como a uno más.

Also, I would like to thanks to POMLAB (Public Opinion & Media Lab) and its director Mauro Barisone for hosting me during my research stay. Many thanks for showing me the beauty and the synergy that lies behind the

international collaboration.

A hombros de gigantes.

Isaac Newton

Abstract

Polarization measures in Social Networks.

This doctoral thesis focuses on the study of the measurement of polarization in social networks. Although there is a relative consensus on the conceptualization of polarization in the literature, this is not the case with its measurement. Throughout the works published in recent decades, there is a wide range of polarization measures that are shown to be in dissonance with each other and, on occasions, in contradiction with the theoretical conceptualization of polarization.

With all this, throughout this work an exhaustive study of the main polarization measures is carried out in order to know its nature and behavior in greater depth. Likewise, polarization scenarios are studied, analyzed and compared to know the main characteristics of the consequences of such polarization in social behavior. As a natural continuation of the above, a fuzzy bipolarization measure is proposed with the aim of overcoming some of the deficiencies found in the literature, presenting a more realistic and consistent polarization measure with the concept. Thus, the proposed measure is applied in the real case of political polarization in social networks during the COVID-19 pandemic.

In addition, other uses of this measure are explored, where by adding the polarization values proposed by our measure to the community detection problems, it is possible to improve these algorithms. In addition, the proposed measure, and its conceptualization, is transferred to the case of graphs in order to measure polarization, not only from an attitudinal perspective, but also from a structural one. Finally, a completely novel approach in the literature is proposed, measuring polarization as a dynamic process over time. To do this, a Markov Chain model is proposed that allows predicting the levels of polarization in the future given a certain attitudinal distribution of the population.

Resumen

Medidas de Polarización en Redes Sociales

Esta tesis doctoral se centra en el estudio de la medición de la polarización en redes sociales. Si bien se encuentra relativo consenso sobre la conceptualización de la polarización en la literatura, esto no ocurre así con su métrica. A lo largo de los trabajos publicados en las últimas décadas, se encuentra una variada oferta de medidas de polarización que se muestran en disonancia las unas con las otras y, en ocasiones, en contradicción con la conceptualización teórica de la polarización.

Con todo ello, a lo largo de este trabajo se realiza un estudio exhaustivo de las principales medidas de polarización con el fin de conocer en mayor profundidad su naturaleza y comportamiento. Así mismo, se estudian, analizan y se comparan escenarios de polarización para conocer las características principales de las consecuencias de dicha polarización en el comportamiento social en redes. Como continuación natural de lo anterior, se propone una medida de bipolarización difusa con el objetivo de suplir algunas de las deficiencias encontradas en la literatura, presentando un modelo de medición de la polarización más realista y coherente con el concepto. Así, la medida propuesta se aplica en el caso real de polarización política en las redes sociales durante la pandemia por la COVID-19.

Complementariamente, se exploran otros usos de la medida propuesta, donde añadiendo los valores de polarización propuestos por la medida a los problemas de detección de comunidades, se consigue mejorar estos algoritmos. Además, se traslada la medida propuesta, y su conceptualización, al caso de los grafos con el fin de medir la polarización, no sólo desde una perspectiva actitudinal, sino también estructural. Por último, se plantea una aproximación totalmente novedosa en la literatura, midiendo la polarización como un proceso dinámico en el tiempo. Para ello, se plantea un modelo de Cadenas de Markov que permite predecir los niveles de polarización en el futuro dada una determinada distribución actitudinal de la población.

Índice

Agradecimientos	VII
Introduction	1
1. Introducción	7
1.1. Objetivos	9
1.2. Contribuciones	10
2. Preliminares	13
2.1. El concepto de polarización.	13
2.1.1. El origen de la polarización.	16
2.2. Medidas de Polarización	24
2.2.1. Medidas de bipolarización tradicionales.	28
2.2.2. Medidas de polarización basadas en redes.	33
2.3. Teoría de grafos y detección de comunidades	36
2.3.1. Concepto de grafo y digrafo.	36
2.3.2. Algunas medidas globales sobre grafos.	40
2.3.3. Medidas sobre nodos: las medidas de centralidad.	42
2.3.4. Detección de comunidades en redes sociales.	43
2.4. Los conjuntos borrosos	45
2.5. Cadenas de Markov.	50
3. Comportamiento de las medidas de polarización	53
3.1. Comportamiento de las medidas de bi-polarización tradicionales.	53
3.1.1. Medidas de bi-polarización basadas en datos continuos.	54
3.1.2. Medidas de bi-polarización basadas en datos categóricos.	57
3.2. Comportamiento de las medidas de polarización basadas en grafos.	62
4. Identificando y Analizando la Polarización	67

4.1. La presencia de los <i>bots</i> políticos en Twitter durante la crisis del COVID-19 en España.	67
4.1.1. Redes sociales y <i>bots</i>	68
4.1.2. Metodología.	69
4.1.3. Resultados.	74
4.1.4. Conclusiones.	78
4.2. Identificación de la Polarización: Polarización e incivilidad en el debate digital español sobre los derechos de la mujer.	80
4.2.1. Metodología y Resultados.	83
4.2.2. Conclusiones.	93
5. La medición de la Polarización a través de los conjuntos borrosos	95
5.1. Una nueva propuesta: medición de la polarización a través de los conjuntos borrosos.	95
5.1.1. Una nueva propuesta de polarización.	97
5.1.2. Una nueva medida de polarización basada en los conjuntos borrosos desde una perspectiva unidimensional y bipolar.	98
5.1.3. Comparación entre medidas de polarización.	102
5.1.4. Conclusiones.	107
5.2. Aplicación en un caso real: Cuando la negatividad es el combustible. Bots y Polarización Política en el debate del COVID-19.	109
5.2.1. Metodología.	110
5.2.2. Análisis y resultados.	114
5.3. Otras aplicaciones: Medida de polarización JDJ y la mejora de algoritmos de detección de comunidades	121
5.3.1. Detección de comunidades a través de una nueva perspectiva.	122
5.3.2. Medida de bipolarización difusa JDJ en grafos.	123
5.3.3. Aplicación a un caso real.	131
6. Medidas de Polarización en grafos y redes	141
6.0.1. Construyendo las funciones de pertenencia a los polos.	143
6.0.2. Medida de Polarización difusa JDJ aplicada a grafos	145
6.0.3. Aplicación y resultados computacionales.	148
6.0.4. Conclusiones.	153
7. Una nueva propuesta: La medición de la Polarización desde	

una perspectiva dinámica	155
7.1. Un nuevo enfoque a la modelización de la Polarización usando Cadenas de Markov	156
7.1.1. Polarización como proceso dinámico.	156
7.1.2. Un nuevo problema: Modelización de la Polarización usando Cadenas de Markov.	157
7.1.3. Experimento y resultados.	160
7.1.4. Conclusiones.	164
Conclusions and Future Work	167
Conclusions	167
Relation between proposed tasks and contributions	169
Future research lines	173
Conclusiones y Futuras Investigaciones	175
Conclusiones	175
Relación entre objetivos propuestos y contribuciones	177
Future research lines	181
Bibliografía	183

Índice de figuras

2.1. Número de artículos de Polarización por áreas en Web of Science. Palabras clave: social polarization* OR political polarization* OR economic polarization* OR ethnic polarization* OR affective polarization*	14
2.2. Número de artículos de Polarización por año en Web of Science. Palabras clave: social polarization* OR political polarization* OR economic polarization* OR ethnic polarization* OR affective polarization*	21
2.3. Grafo simple.	37
2.4. Subgrafo de red de retweets con comunidades.	45
3.1. Funciones de densidad para conjuntos de datos continuos.	55
3.2. Conjuntos de datos continuos.	60
3.3. Comparación de escenarios B y G	61
3.4. Conjuntos grafos generados.	65
4.1. Mensajes asociados a estilos comunicacionales de los principales partidos políticos de España producidos por <i>bots</i>	75
4.2. Análisis de sentimiento de los mensajes producidos por <i>bots</i> agrupados por estilo de comunicación.	76
4.3. Palabras más frecuentes asociadas a sentimientos negativos por cada estilo comunicacional.	77
4.4. Red de usuarios no <i>bots</i>	78
4.5. Red de usuarios <i>bots</i>	79
4.6. Comunidades principales de la red de retweets de “La manada”.	86
4.7. Nube de palabras de mención total. A la izquierda, comunidad de color morado. A la derecha, comunidad de color verde.	87
4.8. Visualización de la red cuéntalo presentada por la profesora y tuitera Dña. Mariluz Congosto.	89
4.9. Grafo de relaciones de los 145 usuarios más influyentes.	92

5.1. Ejemplo de una distribución bipolarizada.	99
5.2. Función de pertenencia (μ_1^T en amarillo y μ_5^T en azul) a los polos $X_A = 1$ y $X_B = 5$	102
5.3. Histogramas para las puntuaciones bidimensionales provenientes de IOV , ER , JDJ_PRO y JDJ_MIN	104
5.4. Correlaciones entre IOV , ER , JDJ_PRO y JDJ_MIN	105
5.5. Relación entre IOV - en deciles - y las puntuaciones observadas de ER , JDJ_PRO y JDJ_MIN	106
5.6. Valores medios para cada medida de polarización (ER -negro-, JDJ_PRO -gris- and JDJ_MIN -blanco-) para diferentes distribuciones de X . (1) 90% de los individuos están agrupados en torno al valor 5. (2) 45% de los individuos están agrupados en torno al valor 5 y el 45% el valor 4. (3) 30% de los individuos están agrupados en torno al valor 5, el 30% en el valor 4 y el 30% en el valor 3. (4) 22.5% de los individuos están agrupados en torno a cada uno de los valores del 2 al 5. (5) 18% de los individuos están agrupados en torno cada uno de los valores de X . (6) 30% de los individuos están agrupados en torno al valor 2, 30% en el valor 3 y el 30% en el valor 4. (7) 15% de los individuos están agrupados en torno al valor 2, 60% en el valor 3 y el 15% en el valor 4. (8) 45% de los individuos están agrupados en torno al valor 2 y 45% en el valor 4. (9) 45% de los individuos están agrupados en torno al valor 1 y 45% en el valor 5.	106
5.7. Distancia a recorrer por una distribución para llegar a un escenario de polarización máxima.	108
5.8. Polarización.	116
5.9. <i>Wordcloud</i> del debate digital.	117
5.10. Polarización por temática y tipo de cuenta.	118
5.11. Proporción de sentimientos por tipo de cuenta y temática.	119
5.12. Grafo difuso extendido de no polarización. $\tilde{G} = (V, E, \mu_{P+})$	127
5.13. Grafo $G = (V, E)$	134
5.14. Partición obtenida por el algoritmo de Louvain del grafo $G = (V, E)$	135
5.15. Particiones obtenidas con el algoritmos de <i>Polarization Louvain</i> con el grafo borroso extendido de no polarización $\tilde{G} = (V, E, \mu_P)$. $\gamma = 0$; $\varphi = max$; $\phi = min$	137

5.16. Particiones obtenidas con el algoritmos de <i>Polarization Louvain</i> con el grafo borroso extendido de no polarización $\tilde{G} = (V, E, \mu_P)$. $\gamma = 0$; $\varphi = max$; $\phi = prod$	137
5.17. Valores de Polarización para la partición $P = \{C_1, \dots, C_s\}$ por operadores de <i>overlapping</i>	139
5.18. Nodos 38, 76, 115 y 203 agrupados por Louvain.	140
5.19. Nodos 38, 76, 115 y 203 agrupados por <i>Polarization Louvain algorithm</i>	140
6.1. En gris, nodos pertenecientes a la comunidad C_1 , en blanco, los referentes a C_2	144
6.2. Red de Kárate con dos comunidades y modularidad $Q = 0.3599$.149	
6.3. Diferentes Medidas de Polarización para la red de Kárate añadiendo y eliminando aristas entre comunidades.	151
6.4. Diferentes Medidas de Polarización para la red de Kárate añadiendo y eliminando aristas dentro de comunidades.	152
7.1. Pasos medios para alcanzar estados de polarización a través de la variación del parámetro α	163
7.2. Probabilidades medias - barras - y pasos - color - para alcanzar estados polarizados.	163
7.3. Valores de polarización y probabilidad de polarización para cada estado inicial.	164

Índice de Tablas

3.1. Resultados de las medidas de polarización de Foster y Wolfson (P_{FW}) y Esteban y Ray (P_{ER}) para cada uno de los conjuntos de datos generados.	56
3.2. Resultados de las medidas de polarización categóricas para cada uno de los conjuntos de datos generados.	61
3.3. Resultados de las medidas de polarización categóricas para los escenarios B y G	62
3.4. Resultados de las medidas de polarización basadas en redes.	64
4.1. Resultados del clasificador SVM-lineal para limpieza de datos.	72
4.2. Resultados del clasificador SVM para la variable Partido político.	73
5.1. Estadísticos descriptivos de IOV , ER , JDJ_Pro y JDJ_MIN	103
5.2. Resultados del clasificador SVM para limpieza de datos.	112
5.3. Resultados del clasificador SVM para la codificación <i>a favor</i> y <i>en contra</i> hacia la gestión del gobierno.	114
5.4. Pruebas de efectos inter-sujetos.	118
5.5. Grados de pertenencia de cada elemento de V hacia los polos X_A y X_B	126
5.6. Valores de las medidas difusas μ_{P-} y μ_{P+}	126
5.7. Resultados de las máquinas de soporte vectorial (SVM) para las variables “TOPIC” y “POSITION”.	132
5.8. Comparación entre las particiones obtenidas. $\varphi = max$ and $\phi = min$	138
5.9. Comparación entre las particiones obtenidas. $\varphi = max$ and $\phi = prod$	138
6.1. Algunas comparaciones por pares y su medición de polarización de la red de comunidades borrosas $(V, E, \widetilde{P}_{gre})$	146

- 6.2. Particiones obtenidas para la red de Karate proveniente del algoritmo *makefuzzy* [104]. 150
- 7.1. Matriz de transición para $\alpha = 0.8$ y $z = 3$ 160

Introduction

Resumen: This chapter makes a brief introduction to the problems found in the literature that supports this thesis, focused on the measurement of polarization. Likewise, the aims proposed by the present work are presented, as well as the scientific contributions published in high impact journals that have arisen as a consequence of the fulfillment of such tasks.

In recent years, thanks to the globalization provided by the Internet, as well as the digitization of reality favored by the appearance of social networks, social behavior has taken a leap into the digital space. In this context, and as a consequence of the rapid spread of information in digital networks, the Social Sciences have directed their research to the study and measurement of phenomena that take place on the Internet. Among these phenomena, it is worth highlighting polarization. Digital polarization occurs in the face of various issues, such as political, social, religious, etc. Given the characteristics of this phenomenon, its appearance has been related as a direct consequence of social conflict or wars, among others (Montalvo y Reynal-Querol, 2003) [76].

Specifically, polarization has represented one of the most cited *topics* in the media and scientific papers in recent years. However, the polarization measurement literature has collected over the years multiple polarization measurement proposals that, as will be seen throughout this thesis, a study on the metric properties that these polarization measurements should satisfy as well as those scenarios in which it can be measured is missing. Additionally, and when trying to mathematically formalize some of the measures that appear in the literature, the term polarization is often confused with dispersion, assuming that the lack of dispersion in a population implies the absence of polarization. This is due to the segmentation found in the different paradigms that have addressed the measurement of polarization. Thus, polarization measures have traditionally been designed from a single discipline, such as economics or sociology, without finding the necessary disciplinary

union for the correct metric design. This fact gives rise to a problem of great importance since if this confusion between polarization and dispersion takes place, two scenarios are presented. In the first place, the measurement of polarization has been approached through an erroneous perspective, giving rise to scientific conclusions that do not present the required and/or expected precision. On the other hand, the need arises to propose a measure of polarization that satisfies the metric and theoretical needs of the phenomenon.

On the other hand, the measurement of polarization, of a traditional nature, has not presented an adequate adaptation to the scenario of *big data*. In this way, the measurement of polarization in social networks has focused on measuring it through graphs, leaving aside the traditional measurement through an ideological axis. Strictly speaking, the measurement of polarization from a structural point of view is as necessary as the measurement of ideological or positional polarization. Although some measures of polarization in graphs intend to face this task, measuring the distribution of two opposite opinions along with the structure of a graph and detecting communities, conceptual deficiencies are found in the proposed measures.

Thus, the main objective of this thesis is to address the current problem of the lack of a metric proposal that addresses, simultaneously and cooperatively, the measurement of polarization from a metric and conceptual point of view. For this, an analysis of the main proposals for measuring polarization in the literature will be made to identify the shortcomings and virtues of all of them. Likewise, studies of identification and analysis of polarization in social networks will be carried out to analyze the discourse and the different agents that influence polarization. Subsequently, a new measure of polarization based on fuzzy sets will be proposed with the aim of reflecting the polarization in a population from a realistic point of view. This polarization measure, as it measures in broad terms, the homogeneity or heterogeneity of a population, will be applied in order to improve community detection algorithms in graphs in order to know the different applications of the designed measure. Likewise, it will be applied to real contexts of *big data* with the aim of testing its applicability and its correct performance. On the other hand, a generalization of the proposed measure to graphs is presented, taking the step to mediate polarization, not only from an ideological point of view but also from a structural one. Finally, since the polarization phenomenon is not static, but dynamic, a mathematical model based on Markov chains will be proposed in order to address the dynamism of polarization.

Tasks

Throughout the literature related to polarization and its measurement, emphasis has been placed on the different characteristics, underlying concepts, and nature of the concept of polarization without reaching a real consensus among authors. This fact raises the need to propose an approach to the concept and metrics of polarization from a multidisciplinary perspective.

Next, the general tasks that are addressed in this thesis are presented. In addition, the specific task will be presented for each of the general tasks.

Task 1: To study and collect the main characteristics and measures of polarization proposed in the literature in order to explore their elements, nature and create a taxonomy around them. Likewise, an analysis will be carried out to determine their performance in different polarization scenarios.

- **1.1.** To make a bibliographical search and compilation around the metric and characteristics of polarization.
- **1.2.** To create a taxonomy around the polarization measures proposed in the literature according to their characteristics, especially in relation to the type of data they need to be computed.
- **1.3.** To compute, and to formalize when it is necessary, the polarization measures collected for their application to different polarization scenarios and to observe their performance against certain population distributions.

Task 2: To analyze and identify polarization in the digital debate in order to detect the origins, consequences, and agents of polarization in networks.

- **2.1.** To study the role of bots in the digital debate as polarizing agents and generators of hate.
- **2.2.** To analyze the polarization around the digital debate in the cases of La Manada and #Cuéntalo.

Task 3: To design a measure of polarization that makes up for some of the shortcomings found in those already proposed in the literature and that addresses the key aspects of polarization detected in the previous objectives.

- **3.1.** To address the measurement of polarization through fuzzy sets in order to propose a new measure of polarization.

- **3.2.** To apply the polarization measure to a real case - the digital debate on networks about COVID-19 - in the social big data analysis context.
- **3.2.** To explore other applications of the proposed polarization measure in improving community detection algorithms in graphs.

Task 4: To adapt the proposed polarization measure to graphs.

- **4.1.** To propose a polarization measure based on fuzzy sets applicable to graphs.
- **4.2.** To compare the proposed polarization measure with the most used polarization measures in graphs.

Task 5: To explore and propose a model that makes it possible to address polarization from a dynamic perspective.

- **5.1.** to model the phenomenon of polarization from a dynamic perspective using Markov Chains.
- **5.2.** To run the model proposed above with a toy example to study its performance.
- **5.3.** To compare the results of the dynamic polarization model with different measures of polarization.

Contributions

The following is a chronological list of the scientific contributions generated as a reflection of the research and work carried out in the preparation of this thesis.

Contribution 1: Robles, J. M., Atienza, J., Gómez, D., & Guevara, J. A. (2019). La polarización de La Manada El debate público en España y los riesgos de la comunicación política digital. *Tempo Social*, 31(3), 193-216.

Contribution 2: Guevara, J.A., Gómez, D., Robles, J.M., Montero, J. (2020). Measuring Polarization: A Fuzzy Set Theoretical Approach. *Communications in Computer and Information Science*, vol 1238. Springer, Cham. https://doi.org/10.1007/978-3-030-50143-3_40

Contribution 3: Gutiérrez, I., Guevara, J. A., Gomez, D., Castro, J., & Espínola, R. (2021). Community Detection Problem Based on Polarization Measures: An Application to Twitter: The COVID-19 Case in Spain. *Mathematics*, 9(4), 443.

Contribution 4: Guevara, J. A., Atienza-Barthelemy, J., Gomez Gonzalez, D., & Robles, J. M. (2021). Polarization and incivility in digital debates on womens rights in Spain. Not just a matter of machismo. *Journal of Gender Studies*, 1-15.

Contribution 5: Robles, J. M., Guevara, J. A., Casas-Mas, B., & Gómez, D. (2022). When negativity is the fuel. Bots and political polarization in the COVID-19 debate. *Comunicar*, 30(71), 63-75.

Contribution 6: Martínez, A., Guevara, J.A., & Jiménez, A. (In press, 2023). La influencia de los bots políticos en Twitter durante la crisis del COVID-19 en España. *REIS: Revista Española de Investigaciones Sociológicas*.

Contribution 7: Simón de Blas, C., Guevara, J.A., Morillo, J., Gómez González, D. (2022). Polarization Measures in Bi-partition Networks Based on Fuzzy Graphs. *Communications in Computer and Information Science*, vol 1601. Springer, Cham.

Contribution 8: Guevara, J.A., Gómez, D., Castro, J., Gutiérrez, I., Robles, J.M. (2022). A New Approach to Polarization Modeling Using Markov Chains. *Communications in Computer and Information Science*, vol 1602. Springer, Cham.

Capítulo 1

Introducción

Resumen: En el presente capítulo se hace una breve introducción a la problemática encontrada en la literatura que sustenta esta tesis, centrada en la medición de la polarización. Así mismo, se presentan los objetivos propuestos por el presente trabajo, así como las contribuciones científicas publicadas en revistas de impacto que han surgido como consecuencia de la cumplimentación de tales objetivos.

En los últimos años, gracias a la globalización proporcionada por Internet, así como a la digitalización de la realidad favorecida por la aparición de las redes sociales, el comportamiento social ha dado un salto al espacio digital. En este contexto, y como consecuencia de la rápida propagación de la información en las redes digitales, las Ciencias Sociales han dirigido sus investigaciones al estudio y medición de fenómenos que tienen lugar en Internet. Entre estos fenómenos, cabe destacar el de la polarización. La polarización digital se da frente a diversas temáticas, como puede ser política, social, religiosa, etc. Dadas las características de este fenómeno, su aparición ha sido relacionada como consecuencia directa del conflicto social o guerras, entre otras (Montalvo y Reynal-Querol, 2003) [76].

En concreto, la polarización ha representado uno de los *topics* más citados en los medios y trabajos científicos en los últimos años. Sin embargo, la literatura de la medición de la polarización ha recogido a lo largo de los años múltiples propuestas de medición de polarización que, como se verá a lo largo de esta tesis, se echa en falta un estudio sobre las propiedades métricas que deberían satisfacer estas medidas de polarización así como aquellos escenarios en los que puede medirse. Adicionalmente y al tratar de formalizar matemáticamente algunas de las medidas que aparecen en la literatura, se confunde con frecuencia el término polarización con dispersión, asumiendo

que la falta de dispersión en una población implica ausencia de polarización. Esto se debe a la segmentación encontrada en los diferentes paradigmas que han abordado la medición de la polarización. Así, las medidas de polarización han sido diseñadas, tradicionalmente, desde una única disciplina, como puede ser la economía o la sociología, sin encontrar la unión disciplinar necesaria para el correcto diseño métrico. Este hecho da lugar a una problemática de gran importancia, pues de tener lugar esta confusión entre polarización y dispersión, se presentan dos escenarios. En primer lugar, la medición de la polarización se ha abordado a través de una perspectiva errónea, dando lugar a conclusiones científicas que no presentan la precisión requerida y/o esperada. Por otro lado, surge la necesidad de proponer una medida de polarización que satisfaga las necesidades métricas y teóricas del fenómeno.

Por otro lado, la medición de la polarización, de carácter tradicional, no ha presentado una adecuada adaptación al escenario del *big data* [94]. De esta forma, la medición de la polarización en redes sociales *online* se ha centrado en la medición de la misma a través de grafos, dejando de lado la medición tradicional a través de un eje ideológico. En términos estrictos, la medición de la polarización desde un punto de vista estructural es tan necesaria como la medición de la polarización ideológica o posicional. A pesar de que algunas medidas de polarización en grafos pretenden afrontar esta tarea, midiendo la distribución de dos opiniones opuestas a lo largo de la estructura de un grafo y la detección de comunidades, se encuentran carencias conceptuales en las medidas propuestas.

Así, el principal objetivo de esta tesis es el de abordar la problemática actual consistente en la falta de una propuesta métrica que incluya, de forma simultánea y cooperativa, la medición de la polarización desde un punto de vista métrico y conceptual. Para ello, se hará un análisis de las principales propuestas de medición de la polarización en la literatura para identificar las carencias y virtudes de todas ellas. Así mismo, se harán estudios de identificación y análisis de polarización en redes sociales para analizar el discurso y los diferentes agentes que influyen en la polarización. Posteriormente, se propondrá una nueva medida de polarización basada en los conjuntos borrosos con el objetivo de reflejar la polarización en una población desde un punto de vista realista. Esta medida de polarización, en tanto que mide en términos amplios, la homogeneidad o heterogeneidad de una población, se aplicará con el fin de mejorar algoritmos de detección de comunidades en grafos con el fin de conocer las diferentes aplicaciones de la medida diseñada. Así mismo, se aplicará a contextos reales *big data* con el objetivo de probar su aplicabilidad y el correcto rendimiento de la misma. Por otro lado, se presenta una generalización de la medida propuesta a grafos, dando el paso a mediar la

polarización, no sólo desde un punto de vista ideológico, sino estructural. Por último, en tanto que el fenómeno de polarización no es estático, sino dinámico, se propondrá un modelo matemático basado en Cadenas de Markov con el fin de abordar el dinamismo de la polarización.

1.1. Objetivos

A lo largo de la literatura relacionada con la polarización y su medición, se ha hecho hincapié sobre las diferentes características, conceptos subyacentes y naturaleza del concepto de polarización sin llegar a un consenso real entre autores. Este hecho despierta la necesidad de proponer una aproximación al concepto y métrica de la polarización desde una perspectiva multidisciplinar.

A continuación, se presentan los objetivos generales que se abordan en esta memoria de tesis. En adición, se presentarán los objetivos específicos que subyacen a cada uno de los objetivos generales.

Objetivo 1: Estudiar y recolectar las principales características y medidas de polarización propuestas en la literatura para conocer sus elementos, naturaleza y crear una taxonomía en torno a ellas. Así mismo, se hará un análisis para conocer su rendimiento ante diferentes escenarios de polarización.

- **1.1.** Hacer una búsqueda y recopilación bibliográfica en torno a las características métricas de polarización.
- **1.2.** Crear una taxonomía en torno a las medidas de polarización propuestas en la literatura de acuerdo con sus características, especialmente, en relación el tipo de datos que necesitan para ser computadas.
- **1.3.** Computar, y formalizar cuando sea necesario, las medidas de polarización recogidas para su aplicación a diferentes escenarios de polarización y observar el rendimiento de las mismas frente a determinadas distribuciones poblacionales.

Objetivo 2: Analizar e identificar la polarización en el debate digital con el fin de detectar los orígenes, consecuencias y agentes de la polarización en redes.

- **2.1.** Estudiar el papel de los *bots* en el debate digital como agentes polarizantes y generadores de odio en la red.
- **2.2.** Analizar la polarización en torno a el debate digital en los casos de La Manada y #Cuéntalo.

Objetivo 3: Diseñar una medida de polarización que supla algunas de las carencias encontradas en las ya propuestas en la literatura y que aborde los aspectos clave de la polarización detectados en los objetivos anteriores.

- **3.1.** Abordar la medición de la polarización a través de los conjuntos borrosos con el fin de proponer una nueva medida de polarización.
- **3.2.** Aplicar la medida de polarización a un caso real - el debate digital en redes sobre el COVID-19 - en entornos del *social big data analysis*.
- **3.3.** Explorar otras aplicaciones de la medida de polarización propuesta en la mejora de algoritmos de detección de comunidades en grafos.

Objetivo 4: Adaptar la medida de polarización propuesta al caso de grafos.

- **4.1.** Proponer una medida de polarización basada en los conjuntos borrosos aplicable a grafos.
- **4.2.** Comparar la medida de polarización propuesta con las medidas de polarización más usadas en grafos.

Objetivo 5: Explorar y proponer un modelo que permita abordar la polarización desde una perspectiva dinámica.

- **5.1.** Modelar el fenómeno de polarización desde una perspectiva dinámica a través de Cadenas de Markov.
- **5.2.** Ejecutar el modelo propuesto anteriormente con un *toy example* para estudiar su rendimiento.
- **5.3.** Comparar los resultados del modelo de polarización dinámico con medidas de polarización.

1.2. Contribuciones

A continuación se enumeran, de forma cronológica, las contribuciones científicas generadas como reflejo de la investigación y trabajo desarrollado en la elaboración de la presente tesis doctoral.

Contribución 1: Robles, J. M., Atienza, J., Gómez, D., & Guevara, J. A. (2019). La polarización de La Manada El debate público en España y los

riesgos de la comunicación política digital. *Tempo Social*, 31(3), 193-216.

Contribución 2: Guevara, J.A., Gómez, D., Robles, J.M., Montero, J. (2020). Measuring Polarization: A Fuzzy Set Theoretical Approach. *Communications in Computer and Information Science*, vol 1238. Springer, Cham. https://doi.org/10.1007/978-3-030-50143-3_40

Contribución 3: Gutiérrez, I., Guevara, J. A., Gomez, D., Castro, J., & Espínola, R. (2021). Community Detection Problem Based on Polarization Measures: An Application to Twitter: The COVID-19 Case in Spain. *Mathematics*, 9(4), 443.

Contribución 4: Guevara, J. A., Atienza-Barthelemy, J., Gomez Gonzalez, D., & Robles, J. M. (2021). Polarization and incivility in digital debates on womens rights in Spain. Not just a matter of machismo. *Journal of Gender Studies*, 1-15.

Contribución 5: Robles, J. M., Guevara, J. A., Casas-Mas, B., & Gómez, D. (2022). When negativity is the fuel. Bots and political polarization in the COVID-19 debate. *Comunicar*, 30(71), 63-75.

Contribución 6: Martínez, A., Guevara, J.A., & Jiménez, A. (Aceptado, Enero 2023). La influencia de los bots políticos en Twitter durante la crisis del COVID-19 en España. *REIS: Revista Española de Investigaciones Sociológicas*.

Contribución 7: Simón de Blas, C., Guevara, J.A., Morillo, J., Gómez González, D. (2022). Polarization Measures in Bi-partition Networks Based on Fuzzy Graphs. *Communications in Computer and Information Science*, vol 1601. Springer, Cham.

Contribución 8: Guevara, J.A., Gómez, D., Castro, J., Gutiérrez, I., Robles, J.M. (2022). A New Approach to Polarization Modeling Using Markov Chains. *Communications in Computer and Information Science*, vol 1602. Springer, Cham.

Capítulo 2

Preliminares

Resumen: En este capítulo de carácter introductorio se presentan los conceptos teóricos bajo los cuales se sustenta la investigación llevada a cabo en la presente tesis. En primer lugar, se hace una introducción al concepto teórico de la polarización así como a la medición de la misma, presentando las diferentes conceptualizaciones métricas propuestas en la literatura. Seguidamente, se hace una introducción a la teoría de grafos así como a los problemas de detección de comunidades, aspectos que en los últimos años se han convertido en un aspecto central del estudio de la polarización. También, se hace una introducción a los conjuntos borrosos, siendo una disciplina que servirá de base para la creación de nuevas propuestas métricas de la polarización. Finalmente, se introducen los conceptos básicos de Cadenas de Markov, modelos que servirán de utilidad en la tesis.

2.1. El concepto de polarización.

El concepto y estudio de la polarización en las Ciencias Sociales ha sido estudiado a lo largo del último siglo en diversas disciplinas como la sociología, economía, ciencias de la computación o comunicación, entre otras - ver Figura 2.1 -. Sin embargo, ha sido en los últimos veinte años que el estudio de la polarización ha presentado un creciente interés en el ámbito científico, encontrando un aumento exponencial en el número de artículos publicados en revistas de impacto tal y como se puede observar en la Figura 2.2. De estos datos se deduce la consolidación progresiva del estudio de la polarización en las Ciencias Sociales, posicionándolo actualmente como un tema de gran interés y repercusión social.



Figura 2.1: Número de artículos de Polarización por áreas en Web of Science. Palabras clave: social polarization* OR political polarization* OR economic polarization* OR ethnic polarization* OR affective polarization*

Sin embargo, a pesar de que el fenómeno de polarización ha sido estudiado a lo largo de los años, el entendimiento de dicho concepto ha variado de acuerdo con la disciplina que lo estudia, encontrando diferencias en la concepción de la polarización en la literatura. Una de las primeras referencias a polarización grupal fue la propuesta por Ross en 1921, donde referente a la cohesión grupal explica que “a veces el espíritu de facción gana misteriosamente la ventaja y la sociedad se polariza en grupos opuestos que pueden no tener más base que los partidos de Verdes y Azules [...] En general, sin embargo, es el antagonismo de la economía, intereses o ideas que dividen a la sociedad en dos” (Ross, 1921, p. 415)[99]. Por otro lado, Deutsch (1971) [23] hacía referencia a la polarización como el proceso en el que, una vez divididas las clases - o los grupos -, éstas se vuelven más homogéneas internamente y más diferenciados entre sí, encontrando en la definición de Deutsch alusiones a las características de los grupos más allá de la propia división. En esta dirección, Esteban y Ray (1994) [28], siendo unos de los primeros autores en proponer una medida de polarización formal en la literatura, proponen la conceptualización de la polarización como división de la sociedad en pocos grupos de tamaño significativo con alto grado de homogeneidad intragrupal y alto grado de heterogeneidad intergrupala. Es también en el marco de la medición de la polarización que se encuentra otra de las definiciones de polarización más tradicionales, propuesta por Wolfson en 1994 [120]. Desde el campo de la economía, Wolfson entiende la polarización de la población

en torno a la renta como la desaparición de la clase media y la agrupación de la población en torno a los extremos persiguiendo, de esta forma, una distribución bipolar.

Merece la pena señalar que la mayoría de las definiciones de polarización en la literatura son propuestas por aquellos trabajos que proponen una métrica para dicho fenómeno. En este sentido, se encuentra la falta de una conceptualización consensuada que incluya ambos aspectos: teóricos y métricos. En este sentido, tradicionalmente se encuentra una “polarización” en la propia conceptualización de la polarización. Así, de acuerdo con Bauer (2019) [6] cabe distinguir entre aquellos autores que entienden la polarización en términos de distanciamiento y aquellos que la entienden como separación o segregación. De hecho, cabe observar en la literatura que la conceptualización de la polarización está estrechamente ligada a la medición de la misma, donde determinados autores optan por un entendimiento unidimensional de la polarización, mientras que otros la entienden desde un punto de vista multidimensional. Por un lado, un entendimiento unidimensional implica la medición y entendimiento de la polarización respecto a un único eje, como por ejemplo la posición política de izquierdas a derechas. En contraste, la conceptualización multidimensional implica la consideración de más de un eje o aspecto, como es el caso de la polarización étnica, donde se considera, no sólo la pertenencia étnica de los participantes a determinados grupos, sino que también la renta de los mismos (Bauer, 2019) [6].

Cabe señalar, además, que la conceptualización de la polarización presenta otra división en la literatura en torno a la naturaleza temporal del fenómeno. Desde este punto de vista, algunos autores entienden la polarización como un proceso estático, o una fotografía de un determinado momento, mientras que para otros supone un proceso dinámico en el tiempo (DiMaggio, Evans y Bryson, 1996) [25]. Las connotaciones, por supuesto, suponen diferencias significativas. Entender la polarización como un proceso dinámico implica la consideración de la ausencia de una división en la población en un momento previo a la aparición de la polarización, por lo que se sobreentiende que, a través de la aparición de un determinado fenómeno social, político, etc., se ha producido una ruptura en la sociedad. Por otro lado, aquellas posiciones que entienden la polarización de forma estática únicamente pueden delimitarse a considerar la actual “fotografía” de la polarización, no pudiendo contemplar distribuciones anteriores o futuras.

Así, en esta sección se propone una definición formal de bipolarización:

Definición 2.1.1 (Polarización). Se entiende como polarización en las Ciencias Sociales a la división de la sociedad en dos grupos antagónicos y radica-

lizados, de tamaño significativo y similar, donde se rompe la comunicación pacífica entre los mismos.

2.1.1. El origen de la polarización.

Llegados al punto en el que se ha establecido una definición de la polarización en las Ciencias Sociales, el siguiente paso natural es el de conocer los inicios, detonantes y orígenes de la polarización. Si bien los inicios de este fenómeno pueden tener diferentes bases en función al tipo de polarización que se está estudiando - e.g.: política, social, afectiva, etc. - en las siguientes líneas se explicarán los procesos subyacentes a la identidad individual, colectiva y social para que dicho fenómeno tenga lugar, en términos generales. De acuerdo con ellos, estos orígenes podrán adaptarse a los diferentes detonantes específicos de cada tipo de polarización.

2.1.1.1. La naturaleza del individuo.

En primer lugar, para comprender de manera adecuada la forma en la que los seres humanos se comportan en sociedad es necesario comenzar por el individuo. Cómo éste se entiende y percibe así mismo para que, posteriormente, pueda sentirse parte de un conjunto y conocer su posición en la sociedad. Es posible identificar tres fenómenos básicos del yo (Baumeister, 2010)[7]: Conciencia reflexiva, ser interpersonal y función ejecutiva. A continuación, se explicarán brevemente los dos primeros en virtud de comprender los orígenes de la división grupal.

Una de las herramientas más características del ser humano es su capacidad para ser percibido por sí mismo, denominada conciencia reflexiva. La teoría de la autoconciencia objetiva (Duval y Wicklund, 1972)[26], mantiene que la conciencia es binaria en el sentido de que puede dirigirse tanto hacia fuera del individuo como para dentro, implicando así la comparación con cierta norma. Gracias a esto, se construye el autoconcepto. Este concepto subyace a la percepción del “yo” y, por lo tanto, a nuestro papel como individuos en una sociedad.

Una vez que el ser humano toma conciencia de sí mismo en la sociedad, como parte de algo más grande, se desarrolla una identidad más allá de la individual, denominada y propuesta por Tajfel como Identidad social (Tajfel, 1974) [113]. La principal idea de la Teoría de la Identidad social apunta a que la identidad del individuo no es estática, sino que cambia en función al entorno social que le rodea. Cuando una persona se encuentra partícipe de una masa, cambia su forma de verse y, en consecuencia, también su forma de actuar. La idea principal es, por lo tanto, que la identidad no es algo fijo

e inamovible, sino que cambia en función a nuestra situación. Así pues, las personas no se perciben así mismas como individuos aislados en entornos sociales, sino como parte de un grupo o determinadas categorías sociales. La conducta, por lo tanto, se encuentra bajo las influencias del grupo. La Teoría de la Identidad social defiende que la filiación a diferentes grupos ocurre gracias a un proceso psicológico llamado categorización. La categorización es el proceso que subyace a la creación y formación de los prejuicios y los estereotipos. El requisito mínimo para que una persona se sienta parte de un grupo es que, a través de una serie de comparaciones, la persona se perciba similar a los miembros de una categoría y se incluya dentro de los mismos. La teoría de la categorización del yo (Turner, 1983)[114] defiende que la identidad social está compuesta por el conjunto de categorías sociales a las que pertenece el yo. La categorización del yo implica consecuencias sobre la percepción de uno mismo y de los miembros del endogrupo, así como los del exogrupo.

Categorización social y el grupo interno.

- Percepción de los miembros del endogrupo: Cuando se analizan personas que pertenecen al endogrupo - mismo grupo social -, se tiende a resaltar las similitudes que hay entre dichos individuos, adoptando una actitud amistosa y abierta de carácter positivo.
- Evaluación de los miembros del endogrupo: Dado que aquellas personas que forman parte del endogrupo comparten los mismos atributos, tiene lugar el yo y lo mío, por lo que se tiende a valorar de una forma más positiva a dichos individuos de lo que podría hacerse con un miembro del exogrupo. En estos casos, por ejemplo, el agrado que producen propicia conceder el beneficio de la duda ante ciertos tipos de comportamientos o juicios, cosa que no ocurre con los miembros de otros grupos.

Categorización social y el grupo externo.

- Percepción del exogrupo: Existe una tendencia llamada homogeneización del exogrupo a través de la cual se tiende a percibir a los “otros” como iguales entre sí, generalizando así las características que poseen, ya sea para bien o para mal.
- Transformación de los otros en “ellos”: Si se cree firmemente que todos los que pertenecen al exogrupo son iguales, las acciones de un único individuo generalmente bastan para poder juzgar las del resto de individuos del exogrupo.

- Efecto de la mera clasificación: Dado que la existencia de una motivación que propicia el mantenimiento de una identidad social positiva, las valoraciones del exogrupo serán vistas de una forma menos favorable, en tanto que, como consecuencia, coloca al endogrupo con una mejor valoración.
- Efectos de una amenaza extrema: En aquellos casos en los que el juicio previo pasa del desagrado al odio extremo, se refleja la percepción de que el exogrupo es, o representa, una amenaza a lo que el endogrupo representa.

De acuerdo con lo anterior, el ser humano posee una naturaleza social que facilita la valoración positiva del endogrupo y rechaza con mayor facilidad al exogrupo. Esta combinación de factores suponen el caldo de cultivo para un comportamiento social proclive a la división en sí mismo. Todo lo contado anteriormente puede resumirse en los estudios realizados en el campamento de “la cueva de los ladrones”, realizado por Muzafer Sherif, estableciendo la base para el desarrollo de la teoría del conflicto realista (Sherif, 1954)[106].

Estos estudios, que duraron tres semanas y fueron diseñados para mostrar cambios sistemáticos en la conducta como resultado de cambios en las relaciones inter-grupales, fueron divididos en tres fases: formación del grupo, conflicto intergrupal y reducción de conflicto.

Formación del grupo: Los niños fueron divididos en dos grupos experimentales: los águilas y los serpientes. Acamparon, inicialmente, a cierta distancia entre sí, sin el conocimiento de la existencia mutua. Tras realizar actividades, por separado, durante algunos días, los propios grupos comenzaban a desarrollar una estructura interna y mini culturas con sus propios símbolos, nombres y normas específicas. En el momento en el que un grupo fue informado de la presencia del otro, diversos niños sugirieron espontáneamente que retaban al otro grupo a alguna competición deportiva. Es significativo que estas expresiones de favoritismo grupal se dieran antes de introducir la fase de conflicto inter-grupal.

Competición inter-grupal: Se anunció a los grupos que se iban a iniciar ciertas competiciones inter-grupales, introduciendo así un conflicto de intereses entre ambos grupos. En otras palabras, los grupos pasaron de ser mutuamente independientes a negativamente interdependientes: lo que ganaba uno, lo perdía el otro. Esto, obviamente, produjo cambios en las conductas de los niños, que cambiaron drásticamente, convirtiéndose ahora en dos facciones hostiles. De hecho, nunca perdían la oportunidad de ridiculizar al exogrupo

y, a veces, con agresiones físicas. Dentro de los grupos también tuvieron a cabo lugar algunos cambios. Estos se volvieron mucho más cohesivos y la estructura de liderazgo cambió a veces, asumiendo el dominio del grupo el chico más agresivo.

Reducción del conflicto: Tras haber generado tan fácilmente una competición tan encarnizada, se intentó reducir el conflicto a través de una serie de actividades con metas supra-ordenadas, es decir, metas deseadas por ambos, pero imposible de conseguir las por un grupo en solitario. Después de varias actividades como éstas, se observó un cambio drástico en la conducta de los participantes. El nivel de agresividad hacia los miembros del exogrupo disminuyó notablemente y se mostró una clara reducción de la cantidad de favoritismo inter-grupal.

A través de este experimento, Sherif concluyó que el prejuicio y la discriminación surgen de la competición por recursos - prestigio, dinero, bienes, tierras, estatus o poder - que los dos grupos desean. En la medida en la que los miembros de un grupo perciben que el otro representa una amenaza para la consecución de estas metas, la deterioración de las relaciones comienza. La teoría del conflicto realista sostiene que la hostilidad se dirige hacia la fuente de la amenaza.

2.1.1.2. De lo psicológico a lo social.

Sin embargo, es importante señalar que lo establecido anteriormente puede entenderse como el origen de los procesos individuales y sociales que permiten el origen de la polarización. Estos fenómenos, teorías y procesos se sitúan dentro del marco de la psicología social, donde los procesos están centrados en el individuo dentro de un contexto social. Para poder comprender los orígenes de la polarización, es necesario moverse al campo de la sociología la cual, de forma complementaria y conjunta, recoge los fenómenos que, de acuerdo con los procesos presentados de la psicología social, originan la polarización.

Así, los procesos sociales explicados anteriormente dentro del marco de la psicología social, presentan sus homólogos en el campo de la sociología donde el foco de atención es la sociedad en sí misma, y no el individuo. De esta forma, cuando se ha hablado anteriormente de los procesos de categorización, autoconcepto e identidad social como iniciadores de la formación grupal entre individuos, las consecuencias a nivel sociológico concluyen en la propuesta del concepto de *homofilia*, propuesta formalmente por Lazarsfel y

Merton (1954) [68]. De acuerdo con el concepto de homofilia, éste proceso sociológico se refiere al hecho de que el ser humano presenta un sesgo a generar vínculos con aquellos individuos que presentan mayor similitudes consigo mismo de acuerdo con sus características sociales. Desde este punto de vista, el concepto de homofilia se posiciona como uno de los pilares fundamentales en la formación y constitución de los grupos sociales. Así pues, es a través de este proceso que los seres humanos se relacionan de acuerdo a determinadas características en común con sus similares, encontrando un mayor afecto en aquello que es similar para el individuo. Como consecuencia de este proceso, es posible hablar de la naturalidad de la formación de grupos homogéneos en la sociedad.

De acuerdo con Lazarsfeld y Merton [68], cabe categorizar en dos subconceptos diferentes la homofilia, siendo la homofilia de estatus y la homofilia de valores. En primer lugar, cabe denominar la homofilia de estatus como la tendencia a la creación de relaciones entre pares semejantes de acuerdo con atributos de naturaleza personal, como lo puede ser la edad, el color de piel, trabajo, lugar de residencia, etc. Por otro lado, la homofilia de valores, también conocida como *homogeneidad*, se refiere al establecimiento de relaciones entre individuos de acuerdo a una similitud basada en criterios de nivel cognitivo. De acuerdo con la homogeneidad, la personas se agrupan entre sí de acuerdo con la semejanza existente entre aspectos subjetivos, como lo pueden ser las normas, interpretaciones, valores, etc. Es cuando la homogeneidad - homofilia de valores - juega un papel importante en la formación de grupos sociales con alto grado homogeneidad intragrupal y alto grado de heterogeneidad intergrupala que la polarización puede tener lugar. Cuando la sociedad se agrupa en torno a una determinada valoración o interpretación personal de un hecho político o social, se forman grupos sociales de acuerdo a éstas mismas valoraciones o interpretaciones, provocando una fragmentación social en torno a los polos opuestos de dicha valoración, caracterizada por el distanciamiento social de los grupos y la interrupción de la comunicación.

A parte de estas dos principales clases de homofilia, pueden distinguirse igualmente la homofilia de base y la homofilia endogámica [116]. La primera de ellas se refiere al establecimiento de relaciones grupales entre individuos como consecuencia de la composición demográfica de la sociedad, es decir “la proporción de las relaciones entre individuos de una característica dada en un colectivo es igual a la proporción de individuos con tal característica en el colectivo” (Verd et al., 2014, p.9) [116]. Así mismo, se habla de homofilia endogámica cuando la proporción de relaciones presentes de acuerdo con una característica social o personal específica en un determinado colectivo es mayor a la proporción presente en la población a la que pertenece dicho

colectivo. Cuando se da éste último caso, puede concluirse que la formación de grupos a través de la presencia endogámica de acuerdo con una determinada característica supera la aleatoriedad. En este sentido, mientras que la homofilia de base puede subyacer a la polarización entre diferentes poblaciones, en tanto que son diferentes por naturaleza, es la homofilia endogámica la que puede subyacer a la polarización entre dos grupos sociales dentro de una misma población.

Finalmente, cabe mencionar que, gracias a la introducción del internet en las vidas diarias de las personas, la web 2.0 y las redes sociales, el comportamiento de la sociedad ha encontrado un nuevo espacio donde tener lugar. De esta forma, gracias a las facilidades que presenta la comunicación online, no solo nuevos fenómenos sociales de carácter digital han tenido lugar, sino que la velocidad a la que evolucionan fenómenos sociales tradicionales se ha visto impulsada hasta niveles insospechados. Este conjunto de características, posiciona a las redes sociales online como el combustible principal de los orígenes y desarrollo de la polarización. Este escenario coloca a la polarización como uno de los fenómenos sociales emergentes de las últimas décadas, cuyo estudio ha presentado una evolución exponencial respecto a los últimos veinticinco años (ver Figura 2.2).

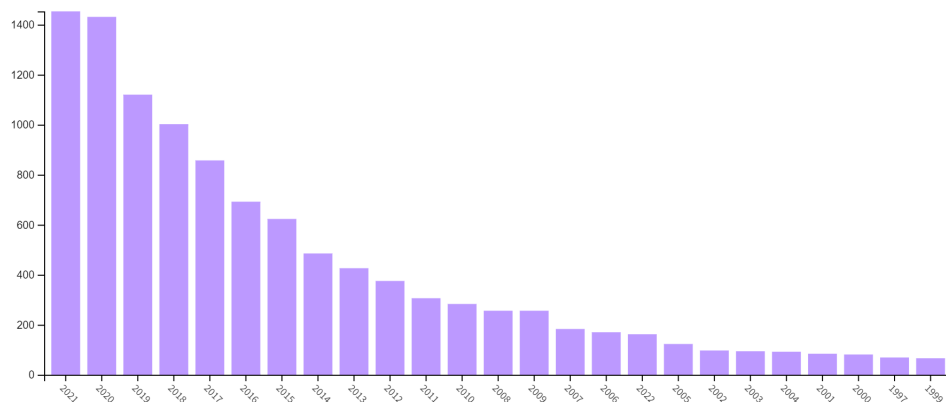


Figura 2.2: Número de artículos de Polarización por año en Web of Science. Palabras clave: social polarization* OR political polarization* OR economic polarization* OR ethnic polarization* OR affective polarization*

Como se ha visto, la polarización en las redes sociales es un fenómeno que está recogiendo gran interés en los últimos años. Las consecuencias de éste fenómeno pueden suponer un gran riesgo para el uso democrático y delibera-

tivo de las redes y del Internet. El distanciamiento y conflicto que conlleva la polarización entre dos grupos puede llevar a la ruptura de la sociedad que, a través del *flaming* o la *incivility*, se asienta el odio entre ambas partes, lo cual presenta el riesgo del establecimiento de un escenario difícilmente reversible. En concreto, la aparición de la polarización es especialmente sensible en las redes sociales online como consecuencia de las facilidades comunicativas que ofrecen tales plataformas. Este hecho, sumado a las facilidades que acompañan a la descarga de datos en internet, ha convertido a la polarización en uno de los *hot topics* de los últimos años.

Uno de los orígenes de la polarización en redes sociales es la exposición selectiva de los usuarios a determinada información. Como consecuencia, tal y como aporta Prior (2013) [91], este hecho promueve el desarrollo de actitudes y valores extremos, ya que los individuos presentan una tendencia al consumo de aquella información que previamente concuerda con sus valores. Así, el refuerzo de tales posiciones extremas es favorecido por el uso de redes sociales, lo que implicaría la estrecha relación entre la aparición de la polarización en estos espacios digitales (Sunstein, 2001[110]; 2018[111]). Como consecuencia, los principales actores de las redes cobran una gran importancia en la esfera pública digital, convirtiéndose en agentes de gran influencia sobre el debate y actitudes digitales. Estos usuarios presentarían el rol de líderes de movimiento, que mostrando posiciones extremas, se produciría un efecto de propagación a lo largo de todos sus seguidores propiciando la aparición de polarización.

Sin embargo, es de vital importancia resaltar que, según Sartori (2005) [102], la existencia de una polarización social no tiene por qué tener repercusiones negativas para la sociedad en la que se da tal polarización. Según el autor, es imprescindible diferenciar entre dos diferentes tipos de polarización. En primer lugar, una polarización negativa, o centrífuga, caracterizada por una posición de alejamiento respecto a ambos polos del eje. En este sentido, se entiende como polarización centrífuga aquella división de la sociedad donde la naturaleza de su interacción tiene un carácter de distanciamiento y diferenciación por parte de las dos comunidades formadas. Por otro lado, Sartori también propone la existencia de una polarización en torno partidos, caracterizada por una interacción entre comunidades de índole colaborativa, donde ambos polos procuran la existencia de un diálogo entre ambas partes, así como un entendimiento, donde no se busca la diferenciación entre los grupos, sino la colaboración. Este último tipo de polarización es también conocido como polarización centrípeta.

Sin embargo, si bien la existencia de polarización política no tiene por qué suponer repercusiones negativas para la interacción, sí que supone una

amenaza respecto a lo que la comunicación política se refiere, ya que el posicionamiento de dos comunidades confrontadas en torno a problemáticas potencialmente conflictivas puede suponer la ausencia de debates (Abramowitz, 2010) [1] que busquen una polarización centrípeta, generando bandos con poca disposición al entendimiento de las partes. Además, de acuerdo con Boxell, et al., (2017) [14], la pertenencia a determinados grupos sociales, así como los sesgos que suponen el consumo de determinada información política, determinan en gran medida los niveles de polarización observados en la red. Como consecuencia de ello, se puede apreciar un fenómeno de contagio donde los usuarios que toman parte en los debates reflejan posturas fragmentadas al transmitir aquella información política sesgada que han estado consumiendo. Entre las consecuencias negativas de este efecto de contagio, se encuentra lo que Lelkes (2016) [69] denominó como polarización afectiva, para referirse al proceso a través del cual los ciudadanos pertenecientes a los diferentes polos del eje radicalizan sus emociones en función a la información sesgada y polarizada producida por los partidos políticos simpatizantes, así como por sus representantes. De acuerdo con Haidt y Hethrington (2012) [55], esta disposición precipitaría la hostilidad entre las comunidades enfrentadas.

Como consecuencia de los efectos negativos de la polarización política, se introduce otro fenómeno conocido como incivildad política - *political incivility* -, caracterizado por un lenguaje inadecuado, incluso ofensivo, en el ámbito de la política. Desde esta perspectiva, cabe señalar que la comunidad científica difiere en cuanto a lo que *incivility* se refiere. Si bien los autores apuntan en una misma dirección, se pueden observar diferencias en cuando a la consideración del fenómeno. En este sentido, según Herbst (2010) [57], se considera *incivility* al uso de expresiones vulgares o irónicas, donde algunos de los interlocutores se muestran de forma poco educada.

Finalmente, cabe señalar que la incivildad supone un proceso de retroalimentación, donde ambientes en los que se observan mayores conductas incívicas se observa un mayor tipo de mensajes con contenido inadecuado (Gervais, 2015) [39]. Además, cabe añadir que, como señala Boxell et al. (2017) [14], en aquellas situaciones en las que el mensaje incívico proviene de un representante político, o fuente de autoridad, los mensajes producidos en torno al mismo son aquellos que contienen mayores niveles de contenido inadecuado.

Además, se ha encontrado que la pertenencia de los usuarios a determinados grupos sociales - e.g.: partidos políticos -, no sólo propicia la aparición de polarización, sino que aumenta los niveles de la misma (Boxell et al., 2017) [14], encontrando un proceso de transmisión actitudinal.

Por otro lado, la comunicación política digital ha cobrado gran prota-

gonismo en las redes sociales. A través de estrategias de marketing, actores políticos ejercen influencia sobre los individuos que consumen su contenido. En los últimos años, estas estrategias han sido afectadas por la revolución tecnológica, donde el uso de cuentas automatizadas, denominadas *bots*, actores tradicionales ejercen su influencia sobre los usuarios que consumen el contenido de la red. A pesar de que el estudio de la utilización de *bots* durante crisis sanitarias apenas ha sido investigado, algunos estudios han arrojado evidencias de la utilización de los mismos como método de difusión para promover dinámicas de conspiración, así como la difusión de mensajes altamente polémicos (Moffit et al., 2021) [75]. De estas forma, y en la misma dirección en la que advierte algunos estudios (Stella et al., 2018[108]; Yan et al., 2020[121]), las cuentas automatizadas aumentan la vulnerabilidad de los usuarios de redes sociales a los mensajes extremos sujetos a la naturaleza de las campañas de comunicación política digital. En concreto, estos *bots* son denominados como *Twitter bombs* (Pastor-Galindo et al., 2020) [88], siendo responsables de la difusión y propagación de mensajes en sintonía de determinadas campañas propagandísticas cuya naturaleza se basa en la información falsa, lenguaje inapropiado y promotor del discurso de odio (Keller et al., 2019[65]; Santana y Huerta-Cánepa, 2019 [101]).

2.2. Medidas de Polarización

Como se mencionó en el apartado anterior, la conceptualización de la polarización presenta un alto grado de heterogeneidad basada en el campo de investigación que estudia dicho fenómeno. El caso de la medición de la polarización no es una excepción a esta tendencia. La principal característica encontrada en la literatura es la falta de consenso en torno a la forma en la que dicho fenómeno debe de ser identificado y medido. Estas diferencias no sólo se encuentran en función a la disciplina que estudia la medición de la polarización, sino que existen diferencias dentro de las mismas disciplinas. Además, la medición de la misma presenta diferencias sustanciales de acuerdo con la polarización que se está midiendo, ya sea política, social, étnica, religiosa, etc.

Por otro lado, la forma de entender la polarización (bipolar o multipolar), así como los elementos que se disponen para su medición, influyen, naturalmente, en el diseño de la medida de la misma. Bramson et al. (2016) [17] propusieron un conjunto de medidas formales de polarización con el fin de estandarizar la medición de dicho fenómeno. Si bien es cierto que a lo largo de la literatura las siguientes medidas han sido usadas como sinónimo de polarización, es importante resaltar que en este trabajo no se consideran medidas

de polarización en sí mismas, sino conceptos que subyacen a la polarización y que han de ser tenidos en cuenta en su conjunto, y no de forma individual. Si bien es cierto que, en escenarios de investigación reales, puede ser difícil disponer de los datos y recursos suficientes medir todos estos conceptos, sí se considera fundamental tenerlos en cuenta para que, en la medida de lo posible, sean aplicados a la medición de la polarización. Así, en este apartado se presentan las aportaciones de Bramson et al. (2016), no como medidas de polarización sino como elementos de la polarización dignos de consideración.

Rango.

Este concepto se considera una de las formas más simples y directas de medir la polarización. De acuerdo con esta idea, sin tener en cuenta la forma de la distribución o su continuidad, cuanto mayor sea la distancia entre el valor actitudinal de los individuos, mayor polarización tiene lugar.

Dispersión.

La dispersión estadística puede, igualmente, ser aplicada en la medición de la polarización en tanto que la dispersión aumenta cuando los individuos de una distribución se alejan del centro hacia los extremos de la distribución. En este sentido, no se tiene en cuenta la noción de grupos en la población. Como ejemplo, los autores proponen la desviación media absoluta, entre otras muchas medidas de dispersión en la literatura.

Cobertura.

La polarización, desde el punto de vista de la cobertura, ocurre cuando todos los valores de una variable se ven reflejados en la población. No importa cuántos niveles hay en la variable, si hay agrupaciones en torno a uno de ellos, o si están situados en los extremos. La máxima polarización se da en aquellos casos donde todos los niveles de la variable están presentes en la población. Se entiende la cobertura como una medida de diversidad, encontrando una relación entre la diversidad actitudinal de una población y su polarización.

Regionalización.

Teniendo en cuenta la cobertura a lo largo de los diferentes niveles de una variable, patrones de espacios vacíos del eje de la variable también aportan información. Así pues, cuánto mayor sea el número de regiones vacías, mayor es la polarización desde un aspecto de regionalización.

Fragmentación comunitaria.

El grado en el que una población puede dividirse en sub-poblaciones aumen-

ta el grado de polarización. En este sentido, cuanto mayor sea el número de grupos presentes, mayor nivel de polarización. La detección de grupos en una población proviene de información extra disponible ajena a la variable actitudinal X , e.g.: sexo, etnia, nacionalidad, etc.

Distinción.

Si la población a estudiar se compone de un conjunto de grupos claramente separados y distinguidos los unos de los otros, se entiende que está polarizada. Cuánto más clara sea esta separación, mayor la polarización presente. En Bramson et al. [17] proponen como medida de distinción el test de Kolgomorov-Smirnov (KS), en tanto que esta prueba se centra en la determinación de que dos conjuntos de datos pertenezcan, o no, a la misma distribución.

Divergencia grupal.

Al contrario que la distinción grupal, la divergencia grupal aporta la máxima polarización cuanto más alejados se encuentren los grupos entre sí, sin tener en cuenta la forma de su distribución.

Consenso grupal.

La polarización, en el sentido del consenso grupal, refleja el grado en el que se forman los grupos de una población. De esta forma, cuando hay presentes múltiples grupos, cada uno de ellos unificados en torno a una actitud - alto consenso intra-grupal -, se encuentra la presencia de polarización. Bramson et al., proponen como medición del consenso grupal cualquier medida de dispersión.

Tamaño grupal.

Manteniendo el número de grupos constante, cuanto más similares sean los grupos en su tamaño, mayor polarización estará presente en la población. De esta forma, poblaciones que presentan dos grupos, pero uno de ellos representa la mayoría (e.g.: un 90 %) apenas estará polarizada, en contraste con el caso en el que ambos grupos presenten el mismo número de individuos, encontrando así el máximo nivel de polarización.

Es importante observar que alguna de las nociones vagamente expuestas anteriormente parecerían insuficientes para medir correctamente la polarización. Para ilustrar esto, se establece el siguiente ejemplo. Se quiere medir la polarización a través de la divergencia grupal a lo largo de una variable actitudinal X con una población de $N = 100$, compuesta por dos grupos,

un grupo A con $n_1 = 90$ y un grupo B con $n_2 = 10$, con un máximo de divergencia grupal. De acuerdo con este punto de vista, cuanto mayor es la distancia entre las actitudes de los grupos, mayor es la polarización existente. Sin embargo, el problema se presenta al no considerar otros parámetros, como lo puede ser el tamaño de los grupos o su consenso. De esta forma, podría darse una gran amplitud entre las actitudes de los grupos, pero presentar un tamaño desbalanceado entre los mismos. En este caso, no se puede hablar de polarización en tanto que nos encontramos ante un escenario donde la mayoría de la población presenta la misma actitud, frente a una minoría que difiere. Además, estos conceptos pueden ser insuficientes si se quiere medir la polarización en un contexto de redes sociales, donde si bien algunos de ellos se pueden rescatar para la medición de la polarización en estos escenarios, conceptos como el de flujo de comunicación o la consideración de la estructura de la red se vuelven imprescindibles.

Así pues, se considera primordial remarcar la importancia que presenta la consideración de los conceptos propuestos por Bramson et al. [17] como actores que deberían influir en la medida de polarización pero que individualmente no recogen la idea global de polarización. Sin embargo, en tanto que la medición de la polarización está estrechamente ligada a la conceptualización y entendimiento de la misma, es gracias a la propuesta de estos conceptos en Bramson et al. (2016) [17] que es posible comprender de una manera más adecuada el concepto de polarización y qué características debe presentar una medida para cubrir la medición de dicho fenómeno de la forma más adecuada posible.

Por otro lado, a lo largo de la literatura el concepto de polarización ha estado fuertemente ligado, y a veces confundido, con el concepto de variación. Variabilidad, dispersión y varianza representan conceptos clave en la estadística, usados para describir tanto distribuciones de variables aleatorias así como valores observados en variables estadísticas. En este último contexto, de acuerdo con Martínez et al. (2019) [73], la medición de la dispersión ha sido tradicionalmente asociada a variables estadísticas continuas. Sin embargo, cuando la variabilidad ha de ser medida en variables ordinales, la aproximación común es la de entender la naturaleza ordinal como numérica a través de la asignación de números a cada uno de los niveles de la variable categórica. Mas, por otro lado, existen propuestas en la literatura que permiten la medición de la variación en variables categóricas.

De esta forma, diferentes medidas de dispersión ordinal han sido propuestas en la literatura (e.g.: Blair y Lacy (2000) [11], Franceschini et al. (2004) [31] o Gadrich y Bashkansky (2012) [36]) para afrontar, de forma apropiada, la medición de la variabilidad en variables categóricas sin la necesidad de

forzar la conversión de una variable categórica a una variable numérica, lo cual puede dar lugar malinterpretaciones.

En esta investigación, a pesar de la existencia de diferentes medidas de dispersión ordinal, se presenta la propuesta por Berry y Mielke (1992) [9], tradicionalmente conocido como *IOV*.

Definición 2.2.1 (Medida de dispersión ordinal IOV). Dada una variable ordinal con valores $X = \{L_1, \dots, L_n\}$ y un vector de frecuencias $f = (f_1, \dots, f_n)$, donde se tiene una población $N = \{1, 2, \dots, i, \dots, n\}$ el índice de dispersión ordinal *IOV* se define como:

$$IOV = \sum_{i=1}^{n-1} \sum_{j=i+1}^n f_i f_j (j - i). \quad (2.1)$$

Finalmente, en los siguientes apartados se presentan un conjunto de medidas de polarización basadas en el entendimiento bipolar de la polarización. Así mismo, se hará una distinción en torno a la agrupación de las diferentes medidas, de acuerdo a su ámbito de aplicación, ya sea el tradicional o basadas en redes.

2.2.1. Medidas de bipolarización tradicionales.

En este trabajo se denomina medidas de polarización tradicionales a aquellos modelos matemáticos que miden la polarización desde una perspectiva nítida - sin contemplar incertidumbre - y que adoptan un enfoque bipolar, donde la polarización es medida a través de una perspectiva estática. Así pues, estas medidas contemplarán la medición de la polarización a través del supuesto de existencia de dos polos como una fotografía en el continuo del tiempo. A continuación, se definen algunas de las medidas de polarización más famosas en la literatura.

2.2.1.1. Foster y Wolfson (1992) [30].

Foster y Wolfson se sitúan como unos de los primeros autores en medir la polarización en la literatura. Los autores, situados bajo el paradigma de la economía, entienden la polarización económica en torno a la renta de la población como la desaparición de la clase media. A su vez los autores establecen que es necesario la presencia de bipolaridad, es decir, la renta de los individuos no sólo debe alejarse de la media, sino que éstas desviaciones deben presentar valores mayores y menores respecto con la media poblacional. Así pues, para medir la polarización económica, Foster y Wolfson se basaron en el índice de desigualdad de Gini [40] presentando la siguiente medida:

Definición 2.2.2 (Polarización de Foster y Wolfson [30]). Sea $x = \{x_1, \dots, x_N\}$ un vector N-dimensional de rentas positivo, x_i la renta de un individuo i , $m(x)$ la mediana de x y $\mu(x)$ su media. La medida de polarización se define como:

$$P_{FW}(x) = 2(2T - G) / (m(x)/\mu(x)) \quad (2.2)$$

Donde G es el índice de Gini y $T = (\mu(x^+) - \mu(x^-))/2(\mu(x))$. Así mismo, sean $\mu(x^+)$ y $\mu(x^-)$ las medias de x^+ y x^- respectivamente, donde x^+ son las rentas de los individuos que se sitúan por encima de $m(x)$, así como x^- las rentas de los individuos que se sitúan por debajo de $m(x)$.

2.2.1.2. Esteban y Ray (1994)[28].

Esteban y Ray suponen otros de los primeros autores que propusieron una medida formal de polarización en la literatura. Al igual que Foster y Wolfson, los autores se ubican en el marco economía. Estos autores presentan una medición de la polarización que incluye aspectos más allá de los meramente métricos. Así, para los autores, el concepto de polarización representa la aparición de una división grupal en la sociedad, con bajo número de grupos, los cuales representan características específicas. Concretamente, los miembros de un mismo grupo han de compartir ciertas características entre sí - *alta homogeneidad intragrupal* - y los grupos existentes deben presentar diferencias en torno a estas características - *alta heterogeneidad intergrupala* -. Partiendo de estas hipótesis, los autores proponen la definición de antagonismo efectivo $T(I, a)$, donde I representa la identificación de un determinado individuo i con su grupo, y A representa la alienación que un individuo i presenta con un individuo j .

Definición 2.2.3 (Polarización de Esteban y Ray [28]). Sea $X = \{x_1, \dots, x_N\}$ un vector N-dimensional de rentas positivo y x_i la renta de un individuo i . Sea π_i la cantidad de individuos que presentan la misma renta que i , siendo π_i el tamaño del grupo de i , y $T(I, a)$ el antagonismo efectivo, entendido como el producto entre I y a , donde I es la identificación de un determinado individuo i con su grupo, determinada por su tamaño π_i , y sea a la alienación que un individuo i presenta con un individuo j , determinada por $\delta(x_i, x_j)$, la cual es la diferencia absoluta entre las rentas de dos individuos $|x_i - x_j|$. Así, se define la medida de polarización de Esteban y Ray como:

$$P_{ER}(\pi, y) = \sum_{i=1}^n \sum_{j=1}^n \pi_i \pi_j T(I(\pi_i), a(\delta(x_i, x_j))) \quad (2.3)$$

De acuerdo con la ecuación anterior, la versión más utilizada es reformulada para aquellos casos en los que la única información disponible para una determinada población $N = \{1, \dots, n\}$ es la variable X , con su distribución relativa $\{(x_1, \pi_1), \dots, (x_n, \pi_n)\}$. Así, la fórmula anterior puede ser representada, sin pérdida de generalidad, como:

$$P_S(X, \alpha) = \sum_{i=1}^n \sum_{j=1}^n \pi_i \pi_j (\pi_i^\alpha |x_i - x_j|) \quad (2.4)$$

Donde $\alpha \geq 1$ - ver Esteban y Ray (1994)[28] -.

2.2.1.3. Montalvo y Reynal-Querol (2003)[76].

A diferencia de las medidas propuestas anteriormente, Montalvo y Reynal-Querol propusieron un índice de polarización situado en el marco de la polarización religiosa. Para los autores, la polarización es un fenómeno que tiene lugar en aquellos escenarios en los que una sociedad presenta únicamente dos grupos religiosos diferentes cuyo tamaño es idéntico. En definitiva, en Montalvo y Reynal-Querol (2003) [76] miden la polarización como la distancia normalizada de una distribución particular de grupos étnicos o religiosos a partir de una distribución bimodal:

Definición 2.2.4 (Polarización de Montalvo y Reynal-Querol [76]). Sea $N = \{1, 2, \dots, n\}$ un conjunto de individuos, $C = \{C_1, C_2, \dots, C_n\}$ un conjunto de grupos religiosos o étnicos, se define la medida de polarización como:

$$P_{MR-Q} = 1 - \sum_{i=1}^C \left(\frac{0.5 - \pi_i}{0.5} \right)^2 \pi_i = 4 \sum_{i=1}^C \sum_{i \neq j} \pi_i^2 \pi_j \quad (2.5)$$

Donde $\pi_i = n_i/N_i$ representa la proporción de personas que pertenecen a una determinada religión o grupo étnico.

2.2.1.4. Apouey (2007)[4].

Apouey se basa en la concepción de la polarización inicialmente propuesta por Foster y Wolfson, bajo la premisa de la desaparición de la clase media y el aumento de bipolaridad. Los autores Wang y Tsui (2000) [118] propusieron una generalización del índice de polarización de Foster y Wolfson (1992) [30] en el que se basa Apouey para construir un índice de polarización para escalas categóricas u ordinales en el ámbito de la polarización sanitaria. Bajo estas premisas, Apouey defiende que el escenario de mínima polarización se da en

aquellos casos en los que todos los individuos presentan una misma categoría, mientras que los niveles altos se presentan cuando la población está dividida en dos mitades, representando una de éstas la categoría más baja y la otra mitad los valores más altos de la variable categórica.

Definición 2.2.5 (Polarización de Apouey [4]). Sea $N = (N_1, \dots, N_{C-1}, N_C)$ una distribución acumulativa, y C las diferentes clases existentes, donde $F = (F_1, \dots, F_{C-1}, F_C)$ es la frecuencia acumulada, K_1 es una constante estrictamente positiva y $\alpha \in [0, 1]$, se define la medida de polarización de Apouey como:

$$P_A(F) = K_1 \left[\left(\frac{1}{2} \right)^\alpha - \frac{1}{C-1} \sum_{c=1}^{C-1} \left| F_c - \frac{1}{2} \right|^\alpha \right] \quad (2.6)$$

2.2.1.5. Dalton (2008)[21].

La medida de polarización propuesta por Dalton se centra en la medición de la polarización política en un sistema en el que tienen lugar múltiples partidos. Así, el autor entiende que un sistema político está fuertemente polarizado en aquellos casos en los que la fuerza centrífuga distribuye los votos de la población hacia los partidos situados en los extremos de un eje ideológico. Para ello, tienen en cuenta dos ideas principales: (1) la posición de un determinado partido político a lo largo del eje ideológico izquierdas-derechas, y (2) la posición del partido político en este eje ponderada por el tamaño o representación del mismo.

Es importante mencionar que la siguiente medida se computa de acuerdo con la escala ideológica categórica de izquierdas-derechas que presenta un rango de 0 – 10.

Definición 2.2.6 (Polarización de Dalton [21]). Sea $C = \{1, 2, \dots, i, \dots, c\}$ un conjunto de partidos políticos, $I = \{1, 2, \dots, 10\}$ una escala política/ideológica de izquierdas-derechas, donde I_i representa la puntuación en dicha escala de un partido i y sea $v = \{v_1, v_2, \dots, v_i, \dots, v_p\}$ un vector del número de votos que han recibido los partidos C , se define la medida de polarización de Dalton como:

$$P_D = SQRT \left\{ \sum_{i=1}^c (v_i) \left(\frac{I_i - \mu_I}{5} \right)^2 \right\} \quad (2.7)$$

2.2.1.6. Permanyer y D'ambrosio (2015)[89].

Esta propuesta plantea la unión de dos concepciones de la medición hasta el momento, contempladas por separado. En primer lugar, se tiene en cuenta la medida tradicional de bipolarización en torno a la renta de Esteban y Ray (1994) [28]. Por otro lado, a esta idea, añade la existencia de grupos predefinidos *a priori* implícitos a las características de los individuos. Si bien ya Esteban y Ray hablaban de grupos, estos grupos venían dados por las rentas que los individuos presentaban. En esta ocasión, se plantea la agregación de información extra a la variable con la que se mide la polarización. De esta forma, si se mide la polarización en torno a la renta, se estará teniendo en cuenta la posible agrupación de la población en torno a e.g.: sexo o etnia.

Definición 2.2.7 (Polarización de Permanyer y D'ambrosio (2015) [89]). Sea $C^c = \{C_1, \dots, C_i, C_j, \dots, C_c\}$ los grupos exógenos presentes en la población, y sea $N = \{\pi_1, \pi_2, \dots, \pi_i, \pi_j, \dots, N_c\}$ el tamaño proporcional de dichos grupos. Finalmente, sea $T(\pi_i, \theta_{ij})$ el antagonismo efectivo entre un grupo i y un grupo j , que viene dado por el producto entre π_i y θ_{ij} , donde $\theta_{ij} = 1$ cuando la comparación se efectúa entre grupos diferentes y $\theta_{ij} = 0$ cuando la comparación se da entre un mismo grupo.

$$P_{PD}(C^c) = \sum_{i=1}^c \sum_{j=1}^c \pi_i \pi_j T(\pi_i, \theta_{ij}) \quad (2.8)$$

Observación 1. Llegados a este punto, resulta de gran interés analizar en mayor profundidad las medidas de polarización de Montalvo y Reynal-Querol (2003) [76] y la medida propuesta por Permanyer y D'ambrosio (2015) [89]. La medida P_{MR-Q} definida en 2.2.4 es equivalente, o un caso particular, a P_{PD} cuando $\alpha = 1$.

$$\begin{aligned} P_{PD}(C^c) &= \sum_{i=1}^c \sum_{j=1}^c \pi_i \pi_j T(\pi_i, \theta_{ij}) = \sum_{i=1}^c \sum_{j=1, j \neq i}^c \pi_i \pi_j T(\pi_i, 1) = \\ &= \sum_{i=1}^c \sum_{j=1, j \neq i}^c \pi_i \pi_j \pi_i^\alpha = \sum_{i=1}^c \sum_{j=1, j \neq i}^c \pi_i^2 \pi_j = \\ &= \sum_{i=1}^c \pi_i^2 (1 - \pi_i) \approx P_{MR-Q} \end{aligned}$$

Recuérdese que en la ecuación de P_{MR-Q} , para convertir la medida en

índice, la presentan como:

$$P_{MR-Q} = 4 \sum_{i=1}^C \sum_{i \neq j} \pi_i^2 \pi_j$$

Por lo que:

$$P_{PD} = \frac{1}{4} P_{MR-Q}$$

A pesar de que las medidas anteriores pueden considerarse entre las más citadas en la literatura, existen otras medidas no definidas en esta sección que y que presentan un potencial interés, como la propuesta por Wagner (2021) [117], la cual aborda la medición de la polarización desde el marco de la afectividad, planteando un índice de polarización afectivo para sistemas políticos con múltiples partidos. En ella, Wagner indica que basa la medición de la polarización desde la fragmentación, cuestión que antecede un comportamiento de la medida distinto al que cabe esperar de la definición general de polarización “existencia dos polos opuestos y antagónicos de tamaño significativo y similar”. Otro ejemplo es la medida de *love and hate* propuesta por Losada et al., (2021) [70], la cual se basa en medir la polarización de forma bidimensional a través de los continuos amor-odio hacia dos candidatos o representantes políticos.

2.2.2. Medidas de polarización basadas en redes.

El estudio de la sociedad a través del estudio de redes sociales con grafos permite a los investigadores acceder, conocer, estudiar y analizar la estructura social que subyace a determinados eventos de interés. De esta forma, la teoría de grafos ha proporcionado al estudio de la sociedad un nuevo enfoque a través del cual analizar aspectos que, anteriormente, se antojaban inalcanzables. Esta nueva herramienta se ha trasladado al estudio de la polarización, donde se plantea un nuevo paradigma en la medición de la misma, migrando de una polarización basada en actitudes, ideologías o variables, en general, con una determinada distribución en la población, al estudio de una polarización estructural, donde cada individuo presenta una posición de carácter relacional dentro de la sociedad, cuya actividad se basa, principalmente, en la interacción con otros individuos. Si bien es cierto que al estudio de grafos se puede agregar información de otras variables externas a la estructura del grafo, como actitudes o ideologías, esta información es complementaria - al contrario que en la medición tradicional de la polarización -, donde la naturaleza de los datos con los que se mide la polarización es relacional. De esta

forma, en los últimos años han emergido diferentes medidas de polarización basadas en grafos, donde los nodos representan individuos y las aristas sus relaciones entre ellos. Así, cabe presentar el concepto de polarización estructural que, frente a las medidas de polarización comentadas con anterioridad, que miden la polarización ideológica o actitudinal, se mide la polarización desde una perspectiva relacional.

2.2.2.1. Waugh et al. (2009)[119].

Una de las primeras propuestas de medición de polarización en grafos fue propuesta por Waugh et al. en 2009. Los autores, con el fin de abordar la medición de la polarización en el congreso de los Estados Unidos, propusieron utilizar el concepto de modularidad (Newman, 2006) [81] como medida de polarización. En términos generales, se entiende por modularidad en la teoría de grafos como la medida en la cual las comunidades detectadas en una red están bien delimitadas y construidas entre sí. Bajo esta premisa, si una determinada partición presenta comunidades cuyos nodos han generado un alto número de aristas o relaciones dentro de su propia comunidad, y pocas relaciones fuera de la comunidad a la que pertenecen, y este hecho es común a todas las comunidades detectadas dentro de la partición, se obtiene un alto grado de modularidad.

Definición 2.2.8 (Modularidad [81]). Sea $G = (V, E)$ un grafo no dirigido A su matriz de adyacencia. Sean $i, j \in V$ dos nodos y $m = |E|$ el número de aristas de G . Sea P un subconjunto del de nodos de V , se tiene que :

$$Q(P) = \frac{1}{2m} \sum_{i,j \in V} \left[A(i, j) - \frac{k_i k_j}{2m} \right] \delta(C_i, C_j) \quad (2.9)$$

siendo k_i el grado del nodo i y C_i el grupo de dicho nodo; $\delta(C_i, C_j) = 1$ si $C_i = C_j$, y $\delta(C_i, C_j) = 0$ en otro caso.

2.2.2.2. Guerra et al. (2013) [49].

Guerra et al. presentan en la literatura una medida que plantea la medición de la polarización centrada en las fronteras que tienen lugar entre las comunidades identificadas en un determinado grafo. Para ello, se consideran todos aquellos nodos - o individuos - que presentan una posición topológica en torno a dos comunidades diferentes. Se determina que un nodo v pertenece a la frontera entre dos comunidades C_1 y C_2 si, perteneciendo a una de estas dos comunidades presenta aristas con, al menos, otro nodo de una comunidad distinta a la suya, a la vez que presenta relaciones con, al menos,

otro nodo de su misma comunidad que no presenta relaciones con otros nodos exogrupales.

Definición 2.2.9 (Polarización basada en la frontera [49]). Sea $G = (V, E)$ un grafo no dirigido y $v \in V$ un determinado nodo o vértice. Sea B el conjunto de nodos que se encuentran en la frontera entre dos comunidades C_i y C_j . Sea $d_i(v)$ el número de relaciones que presenta el nodo v con nodos de su misma comunidad y que no pertenecen a la frontera y $d_b(v)$ es el número total de aristas que v presenta con nodos de otra comunidad. Así, se define la medida de polarización basada en la frontera como:

$$P_{GMCK} = \frac{1}{|B|} \sum_{v \in B} \left[\frac{d_i(v)}{d_b(v) + d_i(v)} \right] - 0.5 \quad (2.10)$$

Los valores de P_{GMCK} oscilan entre -0.5 y 0.5 . Valores < 0 indican ausencia de polarización, indicando que los nodos que se encuentran en la frontera tienden a interactuar más con los nodos de la otra comunidad que con la suya.

Las medidas de polarización basadas en redes propuestas en los párrafos anteriores han supuesto las más usadas a lo largo de la literatura. Sin embargo, otras menos usadas, pero igual de interesantes, han sido propuestas en los últimos años, cuya computación requiere, en ocasiones, de información extra exógena al grafo, como es el caso de la medida propuesta por Morales et al. (2015)[78]. Estos autores definen la polarización política máxima cuando la sociedad se divide en dos grupos de mismo tamaño y opiniones opuestas. Las perspectiva a través de la cual los autores miden la polarización política se basa en la consideración de formaciones grupales de personas en torno a posiciones extremas, donde la clase moderada se reduce al mínimo. Para medir esta polarización, estudian la propagación de información entre nodos a lo largo de un grafo, donde consideran tres factores principales: (1) calculan la opinión de la población, (2) computan el centro gravitacional del grafo de acuerdo con las diferentes opiniones existentes, donde posteriormente calculan la distancia normalizada entre ambos centros (dos opiniones). Finalmente, (3) calculan la diferencia entre el tamaño poblacional y la distancia entre opiniones.

Otros autores, sin embargo, bordan la problemática desde la perspectiva de la controversia, como es el caso de Garimella et al. (2018). Para ello, se basan en el concepto de caminos aleatorios - *random walks* - a través de un grafo. Partiendo de la idea de que existen líderes de opinión en cada comunidad, computan la probabilidad de que un determinado nodo, perteneciente a una comunidad, acabe siendo expuesto a la influencia de la mayor autoridad

de la comunidad contraria. A su vez, los autores [37] proponen una segunda medida de controversia basada en la centralidad en grafos. Los autores parten de la idea de que, si dos comunidades están bien delimitadas y separadas entre sí, si se cortan las aristas entre las mismas, únicamente afectará a esas aristas que actúan como puente entre comunidades.

2.3. Teoría de grafos y detección de comunidades

Uno de los objetivos planteados en esta memoria es el estudio de la polarización en redes sociales. Probablemente, el objeto matemático más habitual para modelizar relaciones entre objetos y, en particular, las relaciones entre individuos en una red social son los grafos. Por este motivo, en esta sección se presentan algunas nociones básicas sobre esta importante parte de las matemáticas así como sobre uno de sus problemas más relevantes: la detección de comunidades.

A lo largo de los últimos años, la esfera pública digital se ha convertido en un escenario virtual que representa una fuente inagotable de recursos donde se producen millones de datos a cada segundo. Más específicamente, las redes sociales online suponen espacios virtuales donde las personas hacen de su uso con la intención de interactuar con otros usuarios. En contraposición a la metodología tradicional de recogida de datos, Internet permite a los investigadores la recolección de millones de datos de forma prácticamente instantánea, bien estructurados y preparados para su análisis. De esta forma, el análisis de estas nuevas fuentes de información ocasionalmente presentes en entornos Big Data, ha tomado un gran protagonismo en los últimos años por dos motivos principales: (1) la gran accesibilidad de los datos que proporciona y (2) la cantidad de los mismos que puede ser recolectada. Junto con el *Social Big Data Analysis*, se presentaron nuevas técnicas de aplicación a este contexto, como el análisis de redes sociales o *Social Network Analysis (SNA)* en la que se centrará el presente trabajo que, si bien en su mayoría ya estaban presentes en la literatura años antes, estas técnicas alcanzan su mayor potencial en los contextos actuales de *Open Data*.

2.3.1. Concepto de grafo y digrafo.

Se define matemáticamente un grafo G como un par ordenado donde $G = (V, E)$, donde V representa el conjunto de nodos o vértices y E representa las aristas. Un aspecto importante sobre los grafos es si se contempla la

direccionalidad de las aristas, es decir, si se contempla si un determinado nodo v_i se relaciona con otro nodo v_j , o si simplemente se contempla la existencia de una relación entre v_i y v_j independientemente del origen de la acción. Para este último caso, se habla de grafos no dirigidos, mientras que para los casos donde se contabiliza la direccionalidad y, como consecuencia, el origen de la relación, se llaman grafos dirigidos o digrafos, donde las aristas pasan a denominarse arcos.

Definición 2.3.1 (Grafo [112]). Sea $V = \{1, \dots, n\}$ un conjunto finito de elementos, llamados vértices. Sea $E = \{\{i, j\} \mid i, j \in V\}$ un conjunto no ordenado de pares de vértices, llamados aristas o arcos. El par $G = (V, E)$ es llamado grafo o red.

Se dice que dos vértices $i, j \in V$ están conectados directamente en G si existe una arista $\{i, j\} \in E$. El número de vértices se denota como $n = |V|$, y el número de aristas como $m = |E|$. Para cada vértice $i \in V$, $H(i)$ representa el conjunto de vértices vecinos o adyacentes a i . El grado de i , denotado por k_i , es la cardinalidad de su conjunto de vecinos, $k_i = |H(i)|$.

Así pues, para visualizar lo anterior, supóngase que se tiene un determinado grafo $G = (V, E)$ con cuatro nodos representados como $V = \{1, 2, 3, 4\}$ y cuatro pares de relaciones determinadas por $E = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{1, 3\}\}$. De esta forma, los nodos 1 y 2 presentan una relación entre ellos, así como el 2 y 3, 3 y 4, y finalmente, el 1 y el 3. Una posible representación gráfica de este grafo se puede ver en la figura 2.3.

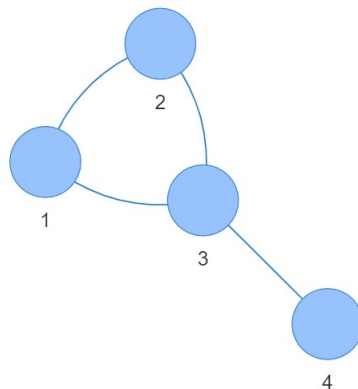


Figura 2.3: Grafo simple.

Desde el punto de vista del almacenamiento como conjunto de datos de un grafo, existen diferentes alternativas que se mencionan brevemente a continuación, todas ellas presentan sus pros y sus contras. Entre las estructuras más conocidas podemos citar las listas y las matrices. En primer lugar, (1) las listas de incidencia representan una lista de relaciones, ya sean dirigidas o no dirigidas, donde el origen y destino de una relación entre nodos es representado. De esta forma, si tenemos el ejemplo de $V = \{1, 2, 3\}$; $E = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$, se tienen tantas filas en la matriz como pares de interacciones, con tres columnas, el nodo origen, el nodo destino y los pesos de las interacciones. Por otro lado, (2) las listas de adyacencia representan, en cada fila, un determinado nodo, junto con sus relaciones correspondientes. Siguiendo el ejemplo anterior, la estructura de datos tomaría forma de una matriz con tantas filas como nodos origen haya y tres columnas diferentes, una para identificar el nodo origen, una segunda columna para reflejar las conexiones de este nodo, y la tercera que indicará los pesos de las interacciones del nodo origen por cada nodo destino. Finalmente, (3) las matrices de adyacencia suponen otra forma alternativa de representar un grafo. En este caso, esta matriz de formato $n \times n$ será una matriz cuadrada, donde las filas y las columnas representan los nodos de la red por igual. El contenido de la matriz será, en este caso, un compendio de w_{ij} y 0, donde el valor A_{ij} tomará el valor de w_{ij} cuando exista una relación entre los nodos i y j y 0 cuando no, siendo w_{ij} el peso o de la relación entre ambos nodos. Para redes no valoradas, éste valor es igual a 1.

Definición 2.3.2 (Matriz de adyacencia). Sea $G = \{V, E\}$ un grafo, se denomina matriz de adyacencia de ese grafo $A(G)$ para cada par de vértices o nodos $i, j \in V$ como:

$$A_{ij}(G) = \begin{cases} w_{ij} & \text{si } \{i, j\} \in E \\ 0 & \text{si } \{i, j\} \notin E \end{cases} \quad (2.11)$$

donde w_{ij} es el peso asignado a la relación entre los nodos $\{i, j\}$.

Por ejemplo, dado un grafo no valorado $G = (V, E)$ con nodos $V = \{1, 2, 3, 4\}$, y una lista de aristas $E = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{1, 3\}\}$, se corresponde con la siguiente matriz de adyacencia:

$$A(G) = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Una generalización del concepto de grafo puede darse en el caso en el que para cada arista $\{ij\}$ tengamos un peso w_{ij} que represente la fuerza o el coste de la relación entre $\{ij\}$. Esta extensión es una representación más realista de la realidad ya que no todas las relaciones entre elementos de un conjunto tienen que modelizarse con la misma intensidad.

Definición 2.3.3 (Grafo Valorado [112]). Un grafo $G = (V, E)$ es valorado si cada una de sus aristas tiene asignado un peso e.g. existe una función definida sobre el conjunto de aristas que a cada una de ellas le asigna un valor, $w : E \rightarrow \mathbb{R}$, donde $w(\{i, j\})$ es el peso de la arista $\{i, j\}$. Se puede considerar $w_{ij} := w(\{i, j\})$.

Dado un grafo $G = (V, E)$ se dice que G es conexo si existe un camino entre cualquiera de los nodos que pertenecen al grafo. Asimismo, para los grafos dirigidos, se denomina grafo no dirigido asociado al obtenido de G sin tener en cuenta la dirección de las aristas. En este escenario, para aquellos casos en los que el grafo asociado es conexo, lo es también G . En aquellos casos en los que se presentan grafos no conexos, estos pueden ser segmentados en subpartes más pequeñas y conexas, denominadas componentes conexas.

Otro concepto relevante en la teoría de grados es el concepto de camino. Se entiende como camino \neq_{ij} entre dos nodos i y j de V a una serie de vértices pertenecientes a V donde existe una arista entre cada vértice y el siguiente, con una longitud k . De esta forma, se puede concluir que dos vértices están conectados entre sí si existe un camino entre ellos. Formalmente, lo explicado anteriormente se puede expresar como $\pi_{ij} = (i_0 = i, i_1, \dots, i_k = j)$, representando una serie de vértices donde $i_l, i_{l+1} \in E \forall l = 0, \dots, k - 1$. En aquellos casos en los que no existe un camino que vaya de un nodo a otro, se dice que están desconectados.

También es posible generar redes aleatorias, o no deterministas, a través de métodos de generación de redes reales no deterministas. Entre todos los existentes, hay tres de ellos que son de gran interés, siendo los modelos más estudiados en la literatura. En primer lugar, (1) las redes completamente aleatorias representan estructuras donde las conexiones entre nodos no se corresponden con ningún patrón en específico, siendo principalmente utilizadas para la generación de otros tipos de grafos. El modelo de Erdos-Renyi (1969)[27] se encuentra entre uno de los más conocidos. Por otro lado, (2) las redes de pequeño mundo son aquellas que simulan situaciones reales relacionadas con el experimento de seis grados de separación, entre otras. Dentro de este tipo de modelos, es más famoso en la literatura es el propuesto por Watts-Strogaz (ver Newman y Watts, 1999)[82]. Finalmente, (3) las redes libres de escala con conexión preferencial son aquellas que simulan los escenarios en los que los nodos de la red muestran una tendencia a conectarse con mayor

probabilidad a los nodos que presentan el rol de líderes. Finalmente, tal y como se mencionó anteriormente, la principal característica de los gráficos no dirigidos es que no tienen en cuenta la direccionalidad de las relaciones. Sin embargo, tener en cuenta este aspecto puede ser de gran interés en determinadas situaciones. Este matiz es muy importante, pues la naturaleza de la red es significativamente diferente, dado que nos permite conocer la dirección de los arcos, teniendo un mayor, y más preciso, conocimiento de las relaciones entre nodos.

Definición 2.3.4 (Grafo dirigido). Un grafo es dirigido si sus aristas tienen un sentido definido, definiendo un nodo, o vértice, de origen y otro vértice de destino. Así, la arista (i, j) del digrafo G presenta i como vértice origen, y a j como vértice destino. En los digrafos, las relaciones entre nodos no tienen por qué ser simétricas, a diferencia de lo que ocurre con grafos no dirigidos. Cada enlace de un grafo no dirigido es un par no ordenado de nodos, de modo que la arista $\{i, j\}$ es la misma que $\{j, i\}$.

En algunas ocasiones, es común seccionar la red en diferentes partes de acuerdo con los conceptos de grafo parcial y subgrafo. En este sentido, se entiende por grafo parcial como un grafo H que presenta la misma cantidad de nodos que el grafo original G , así como sus aristas representan un subconjunto del vector de relaciones E . De esta forma, únicamente se considera $H = (V, L)$ como un subgrafo de G si la totalidad de L está representada en E . Así mismo, se define como subgrafo al subconjunto de nodos que pertenece al vector original V y que mantiene sus mismas aristas. De esta forma, se define un subgrafo G_S como el par $(S, E|S)$ donde $E|S = \{i, j \in S \text{ con } (i, j) \in E\}$.

2.3.2. Algunas medidas globales sobre grafos.

Una de las medidas, indicadores o características, más sencillas asociadas a una red es su tamaño. El tamaño de una red o grafo cualquiera $G = (V, E)$ está determinado por el cardinal de los conjuntos de sus nodos $V(|V|)$ y sus aristas $E(|E|)$. Por otro lado, la densidad de una red nos proporciona la información correspondiente al número de interacciones presentes de un grafo en relación con todas las que podría tener dado un número n de nodos. En este sentido, dado un grafo no dirigido, la densidad se define a través de la siguiente fórmula, interpretándola como el porcentaje de relaciones que se observan en el grafo respecto a todas aquellas que podría tener teóricamente.

Definición 2.3.5 (Densidad). Sea G un grafo, E el conjunto de aristas y V

los nodos, se entiende por densidad de una red como:

$$d(G) = \frac{|E|}{|V| \frac{|V|-1}{2}}$$

Por su parte, el coeficiente de agrupación, también conocido por su término en inglés como *clustering coefficient* ofrece la información relativa a si, para un determinado nodo i , todos sus contactos o vecinos están en relación entre sí. De esta forma, si un determinado subgrafo generado a partir del nodo i presenta las características de un grafo completo, el *clustering* coeficiente muestra su valor más alto. Los valores más bajos indican que un determinado nodo i presenta poca conectividad con sus vecinos.

Definición 2.3.6 (Coeficiente de *clustering*). Sea k_i como el número de conexiones directas de i tiene con otros nodos, donde $i \in V$, y sea L_i las relaciones de estos otros nodos, se define el *clustering coefficient* para el nodo i (C_i) de la siguiente forma:

$$C_i = \frac{L_i}{k_i \frac{k_i-1}{2}}$$

En aquellos casos en los que se agregan los diferentes coeficientes individuales de todos los nodos, se obtiene un coeficiente de agrupación del grafo, representado como:

$$C(G) = \frac{1}{|V|} \sum_{i \in V} C_i$$

El concepto de homofilia (Lazarsfeld y Merton, 1954) [68] ya mencionado anteriormente está fuertemente relacionado con el análisis de redes sociales en tanto que este analiza y representa las interacciones entre individuos. Cuando en un determinado grafo se observan claras agrupaciones de nodos entorno a una característica en común, se dice que su grado de homofilia es alto. Sin embargo, aunque esta homofilia se puede observar a simple vista, es fundamental disponer de información asociada a los nodos que permita conocer y comprobar que, efectivamente, se agrupan en torno a características en común. En aquellos casos en los que no se dispone de esta información adicional, se habla de una red con alta homofilia cuando los nodos que presentan mayor poder en la red tienden a agruparse entre sí con mayor probabilidad, mientras que los nodos con menor poder se juntan, de igual manera, entre ellos mismos. Para poder identificar este escenario, es necesario calcular el coeficiente de correlación entre dos variables. Para calcular esto, dado un grafo $G = (V, E)$, se entiende por X_i como el grado del nodo origen a través

de una determinada arista e Y_i como el grado del nodo destino a través de la misma arista. De esta forma, calculando los grados de poder entre cada uno de los nodos interconectados por una misma arista se obtendrán correlaciones altas y positivas cuando los nodos con grados muy similares se relacionan entre sí. Finalmente, la detección de las componentes conexas supone otro de los aspectos a tener en cuenta a la hora de analizar la topología de la red. Se denomina componente conexa a, dado un grafo no dirigido $G = (V, E)$, las componentes conexas establecen una partición maximal del conjunto de nodos en V (Robles et al., 2020, p. 264)[95]. Calcular el número de componentes conexas permite conocer el número de grupos que pertenecen a un determinado grafo, donde la mayor componente conexa es denominada *Giant connected component*.

2.3.3. Medidas sobre nodos: las medidas de centralidad.

El concepto de centralidad ha sido definido tradicionalmente de forma abstracta, refiriéndose a que un nodo tiene una centralidad alta si presenta la capacidad de interactuar directamente con un alto número de otros nodos, siendo esta interacción de forma rápida y eficaz, y el resto de nodos necesitan de él para comunicarse en la red. Fue a raíz de estas primeras ideas que formalizaciones sobre los conceptos de centralidad llegaron a la literatura (Freeman, 1978[32], 1979[35], 1983[33]), como centralidad por grado, centralidad por cercanía y centralidad por intermediación.

La centralidad por grado se basa en la medición y cuantificación de la cantidad de interacciones que presenta un determinado nodo. Debido a la sencillez conceptual, así como computacional, esta medida de centralidad tiene un uso muy extendido. Como apuntan los autores en Robles et al. (2020)[95], redes sociales online como Facebook o Instagram utilizan la centralidad por grado para conocer el ranking de influencia de los usuarios.

Por otro lado, la centralidad por cercanía se basa en el principio de la distancia que tendría que recorrer un nodo para acceder al resto de nodos de la red. Cuanta menos distancia tenga que recorrer más cerca está del resto. Parte de la idea de tratar de cuantificar el esfuerzo que se tendría que realizar si desde un nodo nos tuviéramos que comunicar con el resto de nodos de la red.

Otras medidas de centralidad son las denominadas medidas espectrales de centralidad, las cuales entienden la importancia de un nodo i como la importancia que presentan los nodos con los que i tiene contacto directo. Entre ellas se encuentra el autovector dominante izquierdo, el índice de Katz/

Alpha centrality, Page Rank, Hub and authorities entre otros (ver por ejemplo Robles et al., (2020) [95] para más detalle).

Similar a las medidas de centralidad por cercanía, las medidas basadas en intermediación por caminos no sólo buscan cuales son las rutas más cortas, sino todas las rutas más directas que acaban en un determinado nodo, siendo el grado *indegree* la medida que refleja esta consideración. La centralidad de *betweenness* es uno de los indicadores más aplicados en el análisis de redes sociales. La medida de *betweenness* se basa en el principio de que, si la comunicación entre nodos se lleva a cabo a través de la ruta más corta, el nodo que esté presente en un mayor número de rutas cortas necesarias para la comunicación entre otros nodos tendrá un alto grado de *betweenness*.

Por otro lado, otra medida de centralidad es la conocida como *Flow betweenness* propuesta por Freeman en 1991 [34]. Esta media se centra en el grado en el que cada nodo k contribuye a la suma de todos los flujos posibles entre otros pares de nodos i y j . En otras palabras, si bien la comunicación o interacción entre nodos se da entre los caminos más cortos existentes entre ellos, supongamos que el nodo k entre ambos no está dispuesto a contribuir a la comunicación.

2.3.4. Detección de comunidades en redes sociales.

Cuando se habla de análisis de redes sociales, uno de los problemas más estudiados y conocidos es el de la detección de comunidades dentro de la propia red. Los problemas de detección de comunidades son sinónimos de problemas tradicionales de *clustering* o agrupación, pues en base a las interacciones mostradas entre los nodos, se agrupan los mismos en diferentes grupos o comunidades. Al agrupar los nodos en comunidades en función a sus interacciones, es de esperar que tales nodos compartan determinadas características o atributos entre sí, aunque esta situación no tiene por qué darse.

Desde un punto de vista matemático, una solución a un problema de detección de comunidades es una partición del conjunto de nodos. Es decir, el objetivo del problema es el de dividir el conjunto de nodos de un grafo en comunidades donde todos los nodos pertenezcan a alguna comunidad y no sea posible que un nodo esté en varias comunidades. Formalmente, y dado un grafo $G = (V, E)$, la idea es identificar una buena partición del grafo $P = \{C_1, \dots, C_k\}$, donde C_i representan cada uno de las comunidades del grafo, y se tiene que $V = \bigcup_{i=1}^k C_i$ es decir, todos los nodos deben estar en alguna comunidad, y $C_i \cap C_j = \emptyset$ para todo i distinto de j - un nodo no puede estar en más de una comunidad -.

Al igual que pasa con los problemas clásicos de *clustering*, existen varian-

tes al problema anteriormente planteado modificando alguna de las condiciones o restricciones anteriores. Esto ha dado lugar a otros nuevos problemas como el de detección de comunidades con solapamiento entre comunidades (Gómez et al., 2016) [45], donde se permite que un nodo esté en más de una comunidad o problemas de detección de comunidades basados en información borrosa - e.g.: Fortunado, 2010 [29]; Gómez et al., 2016)[45] - donde los nodos pertenecen a cada comunidad con cierto grado de pertenencia.

Como se ha mencionado antes, el objetivo del problema de identificar comunidades es el de encontrar una buena partición/agrupamiento del conjunto de nodos. Pero que se entiende por una “buena” o “mala” partición no es un aspecto baladí o trivial. La idea intuitiva que subyace a lo que se entiende por una buena partición es que los grupos que se formen deberían tener muchas relaciones entre ellos y pocas con los demás. Para abordar este problema, en Newman (2006) [81] se propone lo que es hasta la fecha la medida más extendida para cuantificar la calidad de una partición de un grafo: la modularidad, medida ya propuesta anteriormente en la ecuación 2.9.

Debido a lo importante que es identificar comunidades en una red, existen una gran cantidad de algoritmos (ver Fortunato, 2010[29]). Aunque existen muchas otras posibles clasificaciones, una de las más extendidas es la de dividir los algoritmos según el tipo de soluciones que producen que son esencialmente dos grupos mayoritarios: (1) algoritmos jerárquicos y (2) no jerárquicos. Es importante mencionar que la gran mayoría de los algoritmos propuestos en la literatura se centran en la detección de comunidades desde una perspectiva no jerárquica. Entre los algoritmos de detección de comunidades más conocidos, caben destacar los basados en el concepto de disimilitud, división, aglomeración y aleatoriedad. Sin embargo, de entre todos ellos, el más aplicado es el algoritmo de detección de comunidades no jerárquico propuesto por Blondel et al., (2008) [12]. Este algoritmo, conocido más comúnmente como Louvain, es especialmente indicado en aquellos escenarios en los que se tienen grafos de gran tamaño, comunes en los entornos *Big Data*. Este algoritmo se basa en la optimización de la modularidad comparando la densidad de las aristas dentro y fuera de la comunidad. Este algoritmo no sólo tiene la ventaja de ofrecer una detección de comunidades ampliamente aceptado en la literatura, sino que presenta un tiempo de computación veloz en comparación a otros algoritmos, lo que le convierte en uno de los más usados para el SNA en entornos *Big Data*.

Para ilustrar gráficamente un ejemplo de detección de comunidades, se presenta el siguiente gráfico - ver figura 2.4 -. La ilustración representa un subgrafo que contiene las cuatro comunidades de mayor tamaño del grafo principal. Este grafo se ha construido a través de una red de retweets, cuyo

subgrafo presenta 2484 nodos y un total de 18540 aristas. Las comunidades fueron calculadas por el algoritmo de Louvain propuesto por Blondel et al. (2008) [12]. Como se puede apreciar, la densidad de las aristas es mucho mayor dentro de una comunidad que fuera de ellas.



Figura 2.4: Subgrafo de red de retweets con comunidades.

Actualmente, existen algoritmos de detección de comunidades que incorporan información extra agregada con el fin de crear estas comunidades de una forma más realista partiendo de una característica objetivo. En Gutiérrez et al. (2021)[53] los autores utilizan medidas de polarización para crear comunidades más coherentes, y reales, en función a su posición de estar a favor o en contra del gobierno.

2.4. Los conjuntos borrosos

Los conjuntos borrosos han supuesto uno de los inicios de un cambio de paradigma de gran relevancia en el último siglo. Tradicionalmente, la lógica aristotélica de dos valores - e.g.: bueno-malo, alto-bajo o mucho-poco - lleva a los investigadores a contemplar en sus modelos valores nítidos, o *crispy values*, para variables que presentan una naturaleza difusa en sí mismas. De esta forma, la transformación forzosa de la naturaleza de lo que se quiere contemplar puede llevar a imprecisiones de aquello que se quiere medir o modelar. Basándose en este principio, Zadeh (1965) [122] propuso los conjuntos

borrosos, suponiendo una primera alternativa a la teoría de la probabilidad como única herramienta de afrontar la incertidumbre (Klir y Yuan, 1995) [66]. Los conjuntos borrosos permiten representar graduaciones en conceptos que tradicionalmente se han considerado como nítidos como puede ser el hecho de pertenecer a una categoría. A lo largo de los años, los científicos han asumido que la mayoría de los conceptos presentan un entendimiento claro con fronteras bien delimitadas (Belohlavek et al., 2009) [8]. Sin embargo, la ciencia y, especialmente, aquellas áreas que estudian el comportamiento humano, se enfrentan a conceptos que son en su mayoría difusos por naturaleza, dada la complejidad de la naturaleza humana. Para ilustrar esto, supóngase el siguiente ejemplo:

Sea $X = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ una variable tipo likert que mide la ideología política donde 0 se corresponde con una ideología de extrema izquierda, y 10 con una ideología de extrema derecha, donde un participante se sitúa en el valor 5 de la escala, correspondiente al centro de la variable. Bajo esta situación, una aproximación tradicional a el entendimiento de la posición política de este individuo entendería el valor seleccionado por el participante como totalmente cierto. Sin embargo, no es extraño que en escenarios como este, dependiendo de una determinada cuestión específica - e.g.: aborto o economía - una persona pueda adoptar una posición más sesgada hacia la derecha o izquierda política, lo cual entraría en conflicto con la consideración nítida del valor seleccionado en la variable. En este sentido, el entendimiento nítido de los valores puede no ser adecuado para el estudio de conceptos que no presentan delimitaciones específicas, o fronteras claras.

Para solventar este problema, Zadeh (1965) [122] propuso los conjuntos borrosos donde se considera el grado de pertenencia de un determinado objeto hacia un conjunto de atributos o características.

Definición 2.4.1 (Conjuntos borrosos). Un conjunto borroso \tilde{A} sobre el dominio de X se define como:

$$\tilde{A} = \{(x, \mu_A(x)) \mid x \in X\}$$

donde $\mu_A(x)$ representa la función del grado de pertenencia, e.g.: $\mu_A : X \rightarrow [0, 1]$.

La aplicación de los conjuntos borrosos supuso una revolución en cuando al desarrollo de modelos matemáticos que explican fenómenos del mundo real, donde los valores nítidos raramente son aplicados. Los conjuntos borrosos se extienden de igual manera al campo de la clasificación (Amo et al., 2004) [3]. Así, la asignación de un objeto x a un conjunto de clases C ha venido tradicionalmente por la consideración nítida de un único valor, visto en el

ejemplo de la escala de ideología política que se ha expuesto anteriormente. Sin embargo, ya se ha ejemplificado como esto puede suponer una simplificación poco realista de la realidad. Para abordar la clasificación desde una perspectiva borrosa, suele asumirse la existencia de un conjunto de clases C a priori, que pueden venir dadas de diferentes formas, como puede ser una variable tipo likert, por ejemplo. De esta forma, una vez determinadas las clases C , se asigna un determinado grado de pertenencia de un objeto x donde $x \in X$ a una clase c donde $c \in C$. Esta función de pertenencia viene dada por:

Definición 2.4.2 (Clasificación borrosa). Sea x un conjunto de interés, donde $x \in X$, y sea C un conjunto de clases previamente definidas, donde $c \in C$, se define la función de grado de pertenencia de un objeto x a una clase c como:

$$\mu_c : X \rightarrow [0, 1]$$

para cada clase $c \in C$ (ver Amo et al. (2004) [3] y Roubens (1978) [100]). Así, un determinado objeto x puede sustentar determinados grados de pertenencia a más de una clase c , donde el total de sus grados de pertenencia es distribuido a lo largo de todas las clases c , donde $\mu_c(x) > 0$ si el grado de pertenencia de x a una clase c no es nulo. Así, debe cumplirse que:

$$\sum_{c \in C} \mu_c(x) = 1$$

donde $\mu_c(x) \in [0, 1] \forall x, \forall c$.

Así, existe una relación natural entre los conjuntos borrosos y los operadores de agregación de información. La agregación de información supone la simplificación dimensional de la información original sin perder su formato inicial (Montero et al., 2021)[77]. Así, los operadores de agregación (OA) representan unas de las disciplinas más actuales en las ciencias de la información matemática y de sistemas. Originalmente, los OA fueron definidos para la agregación de valores provenientes de funciones de pertenencia asociadas a un conjunto borroso. Por ejemplo, de acuerdo con la clasificación borrosa recién explicada donde un determinado objeto x puede presentar cierto grado de pertenencia a más de una clase, ésta información puede ser agregada para su consideración conjunta en un único valor. Así, la reducción de la cardinalidad permite simplificar la información obtenida, hecho que beneficia, por ejemplo, sistemas de toma de decisiones. Las funciones de agregación presentan diferentes propiedades, como la monoticidad, continuidad, asociatividad, conmutatividad, entre otras (Rojas et al., 2021) [97].

Definición 2.4.3 (Funciones de agregación). Una función $A : [0, 1]^n \rightarrow [0, 1]$ es una función de agregación en n si se cumplen las siguientes condiciones:

- A es creciente en cada argumento, donde para cada $i \in \{1, 2, \dots, n\}$ si $x_i \leq y$, entonces $A(x_1, \dots, x_n) \leq A(x_1, \dots, x_{i-1}, y, x_{i+1}, \dots, x_n)$.
- A cumple con los límites: $A(0, \dots, 0) = 0$ y $A(1, \dots, 1) = 1$.

Otro concepto de gran relevancia dentro de los operadores de agregación y para el desarrollo de esta memoria son las funciones de *overlapping* y *grouping*. El concepto de solapamiento u *overlapping* fue introducido por Buntince et al. (2009) [18] con el fin de medir el grado de solapamiento que presenta un determinado objeto x a dos clases diferentes pertenecientes a un sistema de clasificación difusa. Cabe mencionar que las funciones de *overlapping* fueron ampliadas posteriormente a escenarios con más de dos clases por Gómez et al. (2016) [44]. De esta forma, dados dos grados de pertenencia $\mu_A(x)$ y $\mu_B(x)$ de un objeto x hacia las clases A y B , una función de *overlapping* ofrece el grado z a través del cual el objeto x pertenece a la intersección de ambas clases.

Definición 2.4.4 (Función de solapamiento u *overlapping*). Una función $O : [0, 1]^2 \rightarrow [0, 1]$ es una función de *overlap* si se cumplen las siguientes condiciones:

- O es conmutativa.
- $O(x, y) = 0$ si y solo si $xy = 0$.
- $O(x, y) = 1$ si y solo si $xy = 1$.
- O es creciente en cada argumento.
- O es continua.

Al igual que las funciones de solapamiento, Bustince et al (2009) [18] propusieron las funciones de agrupación o *grouping*, cuyo objetivo es el de medir el grado en el que la combinación de dos clases A y B es soportado. Estas funciones presentan una gran utilidad, habiendo sido usadas, por ejemplo, en el campo de la detección de comunidades [45].

Definición 2.4.5 (Función de agrupamiento o *grouping*). Una función $G : [0, 1]^2 \rightarrow [0, 1]$ es una función de *grouping* si se cumplen las siguientes condiciones:

- G es conmutativa.

- $G(x, y) = 0$ si y solo si $x = y = 0$.
- $G(x, y) = 1$ si y solo si $x = 1$ o $y = 1$.
- G es creciente en cada argumento.
- G es continua.

En adición, uno de los elementos que, de igual manera, presentan gran protagonismo en los conjuntos borrosos, son las medidas borrosas, propuestas por Sugeno (1993).

Definición 2.4.6 (Medida borrosa [109]). Sea V un *set* no vacío. Una medida borrosa es una función $\mu : 2^V \rightarrow [0, 1]$ que cumple con lo siguiente.

- $\mu(\emptyset) = 0$
- $\mu(V) = 1$
- $\mu(A) \leq \mu(B)$, $\forall A, B \subseteq V$ tal que $A \subseteq B$

Así mismo, la lógica difusa ha sido aplicada a múltiples escenarios, como es el caso de los grafos, presentados anteriormente en la sección 2.3. De esta forma, considérese el grafo nítido $G = (V, E)$, cuya matriz de adyacencia es A , que representa las conexiones directas entre los nodos. Más allá del concepto clásico de gráfico nítido, Rosenfeld (1975) [98] introdujo los grafos borrosos basados en las relaciones borrosas entre los individuos. Esta herramienta resulta de gran utilidad para modelar situaciones en las que existe cierta vaguedad o incertidumbre sobre la representación del conocimiento. No obstante, desde un punto de vista matemático, existen algunas situaciones en las que los grafos borrosos pueden entenderse como una especie de grafos ponderados (Moderson y Nair, 2000) [79].

Con la combinación de la capacidad del grafo para modelar conexiones entre elementos y la capacidad de las medidas difusas para manejar la capacidad relacionada con cualquier conjunto de elementos, se definió el grafo difuso extendido (Gutiérrez et al., 2019) [51]. Esta herramienta es un grafo junto con una medida difusa definida sobre el conjunto de nodos. La incorporación de una medida difusa va más allá de la noción anterior de grafos borrosos, que se limitan a considerar únicamente pares de elementos. En este sentido, mediante una medida difusa definida sobre el conjunto de nodos, se puede representar situaciones en las que están implícitos más de dos nodos, independientemente de cómo estén conectados a través del grafo. Es evidente que la capacidad de representación del grafo difuso extendido dista mucho de la de las herramientas existentes, por lo que se pueden abordar situaciones mucho más complejas, con una adecuada modelización de la realidad.

Definición 2.4.7 (Grafo difuso extendido [51]). Sea $G = (V, E)$ un grafo, y sea $\mu : 2^V \rightarrow [0, 1]$ una medida borrosa sobre un conjunto de nodos V . Entonces, un grafo borroso extendido es un triplete $\tilde{G} = (V, E, \mu)$.

2.5. Cadenas de Markov.

Las Cadenas de Markov representan uno de los modelos matemáticos más simples y tradicionales para fenómenos aleatorios que evolucionan en el tiempo [85]. Una de las principales características de las Cadenas de Markov es que permiten modelar fenómenos basados en el tiempo según un enfoque probabilístico, siendo uno de los procesos estocásticos más utilizados. Así, las Cadenas de Markov tienen un carácter no determinista, donde la evolución del proceso depende tanto de variables casuales como aleatorias. En los siguientes párrafos se explican las propiedades básicas de las Cadenas de Markov (Norris, 1998) [85].

Se entiende como cadena a un proceso en el tiempo en el que una variable aleatoria X_n cambia a lo largo del tiempo. X modela un estado aleatorio cuando toma un valor i con probabilidad λ_i . Estas cadenas se pueden considerar en ambos casos: *tiempo discreto*: $n \in \mathbb{Z}^+ = \{0, 1, 2, \dots\}$ o *tiempo continuo*: $t \in \mathbb{R}^+ = [0, \infty)$.

Este trabajo se centra en los procesos de tiempo discreto donde $(X_n)_{n \geq 0}$. Así, sea I el conjunto completo de estados, donde cada estado $i \in I$. Se denota por p_{ij} la probabilidad de transición de un estado i a otro estado j donde $i, j = 1, 2, \dots, m$, siendo $m \in \mathbb{Z}^+$ la cantidad de estados diferentes de I :

$$p_{ij} = P(P_n = j \mid X_{n-1} = i), \quad \text{con } p_{ij} > 0 \text{ y } \sum_{j=1}^m p_{ij} = 1$$

Cuando $p_{ij} > 0$, se dice que un estado i puede comunicarse con otro estado j . Se denomina *matriz de transición* P con dimensiones de $m \times m$ a:

$$P = [p_{ij}]_{i,j \in \{1, \dots, m\}} = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1m} \\ p_{21} & p_{22} & \dots & p_{2m} \\ \dots & \dots & \dots & \dots \\ p_{m1} & p_{m2} & \dots & p_{mm} \end{bmatrix}$$

Para una probabilidad dada de comenzar $\{p_i^{(0)}\}$ en un estado específico i , podemos conocer la probabilidad de llegar a un estado j desde otro estado inicial i en t pasos $p_j^{(t)}: p_j^{(t)} = \sum_{i=1}^m p_i^{(0)} p_{ij}^{(t)}$

Así, $p_{ij}^{(t)}$ se define como la probabilidad de que la cadena alcance el estado

j después de t pasos, cuando la cadena ha comenzado en el estado i . Entonces, la probabilidad de transición en t pasos $p_{ij}^{(t)}$ se define como:

$$p_{ij}^{(t)} = P(X_t = j \mid X_0 = i)$$

Teniendo en cuenta la propiedad de Markov - *la probabilidad de un valor futuro dado de una variable aleatoria solo depende de su valor presente* - se tiene que, para $t \geq 2$ la cadena debería haber pasado por al menos uno de los s estados posibles en la fase $t - 1$, siendo:

$$p_{ij}^{(t)} = \sum_{k=1}^m P(X_t = j, X_{t-1} = k \mid X_0 = i)$$

Aplicando esta lógica a la matriz de transición P , se puede conocer las probabilidades de transición entre todos los estados en t pasos calculando la potencia t de P .

Además, se denomina estado absorbente a un estado i cuando una vez alcanzado la probabilidad de cambiar a otro estado es cero, de modo que $p_{ii} = 1$ y $p_{ij} = 0$, por $i \neq j$ y $j = 1, \dots, s$.

Capítulo 3

Comportamiento de las medidas de polarización

Resumen: En este capítulo se analizan y formalizan las principales medidas de polarización presentadas en el capítulo anterior para entender su comportamiento. Para llevar a cabo esta tarea, se generan determinadas distribuciones. Se hará una distinción entre aquellas medidas que, en primer lugar, necesitan de un conjunto de datos numéricos para poder ser aplicadas. En segundo lugar, se presentan aquellas que necesitan un conjunto de datos categórico y, finalmente, las que necesitan un grafo para su computación. A lo largo de este capítulo se observa el comportamiento, los puntos fuertes y las debilidades de cada una de las medidas de polarización propuestas en el apartado anterior. Esta información resulta de gran utilidad a la hora de proponer una nueva medida de polarización.

3.1. Comportamiento de las medidas de bipolarización tradicionales.

En esta sección se hace una diferencia entre medidas de polarización de acuerdo con el tipo de *input*, o datos, que necesitan para ser computadas. En este sentido, se encuentra una distinción entre aquellas medidas que necesitan de datos continuos para ser computadas y aquellas que necesitan del conocimiento *a priori* de la existencia de grupos. Esta división permite la comparación adecuada entre medidas así como la interpretabilidad de las mismas.

3.1.1. Medidas de bi-polarización basadas en datos continuos.

Entre las medidas presentadas en el capítulo anterior, son dos las que necesitan únicamente un *input* continuo para su aplicación, siendo la propuesta por Foster y Wolfson (1992) [30] definida en 2.2.2 y Esteban y Ray (1994) [28] también definida en 2.4, donde es importante señalar que para la computación de esta medida se utilizó con $\alpha = 1$. Para llevar a cabo la aplicación de las mismas, se crearon diferentes conjuntos de datos, cada uno atendiendo a determinadas distribuciones poblacionales con respecto a la variable con la que se mide la polarización. En tanto que ambas medidas fueron propuestas en el ámbito de la polarización económica en torno a la renta, los conjuntos de datos se han generado de acuerdo con esta idea, estableciendo una renta mínima de 1000 y una renta máxima de 5000. Con ello, en función a la distribución de la población a lo largo de las rentas, se medirá la polarización con ambas medidas con el fin de conocer su comportamiento en profundidad.

Los diferentes conjuntos de datos se generaron de acuerdo con los siguientes criterios:

- **A:** Se generó una muestra aleatoria de tamaño 1000 siguiendo una distribución normal $N(3000, 500)$.
- **B:** Se genera una muestra aleatoria simple de tamaño 1000 de una variable aleatoria Y con la siguiente distribución:

$$Y = \begin{cases} N(\mu = 1000, \sigma = 3000) & \text{con probabilidad } p = \frac{1}{2} \\ N(\mu = 5000, \sigma = 3000) & \text{con probabilidad } p = \frac{1}{2} \end{cases}$$

- **C:** Para este conjunto de datos se generó muestra aleatoria simple de tamaño 10000 de una variable aleatoria Uniforme $U(1000, 5000)$.
- **D:** Se generó un conjunto de datos de 1000 elementos que sigue una distribución normal $N(3000, 5)$.
- **E:** Se generó un conjunto de datos de 1000 elementos que sigue una distribución normal $N(1500, 5)$.
- **F:** Se generó un conjunto de datos de 1000 elementos que sigue una distribución normal $N(4500, 5)$.

A continuación, se observa gráficamente las funciones de densidad para cada uno de los conjuntos de datos descritos en la figura 3.1:

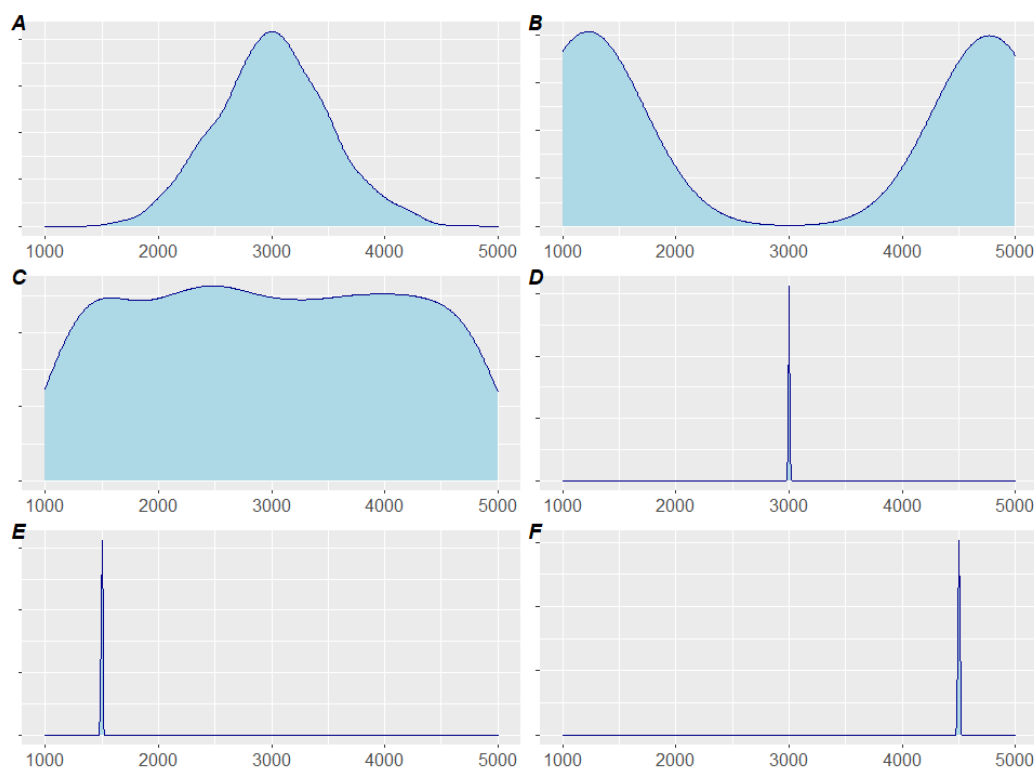


Figura 3.1: Funciones de densidad para conjuntos de datos continuos.

Es importante notar que las medidas de Foster y Wolfson y Esteban y Ray se pueden normalizar entre $[0, 1]$ dividiendo las mismas entre los valores máximos que alcanzan en función del dominio posible de rentas, transformando así la medida propuesta por los autores en un índice que permita la comparación entre ambas.

Nótese que esta normalización no cambia la manera en la que las medidas ordenan los diferentes escenarios de más polarizados a menos, pero permiten comparar los valores brutos de cada medida entre sí. En primer lugar, y como punto de referencia, se computan ambas medidas en sus valores teóricos extremos. Para el caso de polarización máxima teórica de acuerdo con el escenario planteado, equivaldría a 500 elementos situados en la renta más baja, siendo 1000, y otros 500 elementos situados en la renta más alta, siendo 5000. En este escenario ambas medidas deberían mostrar su valor más alto, encontrando valores para $P_{FW} = 2$ y $P_{ER} = 1000$. Por el contrario, el caso de menor polarización se presenta cuando todos los elementos presentan el mismo valor - de acuerdo con las medidas -, siendo $P_{FW} = 0$ y $P_{ER} = 0$. Como el lector puede observar en la tabla 3.1, algunos de los valores ofrecidos por P_{FW} superan los valores máximos teóricos comentados con anterioridad.

Esto implica que los valores máximos de esta medida no se corresponden con los máximos teóricos de polarización. Por su parte, P_{ER} muestra valores extremadamente altos en un escenario “puro” de polarización, por lo que convertir la medida en índice, manejando esta escala de valores, reduciría el resto de puntuaciones a 0. Con ello, y dado que los autores no ofrecen ninguna alternativa, ni los valores máximos y mínimos de sus medidas, sirvan de referencia los mencionados anteriormente para interpretar las medidas.

Como se puede apreciar en la tabla 3.1, ambas medidas muestran sus valores más altos para el caso **B**, escenario que desde un punto de vista teórico es el de mayor polarización, correspondiéndose con la existencia de dos grupos separados de igual tamaño. De acuerdo con ambas medidas, la siguiente distribución con mayor polarización se corresponde con el escenario **C**, correspondiente a una distribución uniforme, seguido por el escenario **A**. Finalmente, y observando los valores, ninguna de las dos medidas distingue entre los escenarios **D**, **E** y **F**. Estos tres últimos escenarios son equivalentes en tanto que siguen una distribución normal con misma desviación típica, encontrando las diferencias entre ellos únicamente en la media poblacional, siendo el escenario **D** equivalente al centro de la variable renta - recuérdese que el rango está entre 1000 y 5000 -, mientras que los escenarios **E** y **F** se corresponden con medias situadas en los extremos, o polos, de la renta con medias 1500 y 4500 respectivamente.

Tabla 3.1: Resultados de las medidas de polarización de Foster y Wolfson (P_{FW}) y Esteban y Ray (P_{ER}) para cada uno de los conjuntos de datos generados.

	A	B	C	D	E	F
P_{FW}	0.347	2.9689	0.8965	0.0033	0.0067	0.0022
P_{ER}	0.5702	3.6904	1.327	0.0056	0.0055	0.0056

En este sentido, se encuentra coherencia entre el comportamiento de ambas medidas, no sólo entre ellas, sino también algunas de las premisas teóricas de la polarización, encontrando los mayores niveles en el escenario **B** y los menores en los escenarios **D**, **E** y **F**.

Sin embargo, uno de los principales problemas que presentan ambas medidas es su interpretabilidad. Por un lado, Foster y Wolfson proponen una medida que, al basarse en el índice de Gini, esta sujeta a determinadas restricciones. De esta forma, cuando la mayor parte de la población se sitúa en torno a cero o, dicho de otra forma, la mediana de las rentas tiende a cero, el valor de la medida tiene a infinito. De esta forma, no es posible la

conversión de esta medida a un índice. Esto supone diversos problemas de interpretabilidad. Además, es importante señalar que la medida de polarización de Foster y Wolfson muestra valores más altos para el escenario C que los teóricamente hablando cuando únicamente hay dos grupos con máxima homogeneidad intragrupal cada uno de ellos con máxima distancia. Esto supone serios problemas de interpretabilidad y de diseño métrico, siendo incoherente el comportamiento de la medida con el concepto de polarización.

Por su parte, la medida Estaban y Ray depende fuertemente del parámetro α , para el cual los autores no proponen una alternativa específica, ni directrices de acuerdo con determinadas situaciones. En este sentido, es necesario un estudio específico para cada caso con el fin de encontrar el valor α más adecuado, pues en situaciones ficticias de máxima polarización la medida presenta diferentes valores, tal es el caso de los escenarios en los que se tiene, por ejemplo $E.G. = \{1, 1, 1, 1, 1, 5, 5, 5, 5, 5\}$ y $\alpha = 1$, $P_{ER} = 1$, mientras que para $E.G. = \{1, 1, 1, 1, 1, 10, 10, 10, 10, 10\}$ y $\alpha = 1$, $P_{ER} = 2.25$. Ya se ha comentado anteriormente, para el caso de máxima polarización del ejemplo computado de la renta, que con $\alpha = 1$, $P_{ER} = 1000$. Estas discrepancias impiden la correcta interpretación de la medida.

Por otro lado, ambas medidas muestran sus valores más bajos en aquellos casos donde toda la población presenta los mismos valores. Sin embargo, no se encuentran diferencias entre la posición de dichos valores a lo largo del eje con el que se mide la polarización. Si bien en términos de distribución los escenarios D , E y F son similares, no ocurre así en términos de interpretación teórica, pues los escenarios E y F presentan distribuciones, como consecuencia del posicionamiento de sus medidas, mucho más alejadas de la distribución teórica de la polarización máxima de lo que presenta el escenario D . Esto puede deberse a la confusión de los términos de polarización y dispersión, ya que lo único que tienen en común estas distribuciones es su dispersión. En este sentido, homogeneizar estos escenarios que en términos teóricos presentan importantes diferencias supone un problema en términos de la medición de la polarización.

3.1.2. Medidas de bi-polarización basadas en datos categóricos.

En este subapartado, se tratarán aquellas medidas de polarización que basan su computación en la existencia *a priori* de grupos definidos en la población, siendo necesario un *input* categórico para la aplicación de tales medidas. Entre ellas, se presentan las propuestas por Montalvo y Rynal-Querol (2003) [76] definida en 2.2.4, Apouey (2007) [4], definida en 2.2.5, Dalton (2008) [21]

definida en 2.2.6 y Permanyer y D'ambrosio (2015) [89] definida en 2.2.7. Para la computación de las medidas recién mencionadas se generarán diferentes distribuciones poblacionales a lo largo de una variable tipo *likert* de 1 a 5. En definitiva, las medidas necesitan la existencia de las categorías definidas por la variable *likert*, así como la frecuencia en la población de cada una de sus categorías.

Para aquellas medidas que requieren de la selección de valores específicos para determinados parámetros, se describen a continuación dicha configuración. Para el caso de la medida de Apouey 2.2.5, con el fin de convertirla en índice y facilitar su interpretación, se estableció la constante $k = 2$ y $\alpha = 1$. Por su parte, la medida de Dalton 2.2.6 requiere de una valoración de cada grupo, o categoría, a lo largo de una variable ideológica de izquierdas a derechas. Así, se entenderán cada una de las categorías de la variable *likert* como partidos políticos, cada uno con una puntuación en la escala ideológica, donde $P = \{1, 2, 3, 4, 5\}$ son las categorías, y $I = \{1, 3, 5, 7, 10\}$ la puntuación en la escala ideológica para las categorías, donde 1 es extrema izquierda y 10 extrema derecha. Finalmente, y para facilitar la interpretación de las medidas, algunas son convertidas a índice, encontrando los valores más altos cuando el 50% de la población se sitúa en la categoría 1 y el otro 50% de la población se sitúa en la categoría 5. Por su parte, los valores más bajos cuando toda la población se presenta en una única categoría. Para ello, se definen los índices de polarización de Dalton y de Permanyer y D'ambrosio a continuación:

Definición 3.1.1 (Índice de polarización de Dalton). Sea P_D la medida de polarización definida en 2.2.6 y sea $vn = \{v_1, v_2, \dots, v_i, \dots, v_p\}$ un vector del número de votos normalizado entre 0 y 1 que han recibido los partidos se tiene que el índice de polarización de Dalton es:

$$IP_D = \frac{P_D}{1.27} \quad (3.1)$$

Donde 1.27 es el valor más alto que ofrece la medida P_D , correspondiente a aquellos casos en los que se dan únicamente dos partidos con mismo número de votos.

Definición 3.1.2 (Índice de polarización de Permanyer y D'ambrosio). Sea P_{PD} la medida de polarización definida en 2.2.7 se tiene que el índice de polarización de Permanyer y D'ambrosio es:

$$IP_{PD} = \frac{P_{PD}}{0.25} \quad (3.2)$$

Donde 0.25 es el valor más alto que ofrece la medida P_{PD} , correspondiente a aquellos casos en los que se dan únicamente dos grupos de mismo tamaño.

Se generaron diferentes conjuntos de datos a lo largo de una variable tipo *likert* de 1 a 5 con una muestra de $N = 1000$ de tal manera que se asemejen lo máximo posible a los generados en el subapartado anterior, mostrando así diferentes distribuciones de interés en el ámbito de la polarización.

- **A:** Se generó una muestra donde el 40 % de la población se sitúa en la categoría central de la variable 3 y se distribuye, equitativamente, al resto de las categorías adyacentes, encontrando un 20 % de la población en las categorías 2 y 4 respectivamente y un 10 % en cada una de las categorías de los extremos 1 y 5.
- **B:** Este conjunto de datos se caracteriza por una población agrupada en torno a los valores extremos, encontrando un 49.7 % de la población en el valor extremos 1, un 49.7 % de la población en el valor extremos 5. y el resto de la población distribuida equitativamente a lo largo del resto de las categorías.
- **C:** Para este conjunto de datos se distribuyeron los 1000 elementos a lo largo de todas las categorías de forma equitativa, siguiendo una distribución uniforme, presentando una acumulación del 20 % de la población en cada valor.
- **D:** Se generó un conjunto de datos donde el 99.6 % de la población se agrupa en torno a la categoría central de la variable 3, y el 0.04 % de la población restante repartida equitativamente a lo largo del resto de categorías.
- **E:** Se generó un conjunto de datos donde el 99.6 % de la población se agrupa en torno a la categoría extrema de la variable 1, y el 0.04 % de la población restante repartida equitativamente a lo largo del resto de categorías.
- **F:** Se generó un conjunto de datos donde el 99.6 % de la población se agrupa en torno a la categoría extrema de la variable 5, y el 0.04 % de la población restante repartida equitativamente a lo largo del resto de categorías.

En la figura 3.2 se observa gráficamente las diferentes distribuciones de cada uno de los conjuntos de datos descritos.

En los resultados observados - ver tabla 3.2 - se encuentra mayor heterogeneidad con respecto a la encontrada entre las medidas de polarización

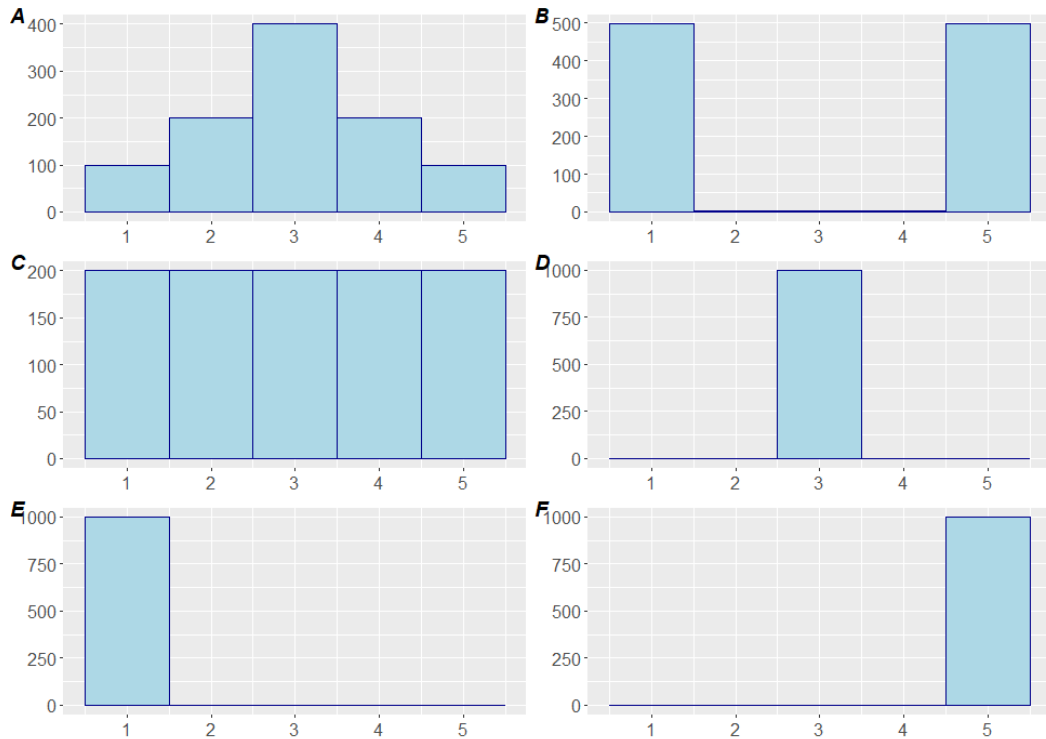


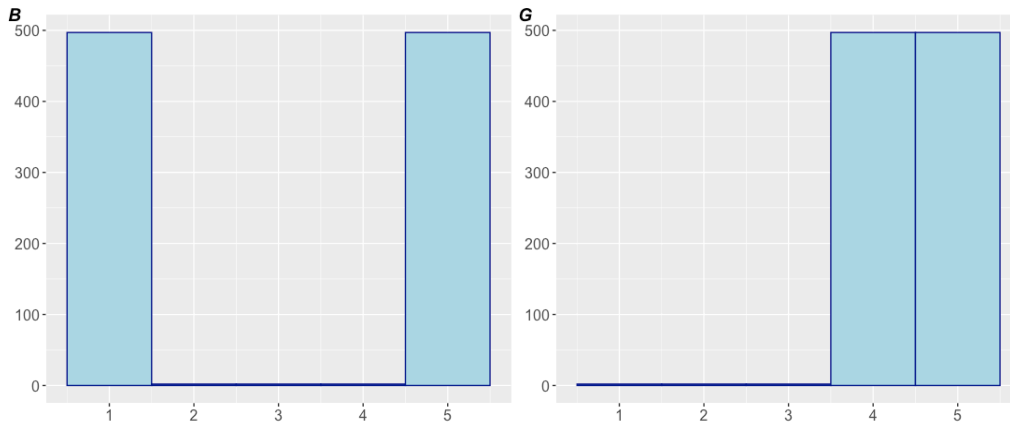
Figura 3.2: Conjuntos de datos continuos.

para variables continuas. Si bien hay consenso entre todas las medidas en mostrar sus valores más altos en el escenario **B**, así como los más bajos para el caso **D** y valores medios altos para el caso **C**, se encuentran diferencias en el resto de soluciones. Para los escenarios **E** y **F**, todas las medidas muestran valores similares al del caso **D** salvo el índice de polarización de Dalton, mostrando valores medios-altos de polarización, mientras que el resto de medidas muestran sus valores más bajos. En referencia al escenario **A**, las medidas P_{M-RQ} y IP_{PD} muestran valores de polarización mayores con respecto al caso **C**, mientras que el resto ofrecen valores más bajos. Así mismo, dentro del propio escenario **A**, de nuevo las medidas P_{M-RQ} y IP_{PD} muestran valores medios-altos, mientras que P_A e IP_D valores medios-bajos.

En adición, es de especial interés comprar de forma separada el escenario **B** con un nuevo escenario no incluido anteriormente, correspondiente al caso **B** con la diferencia de que uno de los grupos, no será el opuesto al contrario, sino adyacente al mismo, encontrando la mitad de la población, por ejemplo, en la categoría 4 y la otra mitad en la categoría 5 - ver figura 3.3 -.

Tabla 3.2: Resultados de las medidas de polarización categóricas para cada uno de los conjuntos de datos generados.

	A	B	C	D	E	F
P_{MR-Q}	0.712	0.994	0.64	0.0158	0.0158	0.0158
P_A	0.4	0.996	0.6	0.003	0.005	0.005
IP_D	0.26	1	0.777	0.0314	0.6614	0.7559
IP_{PD}	0.712	0.994	0.64	0.0158	0.0158	0.0158
IOV	0.12	0.0033	0.14	0.0024	0.0074	0.0074

Figura 3.3: Comparación de escenarios B y G

Si bien ambos escenarios representan el 49.7% de la población situada en una categoría, y el otro 49.7% en otro grupo, ambos escenarios son significativamente diferentes. Se observa como en el escenario B estos grupos presentan polos opuestos, mientras que en el escenario G ambos grupos son adyacentes. Esta diferencia es de especial interés, pues ambos escenarios son sinónimos para una variable categórica nominal, mientras que por el contrario, muy distintos en una variable categórica ordinal. La comparación entre ambos escenarios y los resultados para las medidas de polarización aporta información de gran interés para conocer la naturaleza de las medidas.

Como se puede observar en la tabla de resultados 3.3, las medidas P_{M-RQ} e IP_{PD} muestran los mismos valores para ambos escenarios, mientras que IP_D son muy similares. Por el contrario, únicamente la medida de polarización P_A muestra una clara diferencia entre ambos escenarios, indicando valores bajos de polarización para el caso G .

Como se ha podido observar en los resultados anteriormente expuestos, si bien algunas medidas parecen mostrar sus resultados en la misma dirección, no se encuentra la misma tendencia para todas las medidas de polarización.

Tabla 3.3: Resultados de las medidas de polarización categóricas para los escenarios **B** y **G**.

	B	G
P_{MR-Q}	0.994	0.994
P_A	0.996	0.2545
IP_D	1	0.807
IP_{PD}	0.994	0.994

Estas diferencias son sustanciales, ya que dado un determinado escenario, en función a la medida consultada la interpretación es significativamente diferente. Además, como se ha podido comprobar con la comparación de los escenarios **B** y **G**, no todas las medidas parecen reflejar la misma naturaleza. En concreto, a pesar de que todas ellas presentan como factor común el uso de datos categóricos denominados grupos, asumidos *a priori*, muestran comportamientos muy distintos en esta comparación. De aquí se puede concluir la división entre medidas de polarización nominales y ordinales. Con respecto a las medidas de polarización nominales cabe concluir que la posición de los grupos en la escala tipo *likert* deja de tener sentido ordinal, mostrando resultados iguales, o similares, ante los escenarios **B** y **G**. Esto se puede considerar como una deficiencia notoria, pues ante el mismo tipo de información, la medida no recoge el matiz del posicionamiento grupal de la población a lo largo de la variable, sino únicamente la formación de tales grupos. Por el contrario, de entre todas las medidas analizadas, la propuesta por Apouey - P_A - sí que recoge esta información, categorizando esta medida como una medida de polarización ordinal.

En adición, el lector ha podido observar que los valores de polarización referentes a las medidas P_{MR-Q} y P_{PD} son exactamente los mismos para todos os escenarios y casos. Este hecho tiene su justificación en lo explicado anteriormente en el apartado 2.2, donde se explica que, en realidad, ambas medidas son equivalentes.

3.2. Comportamiento de las medidas de polarización basadas en grafos.

A diferencia de las medidas de polarización presentadas en el apartado anterior, que miden la polarización desde un punto de vista actitudinal o ideológico, cabe rescatar el concepto de polarización estructural. En esta sección, se presentarán diferentes estructuras de grafos con el fin de aplicar las medi-

das de polarización basadas en redes para medir los niveles de polarización estructural presente a tales grafos. Con ello, se observará el comportamiento de estas medidas a diferentes estructuras de nodos.

Para poner a prueba las medidas de polarización basadas en redes, se generan diferentes grafos a través de los cuales se representan diferentes estructuras de relaciones entre los nodos o vértices. Además, se aplican algoritmos de detección de comunidades para asociar cada nodo a su grupo y permitir la computación de la medida propuesta por Guerra et al. (2013) [49]. En tanto que la medida recién citada requiere de la consideración de dos comunidades diferentes, para aquellos casos en los que la fase de detección de comunidades ofrece > 2 , se restringió el número de comunidades a 2. El algoritmo de detección de comunidades aplicado es la función *fastgreedy.community*, incluida en el paquete estadístico *igraph* para R. Todas las redes creadas se componen de un total de 20 nodos a razón de simplicidad gráfica. Así, se presentan a continuación los diferentes grafos conexos creados y sus características.

- **A:** Se generaron un conjunto de nodos de forma artificial que formasen dos cadenas circulares unidas entre sí únicamente por una arista entre dos nodos. Para este grafo generado de forma manual, no se aplicó el algoritmo *fastgreedy.community()* sino que se establecieron las comunidades de forma manual para fines ilustrativos.
- **B:** Se generó un grafo aleatorio según el modelo Erdos-Renyi con la función *erdos.renyi_game()*, indicando una probabilidad aleatoria de arista entre dos nodos aleatorios generada a partir de una distribución uniforme entre 0 y 1.
- **C:** Para este caso, se propone un grafo aleatorio según el modelo de Barabasi con la función *barabasi_game()*. Inicialmente, se encontraron un total de 5 comunidades.
- **D:** Se creó una red aleatoria de pequeño mundo siguiendo el modelo de Watts-Strogatz con la función *sample_smallworld()*, con los parámetros 1 para la dimensión de la red inicial, 5 en referencia al vecindario dentro del cual se conectarán los vértices de la red y 0.05 para la probabilidad de reconexión entre nodos.
- **E:** Para este caso se generan grafos aleatorios e independientes siguiendo el modelo de Erdos-Renyi que son unidos entre sí por un número específico de aristas, usando la función *sample_islands()*. Para ello, se especificó un parámetro de 2 para el número de islas, con 10 nodos cada una de ellas, con una probabilidad de 0.6 de crear una arista dentro de cada isla y un total de 10 aristas entre islas.

- **F**: Se generaron dos grafos aleatorios según el modelo Erdos-Renyi con la función `erdos.renyi_game()` de 10 nodos cada uno, indicando una probabilidad de arista entre dos nodos aleatoria generada a partir de una distribución uniforme entre 0 y 1 para ambos. Posteriormente, se unieron entre sí a través de una arista entre nodos de ambos grafos.
- **G**: Finalmente, se generó un grafo compuesto por un grafo aleatorio según el modelo Erdos-Renyi con la función `erdos.renyi_game()` de 10 nodos y un grafo con forma de cadena circular unidos entre sí por una única arista.

Una vez generados los grafos mencionados anteriormente, así como sus comunidades, se representan en la figura 3.4.

Así, de acuerdo con los escenarios mostrados anteriormente 3.4, se aplican las medidas de polarización de modularidad (Newman, 2006) [81] y basada en la frontera (Guerra et al., 2013) [49], encontrando los resultados de la tabla 3.4. Como se puede observar, se encuentran diferencias notorias entre ambas medidas. En primer lugar, los escenarios con mayor modularidad se dan para el caso **A**, **C** y **F**, mientras que los más bajos para el caso **B** y **D**. Por su parte, la medida de polarización basada en la frontera categoriza el escenario **F** como el más polarizado, alcanzando casi su nivel máximo - 0.5 -, seguido por el escenario **G**, mientras que el escenario **A** y **C** presentan los mismos niveles de polarización, y encontrando valores negativos para los casos **B** y **D**.

Tabla 3.4: Resultados de las medidas de polarización basadas en redes.

	A	B	C	D	E	F	G
Modularidad	0.4523	-0,004	0,447	0,155	0,3676	0,4871	0,3073
P_{GMCK}	0.1666	-0.0277	0.1666	-0.1470	0.1065	0,4	0,2833

En primer lugar, es interesante mencionar que, a pesar de los resultados encontrados, son diferentes escenarios los que recogen el aspecto teórico de máxima polarización, siendo el mismo “la división de la sociedad en dos grupos opuestos de igual tamaño”, siendo el caso de todos los escenarios menos de **B**. Como consecuencia de los resultados mostrados por ambas medidas, una de las primeras conclusiones que se pueden extraer es que ambas medidas no son sensibles al tamaño de los grupos. Por su parte, la modularidad es una medida que ofrece un valor referente a la calidad de las particiones obtenidas del grafo, fuertemente ligada a la densidad de aristas que presentan las comunidades dentro de sí mismas. Si bien este concepto puede tener relación con

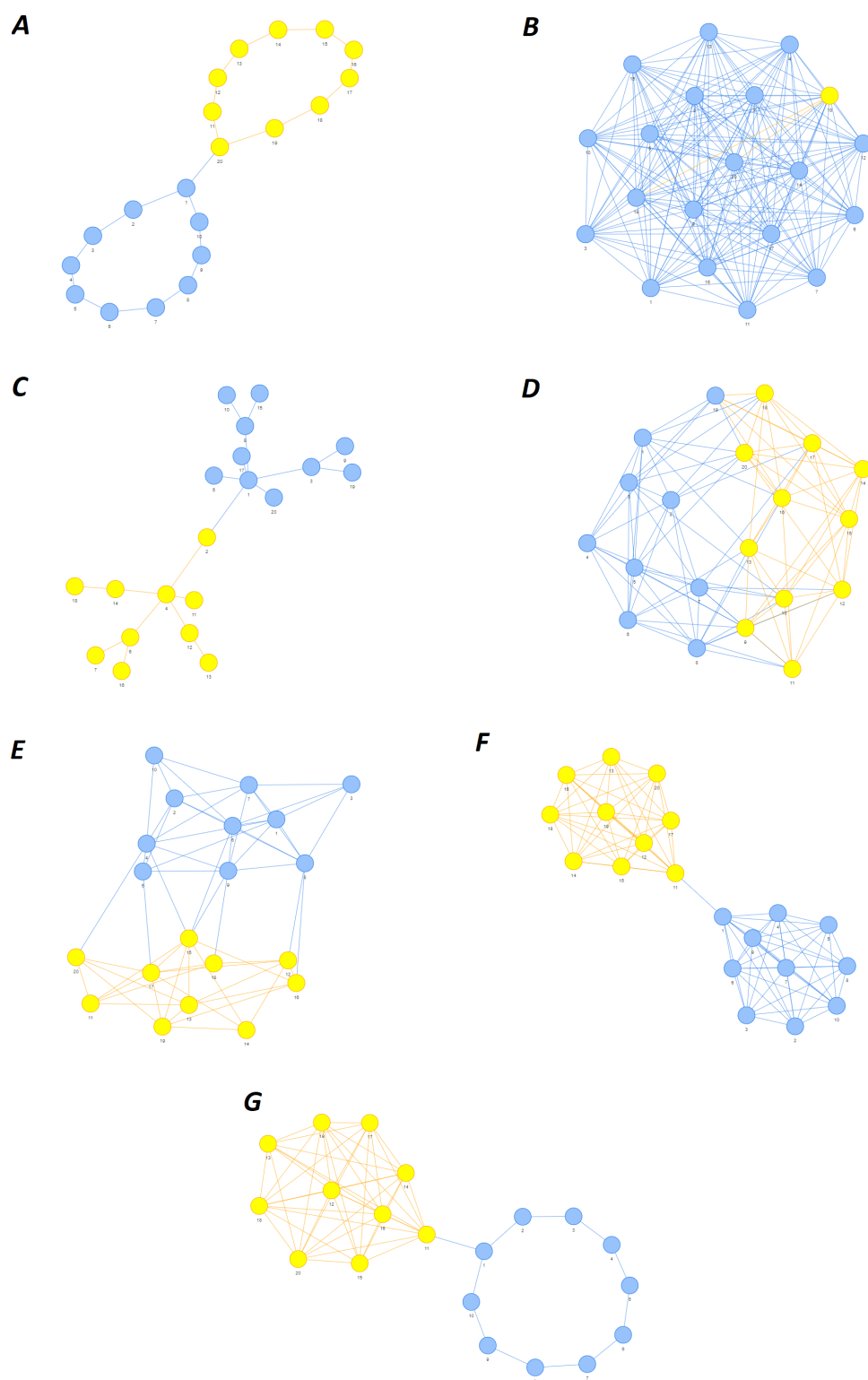


Figura 3.4: Conjuntos grafos generados.

el de polarización, como referencia a la alta homogeneidad intragrupal, parece insuficiente para la medición de la polarización. Escenarios en los que se encuentran claras divisiones entre los grupos detectados, pero que presentan un gran número de aristas en la frontera, derivan en niveles de modularidad más bajos. Así, en un caso real donde existe un alto número de aristas entre comunidades como consecuencia de la comunicación entre ambas comunidades, cuya naturaleza es hostil, la medida de polarización de modularidad mostraría valores irreales. Por otro lado, la medida de polarización fronteriza presenta el problema de tener únicamente en cuenta los nodos fronterizos, obviando el resto de la red en la medición de la polarización. Esto supone un problema doble pues, como se ha comentado anteriormente, presenta insensibilidad ante el tamaño de los grupos, cuestión que resulta ser clave para la polarización. Además, es importante mencionar que los valores de la medida siempre serán sensibles al algoritmo de detección de comunidades empleado. Por último, dadas las características de la medida, es condición necesaria la presencia de únicamente dos grupos diferentes para su computación.

Como se ha mencionado en el capítulo anterior, sí existen otras medidas de polarización en redes en la literatura que tienen en cuenta el tamaño de los grupos, como es el caso de la propuesta por Morales et al. (2015)[78]. Sin embargo, esta medida requiere de cierta complejidad computacional, así como de la necesidad de información extra al grafo para poder ser computada. Por su parte, y dada la naturaleza de un grafo al que se le han aplicado algoritmos de detección de comunidades, medidas de polarización categóricas podrían ser aplicables en estos escenarios, usando como variable nominal las comunidades a las que pertenece cada nodo, en tanto que muchas de ellas únicamente necesitan un conjunto de datos que incluya los grupos existentes y sus frecuencias relativas en la población. Esta opción, aunque de gran interés, aún no se ha visto aplicada en la literatura.

Finalmente, a lo largo de este capítulo se han analizado diferentes medidas de polarización para determinados escenarios. Cabe mencionar que se encuentran algunas deficiencias métricas en las diferentes propuestas, ya sea en el caso de las medidas diseñadas para datos continuos, categóricos o grafos. Todo ello sirve de motivación para proponer una medida de polarización que aborde alguna de estas deficiencias. Para ello, en los próximos capítulos se aborda esta tarea. En primer lugar, a lo largo capítulo 4 se analizan escenarios de polarización con el fin de conocer en profundidad las características de la polarización y el comportamiento de la población con la aparición de este fenómeno. Finalmente, en el capítulo 5 se presenta una medida de polarización basada en los conjuntos borrosos.

Capítulo 4

Identificando y Analizando la Polarización

Resumen: En el presente capítulo, previo a la medición de la polarización, se presentan algunos estudios llevados a cabo con el objetivo de ampliar el conocimiento que se tiene del fenómeno de la polarización en las redes sociales. Así, estas aportaciones se centran en el estudio de la detección e identificación de algunos agentes polarizantes en la red, promotores del origen de la polarización en un entorno digital, como lo son los *bots* políticos en las redes sociales online. Así mismo, también se presentan trabajos en los que se detecta la polarización, y se estudia bajo qué escenarios aparece dicho fenómeno así como las posibles consecuencias de la misma desde un punto de vista comunicacional, así como del comportamiento social. Se considera importante esta primera fase previa a la medición de la polarización con el fin de conocer de forma más profunda cuáles son los posibles orígenes de la misma en entornos online así como de las consecuencias de esta en el comportamiento social digital.

4.1. La presencia de los *bots* políticos en Twitter durante la crisis del COVID-19 en España.

Martínez, A., Guevara, J.A., Jiménez, A. (Aceptado, 2023). La presencia de los bots políticos en Twitter durante la crisis del COVID-19 en España. *Reis. Revista Española de Investigaciones Sociológicas*.

Resumen.

La crisis del COVID-19 conllevó que el gobierno tomase medidas excepcionales como el decreto del Estado de Alarma y un confinamiento domiciliario. En consecuencia, las redes sociales se utilizaron más aún como herramientas de debate político. Sin embargo, como sucedió con otros eventos políticos, el diálogo se ve interferido por herramientas manipuladoras de la opinión pública. En esta investigación se ha detectado mediante un análisis *Big Data* en Twitter, una cantidad significativa de *bots* que operaron durante el Estado de Alarma. Destacan por usar lenguaje incívico y negativo. Consecuentemente, surge una división del debate en torno a la posición política. Se concluye que, a pesar del potencial democrático ofrecido por las TIC, éste se ve coartado por la resistencia de las élites a mantener un diálogo vertical.

Abstract.

The COVID-19 crisis prompted the government to take exceptional measures such as declaring a state of alarm or a severe home confinement. In consequence, social networks were used to a greater extent as tools for political deliberation. However, as has happened with other relevant political events, the dialogue in social networks has been interfered by artificial tools that manipulate public opinion. In this investigation, a significant number of bots that operated during the State of Alarm have been detected through Big Data analysis on Twitter. They stand out for using uncivil and negative language. Consequently, a division of the debate arises around the political position. It is concluded that, despite the democratic potential offered by ICT, it is constrained by the resistance of the elites to maintain a vertical dialogue.

4.1.1. Redes sociales y *bots*.

A lo largo de los últimos años las redes sociales online, y en especial Twitter, han supuesto un cambio fundamental en las vidas diarias de las personas desde una perspectiva de la información y la comunicación. Esta nueva situación ha sido ampliamente aprovechada por la esfera política que, a través de nuevas estrategias de marketing político aprovechan las oportunidades que ofrecen los entornos de comunicación online, hacen uso de herramientas de inteligencia artificial (IA), orientadas específicamente a la alteración y distorsión del impacto de determinados mensajes políticos.

Con la imposición del estado de alarma de marzo a abril de 2020 por parte del gobierno de España por el coronavirus, junto con el período de

desescalada que se prolongó hasta finales de junio del mismo año, los medios sociales cobraron gran protagonismo. Escenarios similares al citado han sido aprovechados por partidos políticos a lo largo de la historia con el fin de cumplimentar sus propios objetivos partidistas (Howard et al., 2016 [59]; Bradshaw y Howard, 2018 [15]). Además, a este hecho se suma el creciente uso de *bots* en redes sociales online con el objetivo de influenciar a los usuarios de los espacios públicos digitales (Yan et al., 2021 [121]; Bradshaw y Howard, 2019 [16]).

Así, Twitter se ha convertido en un espacio de aprovechamiento para estas estrategias donde los usuarios presentan especial vulnerabilidad a sus influencias. La estructura de las redes sociales facilita la aparición de polarización entre grupos enfrentados (Neyazi, 2020) [83] que, sumados a los momentos de gran tensión provocados por la presencia de polarización, aparece el uso de lenguaje denominado *incivility* (Papacharissi, 2004) [87]. El uso de estas herramientas conlleva el planteamiento de la hipótesis de que durante el estado de alarma en España por covid los principales partidos políticos de España utilizaron algoritmos de IA en Twitter para condicionar la conversación política de los usuarios. Así, en este estudio se pretendió conocer si se iniciaron campañas de desinformación lideradas por *bots* organizados y vinculados a agentes sociopolíticos tradicionales con el fin de generar falsas ideas de apoyo o desaprobación sobre el debate digital y político sucedido en Twitter durante la crisis sociosanitaria vivida a lo largo del estado de alarma y las fases de desescalada. La confirmación de esta hipótesis puede suponer la detección de uno de los detonantes más directos de la polarización en entornos digitales, cuyo origen estaría sujeto a las estrategias generadas por partidos políticos cuyo fin es el de crear distorsión en una situación de crisis para su propio beneficio.

4.1.2. Metodología.

4.1.2.1. Hipótesis y objetivos de investigación.

- Durante el periodo del confinamiento domiciliario y las fases de la desescalada en España se utilizaron algoritmos de IA en Twitter para condicionar la conversación política.
 - HS1: Se crearon campañas de desinformación ejecutadas por grupos de bots organizados estratégicamente y vinculados a los agentes sociopolíticos tradicionales cuyo objetivo es el de generar una falsa sensación apoyo o desaprobación de los temas políticos discutidos durante el periodo señalado.

Para abordar las hipótesis de investigación anteriormente expuestas, se proponen los siguientes objetivos de investigación:

- Conocer la presencia de *bots* políticos en la red social Twitter durante la crisis del coronavirus en España.
- Definir el papel de los *bots* en la red social, así como su relación con los principales partidos políticos de España.
- Describir cuáles podrían ser las posibles consecuencias de la presencia de *bots* en el debate público online que tiene lugar en Twitter durante el periodo estudiado.

4.1.2.2. Caso de estudio y fuentes de datos.

A finales del año 2019 y principios del 2020, la COVID-19 llegaba a España para causar, posteriormente, la mayor crisis sociosanitaria del país en mucho tiempo. Como consecuencia, el 15 de marzo se instauró el estado de alarma por parte del gobierno central que conllevó un confinamiento severo acompañado por una desescalada del mismo por fases que duró hasta el 21 de junio de 2020. Bajo esta situación, se sitúa un recién formado gobierno de coalición en España que se enfrentó a la situación es crisis social, económica y sanitaria que acompañó la introducción del virus en el país. Como consecuencia, el debate público incendió las redes sociales online, las cuales tomaron un papel protagonista en la vida diaria de las personas como método de información y comunicación. Como consecuencia de esta crisis, las redes sociales se convirtieron en un lugar de debate entorno a la gestión política de un nuevo gobierno de coalición sobre dicha situación.

En concreto, el caso de estudio de la presente investigación se enfoca en las publicaciones generadas en la red social Twitter cuyo contenido se refiere a la gestión social, económica y sanitaria del gobierno en la pandemia de COVID-19. Los datos se obtuvieron a través de la API de Twitter usada por el paquete estadístico *rtweet* (Kearney, 2019)[63]. Debido a las limitaciones de la API de twitter, y con el fin de abordar todo el proceso, se realizaron cinco tandas de descarga, cada una de ellas de una semana de duración, a lo largo de todo el confinamiento:

- 1 tanda de descarga: "2020-03-16 "2020-03-23".
- 2 tanda de descarga: "2020-04-14 "2020-04-21".
- 3 tanda de descarga: "2020-05-15 "2020-05-22".

- 4 tanda de descarga: "2020-06-08 "2020-06-15".
- 5 tanda de descarga: "2020-06-21 "2020-06-29".

Los criterios de descarga de publicaciones se corresponden con una bolsa de palabras que permite acceder a aquellos mensajes que pertenecen al objetivo de estudio: “gobierno”, “España” y “estadodealarma”, las cuentas de los principales partidos políticos - PSOE, PP, PODEMOS, CIUDADANOS y VOX -, así como los principales representantes de cada uno de los partidos. Una vez finalizado el proceso de descarga, se obtuvo una base de datos con 4895747 tweets.

4.1.2.3. Limpieza de datos.

Debido a las características de la descarga de publicaciones, así como de las coincidencias que presenta el idioma español y, en concreto, la temática de estudio, en cuanto a dobles significados tales como el partido *podemos* - confundido con el verbo poder -, o el partido político *ciudadanos* - confundido con los habitantes de un país -, se considera de vital importancia filtrar la base de datos, eliminando así aquellos mensajes que no pertenezcan al caso de estudio. Para ello, se hizo uso de algoritmos de aprendizaje automático supervisado con el fin de, tras una fase previa de codificación manual y entrenamiento, puedan extraerse todos aquellos mensajes no deseados de la base de datos. Por otro lado, es importante resaltar que, debido al alto grado de dinamismo que presentó el debate digital en Twitter a lo largo de todo el caso de estudio, el contenido de las publicaciones entre tantas puede variar sustancialmente. Por consiguiente, se considera primordial el entrenamiento de los modelos de aprendizaje automático tantas veces como tandas se hayan descargado, con el fin de asegurar la correcta generalización del conocimiento implícito al contenido por parte de los algoritmos. Para ello, se realizó una muestra aleatoria simple de 1500 tweets por cada tanda.

Para permitir la aplicación de algoritmos de aprendizaje automático a datos textuales, resulta fundamental transformar los datos a una estructura adecuada para su procesamiento por los modelos de aprendizaje automático. En primer lugar, se realiza la fase denominada como *tokenización*. Esta primera fase separa un texto en todas y cada una de las diferentes palabras que presenta, convirtiendo cada palabra en una unidad. Posteriormente se aplica la fase de eliminación de las *stopwords*, removiendo así palabras que carecen de significado como determinantes o artículos. Finalmente, los datos se transforman a matrices de frecuencia-término, denominadas *dfm*. En especial, se usaron matrices *tf-idf*, obteniendo una estructura de datos donde

cada fila representa un determinado mensaje y cada columna de la matriz representa cada una de todas las palabras que aparecen en el conjunto de datos. Finalmente, se aplicaron algoritmos de aprendizaje automático usando la librería *caret* [67] y *e1071* [74] en R. Si bien se aplicaron diferentes algoritmos de aprendizaje automático para la realización de esta tarea, fueron las máquinas de soporte vectorial las que presentaron un mejor rendimiento - ver 4.1 -.

Tabla 4.1: Resultados del clasificador SVM-lineal para limpieza de datos.

Tanda	Precisión	Sensibilidad	Especificidad	Kappa	F-Score	AUC
1	0.8017	0.9322	0.3846	0.3670	0.8620	0.6583
2	0.8167	0.5476	0.9213	0.5077	0.6556	0.7344
3	0.8267	0.7027	0.8995	0.6187	0.7596	0.8010
4	0.7867	0.7090	0.8494	0.564	0.7457	0.7791
5	0.7659	0.8758	0.6377	0.5216	0.8171	0.7567

4.1.2.4. Detección de *bots* y su relación con los partidos políticos españoles.

Para la identificación de *bots*, o cuentas automatizadas, en nuestra base de datos se aplicó el algoritmo propuesto por (kearney, 2020) [64] presente en la librería de R *tweetbotornot2*. Este algoritmo se basa en la detección del nombre de usuario de una determinada cuenta y, en función de múltiples variables, como biografía, actividad o publicaciones, entre otras, asocia una probabilidad a que dicha cuenta sea automatizada. Como consecuencia del alto número de cuentas disponibles en la base de datos (445230), se procedió a la aplicación del algoritmo de detección de *bots* llamado *FAST - gradient boosted*, el cual proporciona una mayor velocidad de rendimiento. En tanto que el *output* del algoritmo es una probabilidad asociada a cada usuario, y con el objetivo de mantener un criterio conservar con el fin de evitar las falsas alarmas en el proceso de detección, se decidió categorizar como *bots* únicamente aquellas cuentas que presentan una probabilidad de ser automatizadas situadas por encima del cuartil más alto de la probabilidad, siendo $P > 0.9754$.

Así mismo, para cubrir los objetivos de este estudio, es necesario la asociación de cada *bot* detectado a un determinado partido político. Esta asignación se realizó de acuerdo con el estilo comunicacional que presentan los congresistas de cada uno de los partidos políticos y la similitud encontrada en las publicaciones de una determinada cuenta automatizada. Esta tarea se llevó a través de la aplicación de algoritmos de aprendizaje automático supervisado,

entrenando dichos algoritmos a detectar y discriminar cuales son los estilos de comunicación de los principales partidos de España. Se creó una nueva categoría llamada “otro” en el que se incluyeron el resto de partidos políticos. Así, se comparó el uso del lenguaje de las publicaciones escritas por *bots* con los diferentes estilos comunicacionales de dichos partidos, asignando un *bot* a un determinado partido político con el que encaje su estilo comunicacional.

Para la fase de entrenamiento, se seleccionaron todas las publicaciones en la base de datos escritas por los 350 representantes del Congreso de los Diputados. De nuevo, las máquinas de soporte vectorial fueron aquellas que presentaron un mejor rendimiento, encontrando resultados satisfactorios que permitieron el entrenamiento, y aprendizaje, correcto de los estilos de comunicación para cada uno de los partidos políticos por parte de los algoritmos - ver tabla 4.2 -.

Tabla 4.2: Resultados del clasificador SVM para la variable Partido político.

Variable	Precisión	Sensibilidad	Especificidad	Kappa	F-Score
GENERAL	0.9683			0.9557	0.9632
PODEMOS		0.9490	0.9939		
PSOE		0.9828	0.9794		
CIUDADANOS		0.8591	0.9974		
PP		0.9472	0.9953		
VOX		0.9887	0.9935		
OTRO		0.9587	0.9964		

4.1.2.5. Análisis de sentimiento.

Otra de las metodologías necesarias para cubrir los objetivos de este estudio es el de la aplicación de análisis de sentimiento. Entre los diferentes diccionarios existentes, se decidió aplicar el denominado *afinn*, compuesto por un total de 2477 palabras, las cuales presentan una puntuación de -5 a 5 representando la valencia emotiva. De esta forma, a menor es la puntuación, mayor negatividad emotiva presenta la palabra, mientras que a mayor es el número, existe una mayor positividad asociada al término. Dada la naturaleza del caso de estudio, se incluyó el término “coronavirus”, asociado a un valor de -5 , dadas las connotaciones negativas que se asocian al mismo a lo largo de la pandemia, ya sea para el marco sanitario, económico, social o político.

4.1.2.6. Análisis de redes y detección de comunidades.

En este estudio se ha planteado un análisis de redes basado en una red de retweets adyacente a la base de datos descrita anteriormente. Para facilitar la visualización de dicha red se usó el algoritmo *force atlas 2* presente en el software Gephi (Bastian et al., 2009) [62]. Por su parte, en cuanto a la detección de comunidades se aplicó uno de los algoritmos más usados en la literatura conocido como Louvain (Blondel et al. (2008) [12]. Así, se construyeron dos redes principales.

En primer lugar, se construyó una red compuesta por aquellos usuarios no *bots* con el fin de conocer la estructura y las comunidades de la misma. Esta red se compuso de un total de 194992 nodos - o usuarios que hicieron retweet - y 419130 aristas o relaciones entre nodos.

Por otro lado, se construyó una segunda red centrada en los usuarios *bots* así como los usuarios que interaccionaron con éstos, componiéndose el grafo de un total de 79246 nodos y 118269 aristas. Sin embargo, dada la naturaleza de los objetivos de este estudio, se representó la red con información extra, siendo la misma la asociación definida por los algoritmos de aprendizaje automático presentada en el apartado anterior donde se asocian mensajes a partidos políticos. Así, se asocia un determinado nodo a un partido político específico. Para ello, y dado que son los mensajes el *output* de los clasificadores, se detectaron para cada nodo todos sus mensajes y los partidos a los que se les asocia. A continuación, se computó la moda, y se asoció dicho partido a un determinado nodo o usuario. Por su parte, se hizo lo mismo para aquellos que, en vez de generar el mensaje retweeteado, hicieron la acción de retweetear sobre el mismo, seleccionando todos sus retweets y calculando la moda de los partidos políticos asociados a los mismos.

Por último, y para mejorar la representación visual de grafo, se excluyeron todos los nodos que presentaron un rango menor a 8, simplificando así la red.

4.1.3. Resultados.

En primer lugar, se llevó a cabo la aplicación de algoritmos de detección de *bots*, los cuales proporcionaron la información de que el 19.20% de todos los usuarios que participaron en el debate fueron identificados como *bots*. Por su parte, estas cuentas automatizadas generaron un total del 12.91% de mensajes que compusieron el debate digital en su totalidad. Analizando, a nivel general, el estilo de comunicación en la red teniendo en cuenta la totalidad de los usuarios, cabe concluir que fue el retweet, encontrando que el 80.84% de todos los mensajes publicados fueron retweets.

Una vez que se ha llevado a cabo el proceso de detección de cuentas

automatizadas, se procedió a la aplicación de los algoritmos de aprendizaje automático para codificar, cada uno de los mensajes producidos por dichas cuentas, de acuerdo a su similitud con el estilo comunicacional de los principales partidos políticos de España en dicho momento. Con ello, tal y como se puede observar en la siguiente figura 4.1, se conocen el número de mensajes que presentan contenido textual similar al de los diferentes partidos políticos, encontrando los niveles más altos asociados al PSOE, seguidos por el PP y VOX.

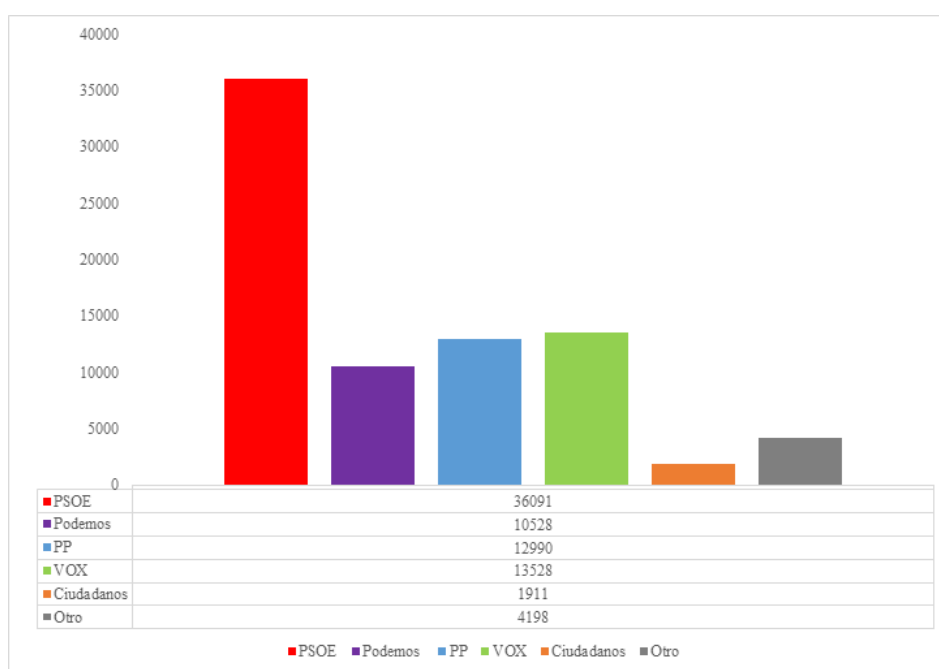


Figura 4.1: Mensajes asociados a estilos comunicacionales de los principales partidos políticos de España producidos por *bots*.

A continuación, se aplicaron diccionarios de sentimiento con el fin de conocer el lenguaje incívico y hostil empleado por las cuentas automatizadas con el fin de romper el debate entre los usuarios de la red social. Como bien refleja la figura 4.2, se encuentran proporciones de palabras asociadas a sentimientos negativos superiores a la de positivos. Además, se encuentran diferencias entre estos porcentajes a lo largo de los estilos comunicacionales de los partidos políticos, encontrando que Podemos presenta el mayor porcentaje de positivos respecto al resto, mientras que VOX presenta la mayor proporción de sentimiento negativo.

Así mismo, para profundizar en el análisis planteado anteriormente, se accedió a las palabras con connotaciones negativas más presentes en el discurso

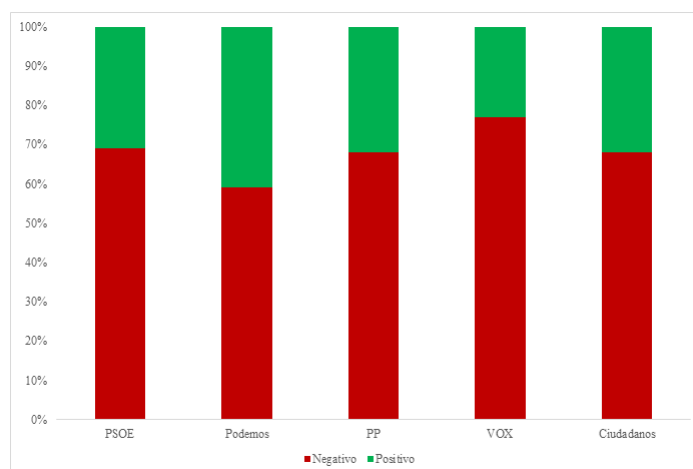


Figura 4.2: Análisis de sentimiento de los mensajes producidos por *bots* agrupados por estilo de comunicación.

producido por *bots* de acuerdo con su asociación a cada partido político - ver figura 4.3 -. En primer lugar, cabe señalar que en todos los casos se observan tres temas recurrentes, a saber sanidad, economía y política. Por otro lado, se encuentran diferencias en las palabras usadas por cada uno de los partidos asociados, identificando los mensajes asociados a PP y VOX como aquellos que presentan un mayor uso del lenguaje incívico y hostil, haciendo uso de palabras malsonantes.

Finalmente, se procedió a la aplicación de análisis de redes sociales junto con la aplicación de algoritmos de detección de comunidades con el fin de observar la estructura de la red para cada uno de los diferentes tipos de usuarios - *bots* y no *bots* -, permitiendo de esta forma conocer la organización de cada conjunto así como la naturaleza de sus relaciones.

En primer lugar, se llevó a cabo la tarea referente a la red de cuentas normales, o no automatizadas, encontrando una modularidad de 0.538 y 582 comunidades. Sin embargo, es importante señalar que de todas las comunidades presentes, destacan tres comunidades principales, las cuales representan un total del 77.73 % de todos los usuarios no automatizados. Como se aprecia en la figura 4.4, la comunidad morada - Comunidad A - se corresponde con el ala izquierda de la política española, representando un total del 40.01 % de todos los usuarios, encontrándose cuentas como la del presidente del gobierno Pedro Sánchez o Pablo Iglesias, los partidos políticos Izquierda Unida, Podemos o PSOE. Por su parte, las comunidades verde - Comunidad B - y azul - Comunidad C - representan el ala derecha de la política, suponiendo un total del 29.03 % y 8.69 % de los usuarios respectivamente. En estas comunidades



Figura 4.3: Palabras más frecuentes asociadas a sentimientos negativos por cada estilo comunicacional.

se encuentran cuentas referentes a los partidos políticos PP y VOX, así como sus líderes Pablo Casado y Santiago Abascal. Se puede observar que estas tres comunidades se dividen y distancian entre sí de acuerdo con la ideología política que representan cada una de ellas, encontrando una división a razón de izquierda-derecha de la política.

Por último, se repitió el proceso para las cuentas automatizadas, encontrando un significativo mayor número de comunidades en la red - 4601 -, con una modularidad de 0.766. Al igual que en el caso anterior, a pesar del alto número de comunidades, se encontraron dos principales, representando el 41.35% de todos los *bots* de la red, los cuales, el 21.83% pertenecen a la Comunidad A y el 19.52% a la Comunidad B. Sin embargo, en la figura 4.5 se representan los nodos coloreados, no por las comunidades a los que pertenecen - separadas visualmente en dos grupos principales -, sino por el color del partido político relativo al estilo comunicacional que presentan sus mensajes. De nuevo, se ve una clara separación en dos grupos principales y opuestos en torno al eje político del país. A la izquierda, se observan aquellos *bots* que presentan un estilo comunicacional referente a la izquierda política del país, compuesto por partidos como PSOE - rojo - o Podemos - morado

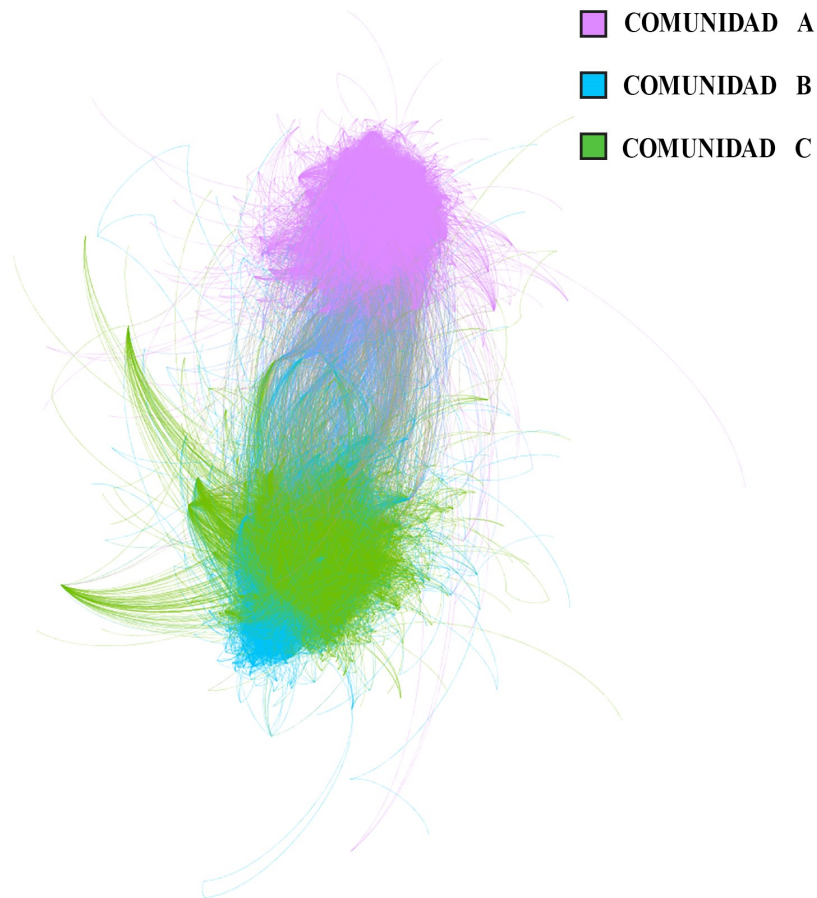


Figura 4.4: Red de usuarios no *bots*.

-, mientras que en la comunidad de la derecha se observa el ala derecha de la política, compuesta por *bots* relacionados con partidos como PP - azul - y VOX - verde -.

4.1.4. Conclusiones.

Como consecuencia de las nuevas oportunidades que ofrece la web 2.0 y el surgimiento de las redes sociales online, como Twitter, se presentan nuevas vulnerabilidades que puede precipitar la aparición de fenómenos como lo es la polarización. En este estudio, se ha perseguido el objetivo de detectar el comportamiento de los usuarios online en la red social Twitter durante la crisis sociosanitaria presentada por la pandemia del COVID-19, así como la presencia y el papel de los *bots* en el debate digital de la misma.

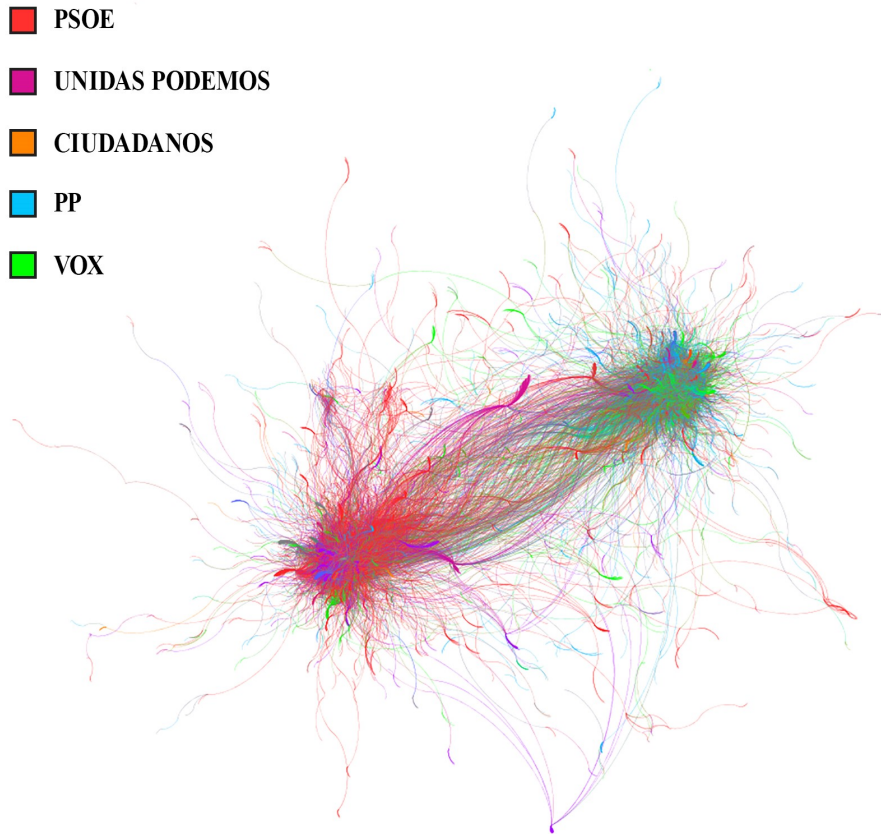


Figura 4.5: Red de usuarios *bots*.

Así, en primer lugar, se puede contrastar la presencia de *bots* en el debate digital, los cuales podrían estar actuando como detonadores del debate a través de campañas de desinformación sujetas a intereses políticos. Como consecuencia, las cuentas automatizadas presentan un *rol* de “megáfono” en la red emitiendo contenido sesgado, mostrando patrones ya definidos anteriormente en la literatura por Hwang y Wooley (2016) [60], denominados como “facilitadores”, cuyo fin no es otro sino el de la difusión de mensajes con el objetivo de conseguir la mayor repercusión posible. Como consecuencia, la presencia de *bots* propicia el enrocamiento y radicalización de las posiciones del resto de usuarios hacia un determinado debate.

Así, tras el estudio en mayor profundidad del contenido publicado por los *bots* se ha observado el uso mayoritario de contenido con carácter negativo así como el uso de incivildad política así como el de palabras malsonantes, cuyo

fin no es otro sino el de propiciar el conflicto y la separación entre usuarios que presentan opiniones heterogéneas.

Finalmente, se observa que lo reflejado anteriormente conduce a una separación y ruptura de la sociedad reflejada en el análisis de redes sociales llevado a cabo, encontrando una ruptura en torno al eje político izquierda-derecha en el debate, tanto de usuarios normales como de cuentas automatizadas. Por consecuente, se concluye la presencia de cuentas automatizadas en el debate digital sobre la crisis sociosanitaria sufrida por España, cuyo *rol* ha sido el de crear distanciamiento social y político entre los usuarios de Twitter. Este escenario supone la antesala al origen de fenómenos como es el de la polarización, concluyendo el papel protagonista que interpretan las cuentas automatizadas para el origen de la polarización, concluyendo una estrecha relación entre la existencia de los mismos y la formación de la polarización política e ideológica en torno a una sociedad virtualizada.

4.2. Identificación de la Polarización: Polarización e incivilidad en el debate digital español sobre los derechos de la mujer.

En la sección anterior se ha estudiado sobre los posibles agentes polarizantes y los orígenes del fenómeno de polarización, conociendo en profundidad la naturaleza de los eventos que han de tener lugar para hacer factible la división de la población. En los siguientes apartados se presentan dos trabajos cuyo objetivo es el de identificar la presencia de la polarización en un determinado y observar las consecuencias de la misma en el entorno digital.

Robles, J. M., Atienza, J., Gómez, D., & Guevara, J. A. (2019). La polarización de La Manada: El debate público en España y los riesgos de la comunicación política digital. *Tempo Social*, 31, 193-216.

&

Guevara, J. A., Atienza-Barthelemy, J., Gomez Gonzalez, D., & Robles, J. M. (2021). Polarization and incivility in digital debates on womens rights in Spain. Not just a matter of machismo. *Journal of Gender Studies*, 1-15.

Resumen.

Este estudio se centra en dos casos de estudio sobre los derechos de las mujeres en España. Se analiza el debate en la red social (Twitter) en torno al caso de abuso sexual La Manada, así como las denuncias públicas de violencia

misógina expuestas bajo el *hashtag* “cuéntalo”. Primero, nuestro objetivo es mostrar cómo la polarización digital y la rudeza, que afecta particularmente a las mujeres, no han logrado generar un debate horizontal sobre la inclusión. Además, a través del análisis de big data y redes sociales, se muestra que los efectos del machismo estructural se pueden evitar en mayor medida cuando se organizan las discusiones. Es decir, ha surgido en un entorno menos monopolizado, dominado por líderes de opinión masculinos que no solo han profundizado y encendido la discusión, sino que también la han fusionado y oscurecido con otras cuestiones más amplias que afectan al debate público en España. Así, los debates inclusivos y horizontales que involucran a las mujeres dependen no solo de las limitaciones del machismo estructural, sino también del diseño de los espacios públicos de discusión.

Abstract.

This study focuses on two case studies of women’s rights in Spain. We analysed the debate on the social network (Twitter) surrounding the sexual abuse case "La Manada" (Wolf Pack), as well as public reports of misogynistic violence exposed under the hashtag “cuéntalo”. First, our aim was to show how digital polarization and rudeness, which particularly affects women, have fallen far short of sparking a horizontal debate on inclusivity. Furthermore, through analysis of big data and social networks, we show that the effects of structural machismo can be avoided to a greater extent when discussions are organized. That is, it has arisen in a less monopolistic environment, dominated by male opinion leaders who have not only deepened and ignited the discussion, but have also merged with and obscured it with other broader issues affecting public debate in Spain. Thus, inclusive and horizontal debates involving women depend not only on the constraints of structural machismo, but also on the design of public spaces for discussion.

4.2.0.1. Redes sociales y comunicación digital.

Desde finales del siglo *XX*, la incorporación del Internet en la sociedad ha revolucionado la forma de interacción del ser humano, creando oportunidades comunicativas nunca antes contempladas y con un potencial revolucionario en cuanto a la sociedad se refiere. La digitalización de la realidad es, hoy en día, un hecho donde a lo largo de las últimas décadas diferentes autores han puesto de manifiesto la aparición de un nuevo fenómeno digital, centrado en las oportunidades de creación de contenido digital, así como la capacidad de divulgación que ofrece Internet del mismo, denominado como producción digital (Schradie, 2011) [103] o participación digital (Lutz, Hoffman y Meckel,

2014) [71].

Entre las diferentes repercusiones que supone la intrusión de éstos fenómenos en la sociedad, la comunicación política se ha visto afectada de manera especial donde, en comparación a la difusión y actividad vertical observada en un marco tradicional, ésta se ha visto forzada a derivarse en una comunicación predominantemente horizontal, gracias a las oportunidades ofrecidas por la red. En este nuevo marco, ciudadanos y políticos interactúan de forma directa y bidireccional, favoreciendo la profundización democrática (Hague y Loader, 2005) [54]. Así pues, en este nuevo escenario de participación ciudadana en el discurso político, se abre la posibilidad de estudiar los diferentes discursos de índole social y político en la sociedad, así como la formación y posicionamiento de diferentes comunidades, la interacción entre ellas y el discurso y naturaleza de las mismas. Conceptos como el de polarización social o *incivility*, suponen un aspecto clave para la comprensión de tales fenómenos ya que, surgidos como consecuencia de ésta desintermediación en la comunicación, suponen un riesgo para un diálogo adecuado entre diferentes comunidades, derivando en la división de la sociedad en dos grupos opuestos, caracterizados por la contrariedad de sus argumentos. Especialmente en este contexto, se quiere resaltar el protagonismo de la cuestión de género en este estudio, donde el papel de las mujeres se ha visto altamente infravalorado e infrarrepresentado en la red a pesar de ser el colectivo con mayor número de usuarios en las redes sociales online.

Como consecuencia de la desintermediación de la comunicación, varios estudios han resaltado los potenciales efectos negativos de la red en el escenario de la comunicación política (Gentzkow, 2016) [38]. Entre los diferentes efectos que han llamado la atención de los autores cabe resaltar la polarización de la sociedad, la *incivility* o la homofilia, los cuales suponen un desequilibrio para la dimensión deliberativa de la democracia (Sunstein, 2017) [111].

Ya en 1954, Lazarsfeld y Merton [68] propusieron la existencia de dos diferentes tipos de homofilia, la homofilia de estatus y una homofilia de valores, explicadas en capítulos anteriores 2.1.1. En este sentido, cabe resaltar que, debido a esa homofilia de valores, categorizada como homogeneidad, se forman diferentes comunidades diferenciadas por sus valores y que, como consecuencia, aparece otro fenómeno conocido como polarización. Una vez situados en este contexto, cuando los usuarios de una red social generan y comentan contenido respecto a una situación de contenido político, los usuarios se dividen en diferentes grupos conforme a sus valores comunes, separando a la sociedad en dos grupos relativamente igual y equilibrados con opiniones opuestas. Este fenómeno es conocido como polarización política, donde una sociedad se ve segregada en función a la opinión de sus ciuda-

danos en cuanto a una cuestión política se refiere. De acuerdo con Robles, Vélez, Marco Rodríguez y Gómez (2020) [96], este fenómeno es aplicable a las redes sociales online, como Twitter, donde los usuarios de las mismas generan opiniones, interactúan y forman comunidades conforme a sus opiniones.

El objetivo de esta investigación es el de, a través de los casos de estudio de #LaManada y #Cuéntalo, de gran relevancia y repercusión social y político en España, observar y conocer la naturaleza, a través de la red social Twitter, de los efectos de la desintermediación en las comunidades que lo componen. En este sentido, nos interesa estudiar las características de las redes de comunicación observadas, los principales agentes y mediadores que las componen, así como observar la polarización política y social de la red consecuente de tales debates. Consideramos tales objetivos de vital importancia para una mejor comprensión de los procesos de generación, procesamiento y transmisión de la información en la red.

4.2.1. Metodología y Resultados.

4.2.1.1. Casos de estudio.

Caso 1: La manada.

La serie de actos que tuvieron lugar durante las fiestas de San Fermín, Pamplona, en la mañana del 7 de julio de 2016, se denominó el caso de “La manada”. Ese día, cinco hombres abusaron sexualmente de una joven de 18 años. La víctima denunció a los cinco hombres, lo que derivó en un proceso judicial y social que supuso el revuelo de la opinión pública. El debate público ganó gran protagonismo en el proceso judicial a medida que surgieron noticias sobre el caso, y ciertamente gana mucho poder una vez que se conoce el veredicto de un caso.

Caso 2: Cuéntalo.

La iniciativa denominada #Cuéntalo impulsada en Twitter la inició Cristina Fallarás, una periodista que, entre otros medios trabajó en Diario 16 digital el 26 de abril de 2018, una vez conocida la sentencia en el anterior caso de La Manada. El principal objetivo de la iniciativa es permitir que mujeres de toda España puedan compartir sus experiencias relacionadas con el sexismo a través de este medio - redes sociales, especialmente Twitter -. La convocatoria de esta iniciativa en la opinión pública española pronto se hizo patente. En los primeros 14 días, más de 800000 mujeres hablaron sobre cómo vivieron el abuso sexista. A partir de ahí, la iniciativa se hizo más popular, con más mujeres anónimas y de alto perfil uniéndose al movimiento.

4.2.1.2. Objetivos de estudio.

- Estudiar las características de una red de comunicaciones, así como sus comunidades sobre un caso de repercusión social y político de índole conflictivo (#LaManada).
- Observar la polarización política existente en dicha red, así como las características de la misma (#LaManada).
- Examinar las características del discurso presentado por ambos extremos de la polarización observada (#LaManada).
- Conocer la naturaleza de una red de comunicaciones sobre un caso de interés, tanto social como político (#Cuéntalo).
- Identificar los actores relevantes en la difusión de información de dichas comunidades, así como las estrategias de difusión (#Cuéntalo).
- Determinar los brokerage roles entre comunidades (#Cuéntalo).

4.2.1.3. Fuente de Datos.

Twitter es una red social donde se publican mensajes cortos de hasta 280 caracteres, llamados tweets. La red social permite diferentes formas de interacción entre sus usuarios, como interacciones entre mensajes directos privados, menciones y tuits, como retuits, citas o respuestas. Numerosos estudios científicos han demostrado que Twitter es un termómetro fiable para analizar el comportamiento humano en relación con determinados temas de interés, como los procesos electorales (Borondo et al., 2012 [13]; Caldarelli et al., 2014 [19]; Martín-Gutierrez et al., 2018 [72]), crisis políticas (Cárdenas et al., 2018) [20] o movimientos sociales (van Haperen et al., 2018) [115].

Los datos sobre los que se realizó este estudio fueron tuits públicos sobre los casos “La Manada” y #Cuéntalo, y las relaciones resultantes. Estos tweets y la información de sus características se descargan en tiempo real utilizando la API de transmisión proporcionada por Twitter.

Se descargaron los tweets de acuerdo con los siguientes filtros:

- Estudiar las características de una red de comunicaciones, así como sus comunidades sobre un caso de repercusión social y político de índole conflictivo (#LaManada).
- Contiene las palabras la y manada, sin importar el orden y sin necesidad de que las dos palabras estén seguidas.

- Contiene el *hashtag*: #manada.
- Contiene el *hashtag*: #LaManada.

Los datos se descargaron durante el mes de febrero de 2019, concretamente entre el 6 al 19 de febrero de 2019, resultando un total de 74951 tweets.

Para el caso concreto de #Cuéntalo, se analizaron todos los tweets y sus relaciones dentro del mismo *hashtag*. Del 26 al 29 de abril de 2018 fue noticia el *hashtag* #Cuéntalo para dar a conocer casos de agresiones sexuales. La información obtenida y recogida fueron los tuits emitidos bajo este hashtag con las siguientes variables: id tweet, date, author, text, app, id user, followers, following, statuses, location, url, geolocation, name, description, url_media, type media, quoted, relation, replied_id, user replied retweeted_id, user retweeted, quoted_id, user quoted, lang.

4.2.1.4. Método y resultados de La Manada.

Red de retweets.

De todos los mecanismos de interacción que pueden darse entre un usuario y los tuits que genera, los retuits son el mecanismo por el que un usuario comparte un tuit escrito por otro usuario. Cuando esto sucede, los seguidores del usuario que retuiteó el tuit verán el tuit original y podrán decidir retuitear de nuevo, repitiendo el paso anterior en sus seguidores. Esta interacción es muy frecuente en Twitter y constituye la actividad número uno en Twitter donde se lleva a cabo el proceso de difusión y difusión. Cuando un usuario retuitea un tuit, se puede entender que está de acuerdo con lo que incluye el tuit. Por tanto, los retuits pueden utilizarse como indicador de influencia y proximidad de opiniones entre los usuarios. Es por esta razón que nos asociamos con la red de reenvío.

En esta red, los usuarios que participan en la conversación de “La Manada” se denominan nodos, mientras que los enlaces son retweets entre usuarios. Las aristas son dirigidas porque se entiende que el usuario del tuit original está influyendo en el usuario del retuit, y no al revés. De esta forma, la dirección del enlace va desde el usuario que escribió el tuit original j hasta el usuario que retuiteó i , que es el sentido de la difusión de información.

Una vez que se establece la red de reenvío, se aplican algoritmos de detección de comunidades. Es decir, un estudio evalúa si existen grupos de nodos con gran cantidad de enlaces. Para ello, lo primero que hay que hacer es reducir la red a su mayor componente conexa. La red de retweets relevantes obtenida consta de 48710 nodos y 71325 enlaces.

A continuación, se procede a calcular la modularidad, aplicando la fórmula ya propuesta en el capítulo anterior - ver 2.9 -. La forma de implementar

esta optimización en el método Louvain es optimizar la modularidad localmente en todos los nodos. Luego, cada pequeña comunidad se agrupa en un nodo, y el primer paso se repite hasta que la modularidad general no se puede mejorar modificando la distribución actual de la comunidad. El resultado del cálculo es una agrupación de nodos de diferentes comunidades, dos de los cuales destacan por su gran tamaño, el mayor contiene 54,34% de usuarios, y el segundo contiene 24,26%. El resto de comunidades fueron eliminadas. De esta forma, la red de reenvíos se divide en dos comunidades bien diferenciadas, en las que existe una clara polarización. Esta red consta de 38.284 de usuarios y 60000 enlaces entre ellos - ver imagen a continuación 4.6 -. Las comunidades que apoyan a las mujeres maltratadas y/o los mensajes críticos con las decisiones judiciales están pintados de púrpura. Se pinta de verde la comunidad formada por el debate en la que se relativizan las acciones de los cinco y/o se ratifican las decisiones judiciales. De esta manera, se tienen dos comunidades distintas en torno a un eje de doble tema.

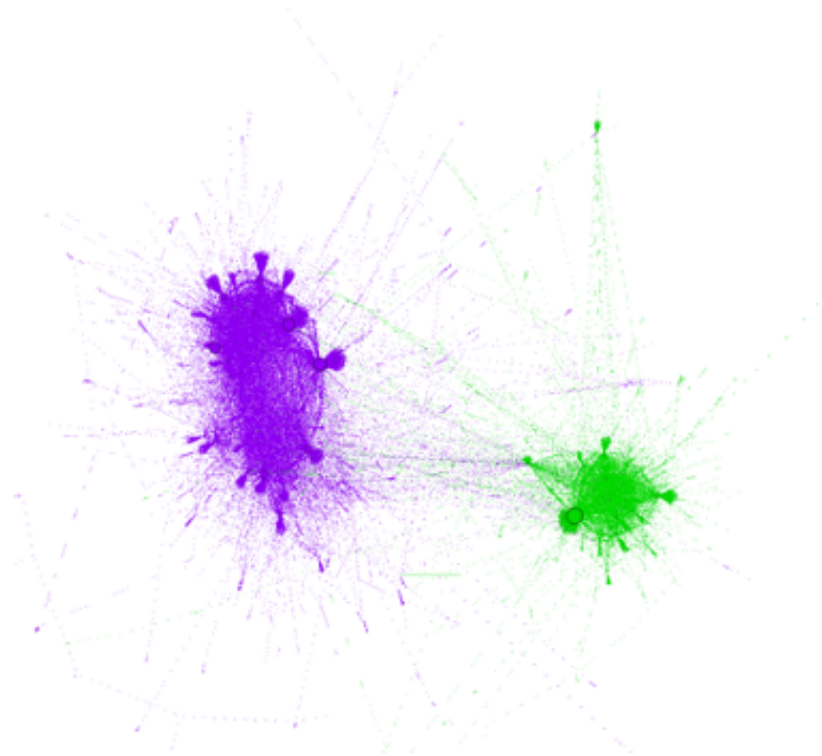


Figura 4.6: Comunidades principales de la red de retweets de “La manada”.

Nube de palabras.

Una vez separados estos 38284 usuarios en dos grupos distintos, se procedió a analizar el uso de palabras por parte de cada uno de ellos. Para inferir las

diferencias se generó una nube de palabras que los usuarios que integraban estos grupos utilizaban con mayor frecuencia en sus tuits sobre “La Manada”.

Por otro lado, téngase en cuenta que un tuit puede ser original, retuiteado o una cita de otro tuit. En este caso, no es lo mismo que se mencione la palabra, ya sea de forma directa - es decir, escrita por el propio usuario - o indirecta - es decir, reenviando el mensaje o haciendo referencia a la ocurrencia de la palabra -. Escribir una palabra la primera vez es más importante para el remitente, sobre todo si la palabra se vuelve viral, le da más poder. Por lo tanto, las palabras se pueden mencionar de tres maneras diferentes:

- Manera directa, si es el propio usuario el que ha escrito el tweet. Se suele entender que, si el usuario ha escrito el tweet, está de acuerdo con lo que dice.
- Manera indirecta, si se trata de un retweet de un tweet escrito por otra persona. Se suele entender que, si el usuario ha retweeteado el tweet, está de acuerdo con lo que dice.
- Manera indirecta, si la palabra aparece en el tweet que el usuario está citando. En el caso de la cita, por el contrario, no es posible concluir que el usuario está de acuerdo con lo que dice el tweet citado.



Figura 4.7: Nube de palabras de mención total. A la izquierda, comunidad de color morado. A la derecha, comunidad de color verde.

Interpretación y análisis de los resultados de #La Manada.

En primer lugar, casi 8 de cada 10 participantes en la conversación apoyaron una de las dos opiniones (78,6%). Por supuesto, la mayoría de las opiniones fueron aquellas que expresaron opiniones fuertes sobre la naturaleza del incidente, la violación y/o la falta de una sentencia fuerte (54.34%). Sin embargo, el soporte para la otra opción también es muy alto (24,26%). En segundo

lugar, significa que las opciones alternativas o de reconciliación no existen o son fundamentalmente marginales. Es decir, sobre este tema de análisis, la opinión pública española se expresa de forma muy fragmentada.

Otro aspecto a considerar es que la mayoría de los líderes de opinión de cada comunidad son figuras públicas - periodistas, académicos, activistas, etc. -. Sin embargo, también hay líderes que claramente no pertenecen a ningún departamento u organismo que defina la estructura y funcionamiento de la opinión pública tradicional. En este caso, podemos observar un grado de desintermediación y, por tanto, un cambio en la opinión pública.

Finalmente, mencionaremos la nube de palabras y, por lo tanto, el contenido del debate entre las dos comunidades. Lo primero que queremos destacar es la diferencia entre las dos comunidades. Es importante observar cómo la comunidad verde implica constantemente elementos fuera del ámbito estricto del debate, permitiendo enmarcar el diálogo para darle el sentido que se pretendía dar. Primero, hay una referencia consistente y muy poderosa a la palabra “Sabadell” por parte de esta comunidad. Entendemos que la referencia anterior se refiere a una violación en la ciudad catalana en 2018. Lo importante del comportamiento anterior es que existen similitudes y diferencias en las circunstancias de los casos aquí tratados. Similitudes; una mujer es agredida sexualmente por un gran grupo de hombres. discrepancias; en el caso Sabadell, la mujer había sido previamente secuestrada y trasladada a la fuerza a un lugar apartado donde se produjo el crimen. La inclusión de la palabra Sabadell podría significar que los dos delitos eran diferentes, y por tanto, la conducta que tuvo lugar en San Fermín en 2016 fue de distinta naturaleza. En la misma lógica, entendemos que la palabra Alicante se refiere a un caso en el que un reincidente dejó claro que se trataba de un caso de violación.

Se debe resaltar que, en general, la *invicibility*, si bien emerge y es fundamental en este debate, no es tan decisiva como la polarización de este debate. Cuando se trata de la comunidad morada, debemos enfatizar que es un discurso que claramente se mueve en la lógica de los derechos civiles. Por lo tanto, las palabras centrales son “libertad”, “justicia”, etc. Se refiere a los derechos de las mujeres y sus demandas de igualdad.

En definitiva, si bien podemos analizar cada palabra en todas las nubes, es fundamental señalar que los temas aquí analizados han generado mucha polarización por intereses más generales y de quiebre típico. la situación política del país. Asimismo, señalar la falta de discusión y calificación - más comúnmente retuiteada - de los discursos de los líderes de los medios exagera esta ruptura en la comunicación política.

4.2.1.5. Método y resultados de #Cuéntalo.

Modelización de la red de tuiteros.

Los nodos de la red corresponden a los diferentes usuarios que estuvieron activos durante la campaña “Cuéntalo”, ya sea porque escribieron sobre algunas experiencias relacionadas con el tema o porque retuitearon las experiencias de otras personas. Un enlace dirigido ocurre entre dos usuarios i y j con dirección j sobre i si: el usuario j es mencionado por la cuenta i , o si el usuario i retweetea al usuario j .

En los casos en que esto se repita más veces durante la campaña, la relación valorada representa el número de veces que j es retweeteado o mencionado por el tuitero i , por lo que la influencia del nodo j aumenta sobre i . En el análisis de redes sociales, esta relación debe entenderse como direccionalidad y dominancia o transferencia de información, entendiendo que si existe una arista entre los nodos i y j , es porque j está recibiendo información de un tuitero i , por lo que i afecta al nodo j de alguna manera.

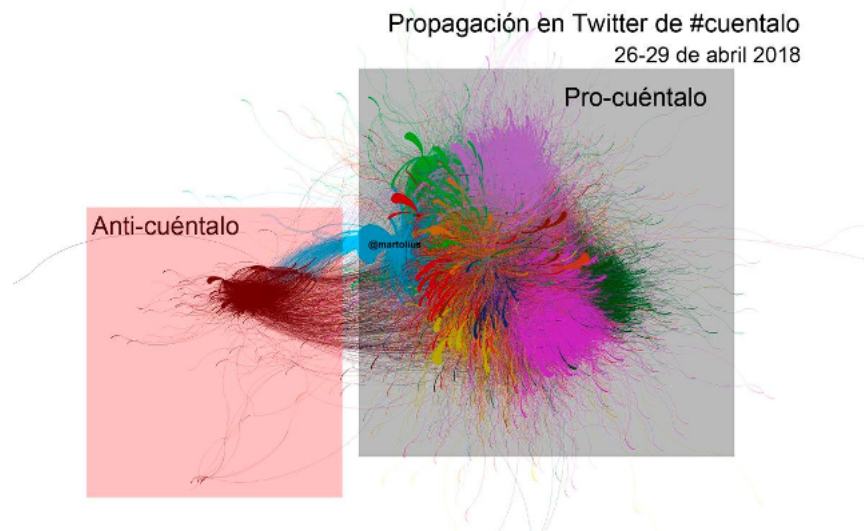


Figura 4.8: Visualización de la red cuéntalo presentada por la profesora y tuitera Dña. Mariluz Congosto.

Finalmente, el número de nodos de esta red es de 150668 con un total de 408576 relaciones.

El número de aristas medio por usuario es de 5.42, pero la varianza es tremendamente alta lo que indica heterogeneidad en los roles de los usuarios dentro del debate digital.

Análisis de la aleatoriedad de la red de tuiteros.

Teniendo en cuenta que el tamaño de la red es $N = 150668$ y el número de aristas $L = 408576$, entonces se tiene un total de posibles relaciones igual a 11350347778, con un ratio de $p = 0.00003599$. Así, el siguiente paso es probar si la variable grado sigue una distribución binomial con parámetros N y p . Para responder a esta pregunta, se realiza cuatro pruebas de bondad de ajuste en las variables de grado en Normal, Exponencial, Poisson y Uniforme, todas las cuales rechazaron la hipótesis nula que concluiría que nuestra red no es una red aleatoria.

Análisis de la escalabilidad de la red y red de pequeño mundo.

Una red sin escala tiene una distribución de grados diferente a la de una red aleatoria. Si bien la probabilidad de encontrar un nodo con un grado mayor en una red aleatoria es efectivamente cero, es común encontrar un nodo con grado alto, o *hubs*, en una red sin escala. Una explicación para llegar a redes libres de escala es la idea de conexiones preferenciales, donde los nodos que aparecen en la red tienden a retweetear - en este caso - más a menudo que los que retweetean menos.

A continuación, se comprueba los grados - para grados mayores que uno - tienen una distribución similar a la potencial. Los grados y sus frecuencias se pueden ajustar a una ecuación latente de la forma $y = x^{-2.975}$ con un coeficiente de determinación 68%, aceptando el comportamiento similar a la de una red libre de escala.

A continuación, se analiza la distancia media entre dos nodos de la red. Para comprobar o contrastar si esta red es una red de mundo pequeño, el primer paso es analizar las componentes débiles conexas. Dado que la red está desconectada, podemos - estrictamente - decir que esta red no es una red de pequeño mundo, dado que hay pares de nodos que no están conectados por una cadena. Aunque hay un número total de componentes conexas 598 en la red de Twitter, lo que es realmente relevante son las llamadas componentes conexas gigantes. Esta componente conexa presenta el 98% de los usuarios de la red ya que de los 158668 vértices de esta componente conexa pertenecen 148967.

Dentro de la componente conexa gigante se calculan los caminos mínimos entre cada par de nodos. Una vez calculados todos los caminos mínimos se computa la distancia media entre dos tuiteros que, para esta red, es de 3.54. Esto, unido a la baja densidad, permite asegurar estadísticamente que nos encontramos ante una red de pequeño mundo. Del análisis de ambas carac-

terísticas podemos concluir que nos encontramos con una red libre de escala.

Estructuras identificadas

Después de un análisis detallado de la red, se puede concluir que la red tiene esencialmente estrellas de tamaño 4 o más, y la segunda estructura detectada es una cadena ordenada de longitud 3 o más. Estos dos tipos de estructuras nos permiten dilucidar la relación entre los usuarios. Por un lado, se tienen los tuits generados por personas con muchos seguidores que se viralizan tras ser retuiteados por un gran número de seguidores. En cambio, cuando estos tuits “virales” los producen personas con pocos seguidores, se establece una cadena entre amigos hasta llegar a un nodo con muchos seguidores y formar una gran estrella.

Detección de comunidades en la red del caso #Cuéntalo.

Las personas suelen asociarse con personas de características similares - redes homofílicas -, pero no siempre es así. En los casos en que tengamos comunidades previamente definidas por algunas características intrínsecas de cada nodo - por ejemplo: ubicación, afinidad política o género -, será interesante ver si las comunidades resultantes de vínculos mantienen a priori estas definiciones comunitarias.

Tras aplicar algoritmos de detección de comunidades, se detecta una red modular con comunidades bien definidas y densas y conexas. En las comunidades de mayor tamaño se detectan diferentes líderes, correspondiéndose con el fenómeno *rich club*, caracterizadas por pocos nodos muy influyentes y conectados entre sí, pero con algo número de seguidores cada uno. El siguiente gráfico muestra a los usuarios de Twitter con más de 1000 conexiones de las cuatro comunidades más influyentes.

Es interesante mencionar que, en la red representada en la figura 4.9, los usuarios más influyentes representan usuarios de carácter *amateur*, mostrándose como líderes de debate.

Interpretación y análisis de los resultados de #Cuéntalo.

En este caso, además de la estructura en forma de estrella, se encontró una estructura en forma de cadena. Esto quiere decir que al pasar de un nodo a otro, las personas con pocos seguidores consiguen que su mensaje llegue muy lejos - a mucha gente - y así ser muy leído. Este resultado tiene un valor analítico muy importante porque significa que la red analizada se

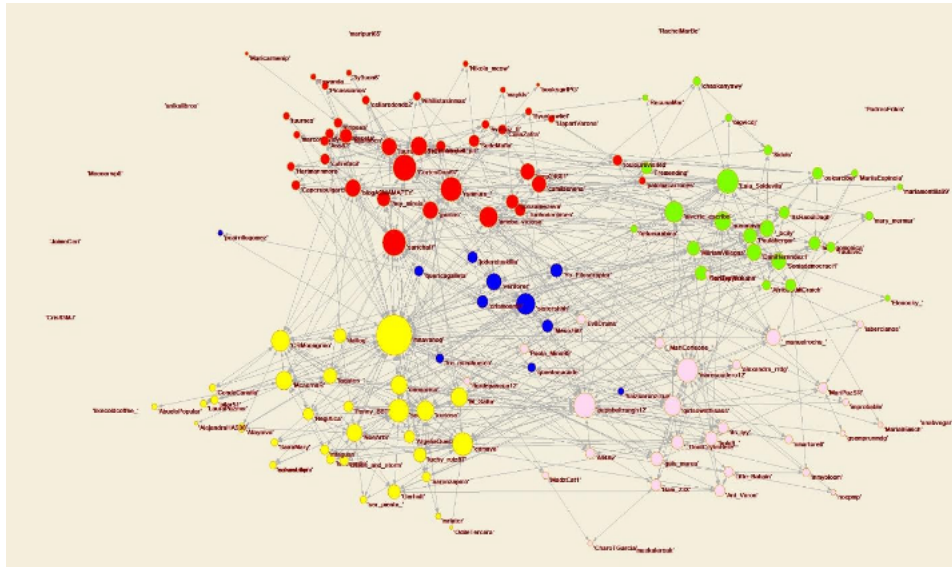


Figura 4.9: Grafo de relaciones de los 145 usuarios más influyentes.

caracteriza por el poder de difusión de personas con muchos seguidores - que en cierto sentido son famosos -, pero también caracterizada por personas *amateur*. Cambiar esta interpretación a los detalles del caso significa que el testimonio de abuso compartido por celebridades - estrellas de Internet - tiene un gran impacto debido a que es ampliamente retuiteado. Sin embargo, las experiencias que las personas anónimas comparten sobre sus experiencias también recorren un largo camino y pueden tener un impacto - en forma de cadenas -.

Además, hay un conjunto de comunidades con múltiples nodos centrales. Dicho esto, la comunidad emisora es mucho más democrática que el caso de la manada. Es una red con una estructura política definida, no desestructurada como “La Manada”. La experiencia de Cuéntalo permite que muchos nodos se sumen al debate en igualdad de condiciones, evitando el monopolio de grandes figuras mediáticas centrando el debate en su temática en sí, y no otras cuestiones distractoras. Otro tema importante es la densidad de la red, que no es solo una estructura de red de comunidades de mundo pequeño que se comunican dentro de la comunidad. En este caso, los retweets de la experiencia son muy altos, tanto dentro como fuera de la comunidad. Las comunidades que pueden formarse por cuestiones como la homogeneidad están abiertas a la entrada y salida de información.

Esto tiene dos consecuencias. Primero, las personas del mismo rasgo (homofilia) que componen una comunidad no están cerradas sino abiertas a las experiencias que han tenido personas de otras comunidades y por lo tanto

tienen otros rasgos. Esto naturalmente, y en segundo lugar, tiene una lectura muy positiva de la opinión pública en términos de estructuras de comunicación y debate público.

En teoría, en este caso, nos interesa la desintermediación. Lo dicho hasta ahora, la cadena de transmisión de información, los múltiples nodos centrales en la comunidad, y la fuerza del flujo de información, nos hablan de una red muy desintermediada. Es decir, una red donde los nodos "no profesionales" son tan importantes como los formados por "celebrities".

4.2.2. Conclusiones.

En cuanto a la polarización se refiere, cabe destacar la presencia de la misma en el caso de la "La Manada" al contrario que lo observado en el caso de #Cuéntalo. La aparición de este fenómeno en uno de los casos y no al otro se debe a las características y estructura del debate de "La Manada" en la red frente al de #Cuéntalo.

En este último caso, se observa un debate desintermediado, en el que tanto usuarios influyentes como *amateurs* participan en la conversación. Además, se observa que la naturaleza de las comunidades es abierta, encontrando una gran interacción entre diferentes usuarios de comunidades, caracterizando el debate por su cooperación. Por el contrario, en el caso de "La Manada" se encuentran dos comunidades principales, opuestas entre sí, donde ambas representan las dos partes contrarias del debate. En adición, cabe señalar la presencia de incivilidad entre los usuarios de la red, caracterizando el debate por la presencia de conflictividad entre comunidades. En este sentido, cabe relacionar la aparición de la polarización, de nuevo, a la presencia de *invility*, así como a un debate desestructurado.

Así mismo, es importante señalar que en esta investigación se ha identificado la polarización, observando sus detonadores y consecuencias en el debate. Tras el estudio del papel de los *bots* en el debate digital, así como los diferentes escenarios del caso de la manada y cuéntalo, se concluye la importancia que representa la radicalización de los individuos en el desarrollo de la polarización. En los capítulos siguientes, se procederá a la presentación de una medida de polarización capaz de medir la polarización basándose en esta premisa de radicalización en entornos de grandes bases de datos procedentes de redes sociales online.

Capítulo 5

La medición de la Polarización a través de los conjuntos borrosos

Resumen: En el capítulo presente se hace una propuesta novedosa en la literatura, presentando una medida de polarización basada en los conjuntos borrosos centrada en la medición del grado de radicalización de una población a los polos extremos de una variable actitudinal o ideológica. Así mismo, se presenta un trabajo donde se aplica dicha medida a un contexto real de grandes bases de datos descargados de la red. Por último, también se incluye un estudio sobre las aplicaciones que la medida de polarización presenta más allá de medir únicamente la polarización, haciendo uso de ésta métrica para mejorar algoritmos de detección de comunidades en redes.

5.1. Una nueva propuesta: medición de la polarización a través de los conjuntos borrosos.

Guevara, J.A., Gómez, D., Robles, J.M., Montero, J. (2020). Measuring Polarization: A Fuzzy Set Theoretical Approach. Communications in Computer and Information Science, vol 1238. Springer, Cham.
https://doi.org/10.1007/978-3-030-50143-3_40

Resumen.

La medición de la polarización ha sido estudiada en los últimos treinta años (Esteban y Ray, 1994 [28]; Foster y Wolfson (1992) [30]). A pesar que se ha abordado desde diferentes disciplinas, en tanto que el concepto de polarización es complejo, se encuentra una falta de consenso sobre como dicho fenómeno es medido. En esta investigación se propone una nueva aproximación de la medición del concepto de polarización basada en los conjuntos borrosos. La aproximación borrosa - *fuzzy* - proporciona la posibilidad de utilizar grados de pertenencia, presentando así una nueva perspectiva en la medición de la polarización. Debido a la naturaleza difusa de la realidad, la medición de la polarización debería reflejar dicho aspecto. En este sentido, se analizan las propiedades métricas de la polarización y se desarrolla una nueva medida de riesgo de polarización compuesta por operadores de agregación así como funciones de *overlapping*. En este capítulo se propone una medida de bipolarización basada en conjuntos borrosos a partir de la existencia de dos polos opuestos. La idea que subyace en esta medida trata de medir la polarización para cada par de individuos en base a su “grado de cercanía” a cada uno de los polos, entendiendo que se produce la polarización máxima entre dos individuos cuanto más cerca estén cada uno de ellos a los dos polos opuestos.

Una vez definida y analizada la medida se muestra su comportamiento sobre una población de tamaño $N = 391315$ para una escala tipo *likert* con diferentes distribuciones para probar la nueva medida. Se usaron otras medidas de polarización para comparar la nueva propuesta a lo largo de diferentes escenarios donde la aproximación borrosa propuesta en este estudio ofrece nuevos resultados donde se ha podido contrastar la importancia de las funciones de pertenencia en la medición de la polarización. Finalmente, se quiere resaltar el gran potencial que ofrecen los conjuntos borrosos en la medición de la polarización, abriendo un nuevo campo en tales líneas de investigación.

Abstract.

The measurement of polarization has been studied for the past three decades (Esteban y Ray, 1994 [28]; Foster y Wolfson (1992) [30]). Because the concept of polarization is complex, there is not a consensus in the literature on how to measure it despite different approaches. In this paper, a new method for measuring polarization phenomenon based on fuzzy sets is proposed. The fuzzy approach offers a new perspective with elements that allow membership degrees. Since reality is not black and white, polarization measures should include this key property. To this end, we analyze the properties of polarization

measures and develop a new risk of polarization measure using aggregation operators and overlapping functions. In this chapter we propose a measure of bipolarization based on fuzzy sets from the existence of two opposite poles. The idea behind this measurement tries to measure the polarization for each pair of individuals based on their “degree of closeness” to each of the poles, understanding that the maximum polarization between two individuals occurs the closer they are to each other to the two opposite poles.

Once the measure has been defined and analyzed, its behavior is shown on a population of size $N = 391315$ on a 5-likert scale with different distributions to test our measure. Other polarization measures have been applied to compare cases where fuzzy set methods give different results, where membership functions have been shown to play an important role in the measurement. Finally, we would like to highlight the new and potential contribution of the fuzzy set method to the measurement of polarization, which opens up a new field of research.

5.1.1. Una nueva propuesta de polarización.

La polarización es uno de los conceptos más estudiados en las ciencias sociales actualmente, donde en las últimas décadas ha presentado un incremento en el interés científico. El concepto de polarización ha sido estudiado desde diferentes perspectivas (e.g.: Estaban y Ray (1994) [28] Wang y Tsui (2000) [118], Montalvo (2001) [76], Apouey (2007) [4], Permanyer y D'ambrosio (2015) [89] o Morales et al (2015) [78]. Sin embargo, tal y como se puede comprobar tras un repaso a la literatura, no existe una medida de polarización universal y aceptada. Ligado a este hecho, no existe un consenso definido en la literatura sobre la verdadera naturaleza de la polarización, así como de su medición.

Sin embargo, una de las más citadas y usadas medidas de polarización fue propuesta en el marco de la economía. Wolfson (1994) [120] y Esteban y Ray (1994) [28] fueron dos de los primeros autores en proponer una medida de polarización formal. Estas medidas presentan una fuerte relación con el concepto de desigualdad tradicionalmente ligado a la economía. Desde entonces, un creciente número de medidas de polarización han sido propuestas en la literatura.

Entre todas ellas, es necesario remarcar la importancia de la propuesta presentada por Esteban y Ray en 1994, la cual incluye conceptualización teórica de la polarización incluida en las propiedades métricas de la medida, tales como identificación del individuo con el grupo, alienación, etc. Sin embargo, estas propiedades carecen de una naturaleza nítida y presentan gradualidad en sí mismas. De esta forma, la medida en la cual un determinado individuo

siente identificación hacia su grupo, o alienación hacia el resto de individuos, puede ser modelado a través de una función de pertenencia borrosa, las cuales representan el grado en el que un determinado individuo siente pertenencia hacia un determinado grupo.

Así, partiendo de la idea propuesta por Esteban y Ray desde una perspectiva bipolar, en esta investigación se propone una nueva medida de polarización a través del uso de los conjuntos borrosos y el uso de las funciones de pertenencia, gracias a las cuales se podrá medir el grado de radicalización de los individuos a los polos extremos de un eje actitudinal. Estas funciones son agregadas a través de operadores de agregación para, finalmente, obtener un valor que indica el riesgo de polarización entre dos individuos.

5.1.2. Una nueva medida de polarización basada en los conjuntos borrosos desde una perspectiva unidimensional y bipolar.

La mayoría de las medidas propuestas en la literatura sobre dispersión o polarización tratan de comparar para cada par de elementos un valor diferencial. Este valor en el caso de ER puede interpretarse como el antagonismo efectivo que existe entre los individuos $\{i, j\}$ en base a las rentas. En el caso de la medida de dispersión de IOV , por ejemplo, se mide la diferencia o distancia que existe entre ambos individuos.

Desde una perspectiva similar, se reflejará la polarización en una población como una agregación - suma - para cada par de elementos, donde el valor diferencial represente el grado de polarización o posibilidad de ruptura que tienen esos dos individuos en base a sus ideas, rentas, o posiciones ideológicas. Para tratar de computar este valor diferencial, en primer lugar se asumirá la existencia de dos polos X_A y X_B que representan los opuestos ideológicos. En este contexto, se usarán conjuntos borrosos para determinar el grado de cercanía con el que los individuos i y j se identifican tanto con el polo X_A como con el polo X_B y se usará esa información para obtener el grado de polarización entre $\{i, j\}$ entendiendo que la polarización máxima se produce cuando i está muy cerca de X_A y j muy cerca de X_B y viceversa. También se asume que fija la posición ideológica de i - supóngase cerca de X_A -, la polarización debería aumentar cuando el individuo j se mueve hacia el polo X_B .

Teniendo en cuenta estas premisas, formalmente se necesita conocer, para cada par de elementos $\{i, j\}$ únicamente su grado de cercanía a los polos X_A y X_B . Así, μ_{X_A}, μ_{X_B} representan las funciones de grado de pertenencia, donde $\mu_{X_A}, \mu_{X_B} : N \rightarrow [0, 1]$ son funciones, y para cada $i \in N$, $\mu_{X_A}(i)$ y $\mu_{X_B}(i)$ se

representa el grado de pertenencia de un determinado individuo i a ambos polos de la variable.

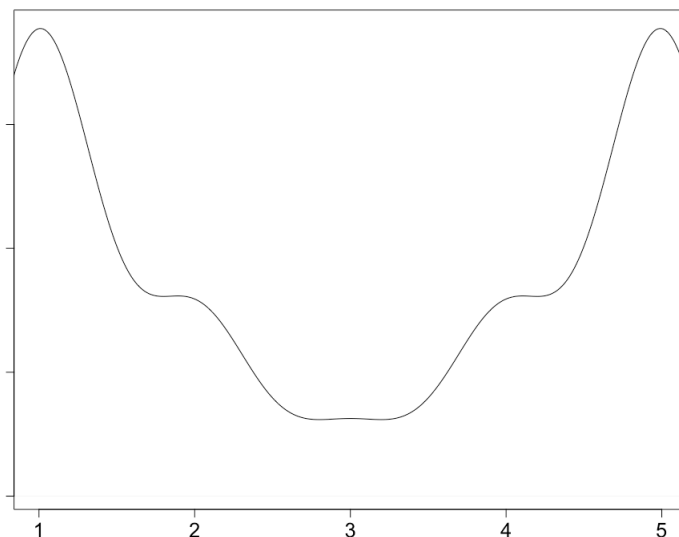


Figura 5.1: Ejemplo de una distribución bipolarizada.

Para este caso de bipolaridad en el que se asume la existencia de dos posiciones radicales o extremas, se entiende que la polarización se asocia a la aparición de las siguientes dos situaciones:

1. Una parte significativa de la población se sitúa cerca del polo A (X_A).
2. Una parte significativa de la población se sitúa cerca del polo B (X_B).

Por otro lado, al igual que ocurre en el caso de Esteban y Ray, se asume que existe la posibilidad de medir la discrepancia entre estos dos polos o actitudes extremas, siendo $\delta(X_A, X_B)$, asumiendo su valor máximo de 1 para el caso bipolar actual.

Finalmente, la medida de polarización propuesta en este capítulo puede ser entendida como la suma de la agregación de tres conceptos clave, siendo entendida como el riesgo de polarización que presenta una determinada población. Así, consideramos importante remarcar que cuando existe un alto grado de polarización, las relaciones entre grupos opuestos, así como su comunicación, se interrumpen.

De esta forma, se considera el riesgo de polarización entre dos individuos como la posibilidad de las siguientes situaciones.

- Cómo de cerca un individuo i está respecto con la posición extrema X_A y cómo de cerca un individuo J está respecto con la posición extrema X_B .
- Cómo de cerca un individuo i está respecto con la posición extrema X_B y cómo de cerca un individuo J está respecto con la posición extrema X_A .

Así, al asumir que la polarización se da cuando ocurren ambas situaciones, se propone la siguiente medida de polarización basada en la lógica difusa.

Definición 5.1.1 (Medida de bipolarización JDJ). Sea $X = \{x_1, x_2, \dots, x_n\}$ una variable actitudinal o ideológica para una población $N = \{1, 2, \dots, n\}$ y sean X_A y X_B los polos de dicha variable, asumidos como los valores máximos y mínimos teóricos de la misma. Sea $\mu_{X_A}(i)$ y $\mu_{X_B}(i)$ los grados de pertenencia de un individuo i a ambos polos, donde $i \in N$. Además, sea $\phi : [0, 1]^2 \rightarrow [0, 1]$ es un operador de agregación *overlapping* y $\varphi : [0, 1]^2 \rightarrow [0, 1]$ es una función de agrupación. Así, se define la medida de polarización como:

$$JDJ(X) = \sum_{i,j \in N, i < j} \varphi(\phi(\mu_{X_A}(i), \mu_{X_B}(j)), (\phi(\mu_{X_B}(i), \mu_{X_A}(j)))) \quad (5.1)$$

La medida JDJ recién presentada en la definición anterior 5.1.1 muestra sus valores máximos en aquellas situaciones en las que el 50 % de la población presenta el máximo grado de pertenencia hacia el polo e.g.: X_A , y el mínimo grado de pertenencia hacia X_B , y el otro 50 % de la población presenta el máximo grado de pertenencia hacia el polo X_B , y el mínimo grado de pertenencia hacia X_A . Por su parte, muestra sus valores mínimos en aquellos casos en los que toda la población no sólo presenta los mismos grados de pertenencia hacia ambos polos, sino cuando estos grados de pertenencia son extremos, es decir, o bien toda la población presenta el máximo grado de pertenencia hacia el polo e.g.: X_A , y el mínimo grado de pertenencia hacia X_B o viceversa.

Ejemplo 5.1.1. Para una comprensión en detalle de la fórmula propuesta previamente, se computa para cada par de individuos $i, j \in N$ el valor de polarización entre ambos.

- **Caso 1. Individuo i está cerca del polo X_A y j está cerca del polo X_B .** Así, se tiene que agregar a través de la función de agrupación φ dos valores: el grado en el que i pertenece a X_A y j pertenece a X_B , donde $a = \phi(\mu_{X_A}(i), \mu_{X_B}(j)) = \phi(1, 1) = 1$ y el grado en el que i

pertenece a X_B y j pertenece a X_A , teniendo $b = \phi(\mu_{X_B}(i), \mu_{X_A}(j)) = \phi(0, 0) = 0$.

Finalmente, se tiene que verificar cuál de estos dos hechos verídicos utilizando la función de agrupación $\varphi(1, 0) = 1$, teniendo, por lo tanto, un caso de alta polarización.

- **Caso 2. Individuo i e individuo j presentan posiciones neutras respecto con los polos X_A, X_B .** Así, se tienen dos individuos situados en el centro de la distribución. De nuevo, se agregan los dos resultados a través de la función de agrupación φ , siendo el grado en el que i pertenece al polo X_A y j al polo X_B , siendo $a = \phi(\mu_{X_A}(i), \mu_{X_B}(j)) = \phi(0.5, 0.5) = 0.5$ y viceversa, donde $b = \phi(\mu_{X_B}(i), \mu_{X_A}(j)) = \phi(0.5, 0.5) = 0.5$.

Se tiene, por lo tanto, que $\varphi(0.5, 0.5) = 0.5$, encontrando un caso de polarización media.

- **Caso 3. Individuo i está cerca del polo X_A e individuo j también está cerca del polo X_A .** Se agregan, a través de la función de agrupación φ , los dos resultados obtenidos, donde se mide la cercanía de i al polo X_A y de j al polo X_B , siendo $a = \phi(\mu_A(i), \mu_B(j)) = \phi(1, 0) = 0$ y viceversa, donde $b = \phi(\mu_B(i), \mu_A(j)) = \phi(0, 1) = 0$.

Finalmente, $\varphi(0, 0) = 0$, siendo un caso de polarización nula.

Observación 2. Nótese que ϕ es una función de agregación, por lo que cualquier incremento en alguno de los tres componentes incrementará los valores de polarización para $i, j \in N$ fijo.

Observación 3. Aquellas situaciones en las que una población N se divida en k grupos, habrá tantos grupos dados como diferentes valores asociados a la variable X (e.g.: $D_X = \{X_1, \dots, X_K\}$). En estos casos, la medida de polarización JDJ puede ser reescrita por cuestiones computacionales como:

$$JDJ_g(X) = \sum_{i=1, K} \sum_{j=1, K} \pi_i \pi_j \phi(\mu_{X_A}(i), \mu_{X_B}(j)) \quad (5.2)$$

Así, la única diferencia con la medida de polarización anterior es la forma en la que $I(i, j)$ es medida para cada par de grupos C_i, C_j .

Por su parte, ésta medida puede ser convertida en índice con el fin de facilitar su interpretabilidad, encontrando los valores mínimos en 0 y los máximos en 1. Para ello, se presenta la siguiente definición:

Definición 5.1.2 (Índice de bipolarización JDJ). Sea JDJ la medida de bipolarización presentada en 5.1.1 obtenida sobre una población de individuos

N . Por otro lado, sea NC el número de comparaciones por pares totales realizadas para su computación, es decir, $\frac{N*(N-1)}{2}$. Es posible obtener un índice entre 0 y 1 de la siguiente manera:

$$IJDJ = \frac{JDJ}{NC} * 2 \tag{5.3}$$

De manera similar, y si la fórmula que se tiene es la definida en 5.2 sobre la distribución de probabilidades asociadas a los grupos, el índice de polarización se define de la siguiente forma:

$$IJDJ_g = JDJ_g * 4 \tag{5.4}$$

5.1.3. Comparación entre medidas de polarización.

Uno de los *handicaps* de esta medida es como construir las funciones de pertenencia μ_{X_A} , μ_{X_B} para cada individuo. En esta sección se propone una vía para construir estas funciones de pertenencia cuando para cada individuo i se tiene una variable estadística que pueda ser ordinal o continua acotada.

En particular, supóngase una población N y el caso de una variable con dominio $D_X = \{1, 5\}$. En primer lugar, es necesario elegir las funciones de agrupación, *overlapping* y de pertenencia asociadas a la medida JDJ con el fin de posibilitar su aplicación. Por simplicidad, en este caso se utilizará como función de agrupación el operador de agregación “máximo”. Por su parte, referente a las funciones de *overlapping*, se aplicaran dos de las más famosas en la literatura, el “mínimo” y el “producto”. Finalmente, respecto con las funciones de pertenencia μ^T , obsérvese la siguiente figura 5.2:

$$(\mu_1^T(x), \mu_5^T(x)) = \left(1 - \frac{x-1}{4}, \frac{x-1}{4}\right) \tag{5.5}$$

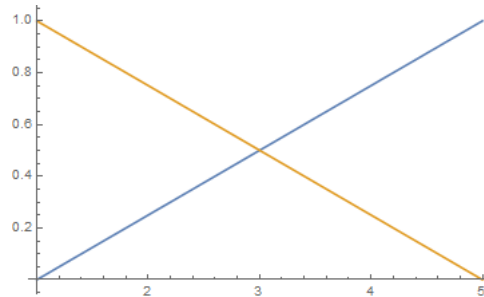


Figura 5.2: Función de pertenencia (μ_1^T en amarillo y μ_5^T en azul) a los polos $X_A = 1$ y $X_B = 5$.

Es importante resaltar la posibilidad de elegir diferentes funciones de pertenencia triangulares. En este sentido, se podría forzar la función μ_1 a 0 para aquellos casos o valores en X mayores que a , así como para μ_5 menores que b .

Finalmente, y referente a JDJ , en este experimento siempre se ha considerado $\varphi(u, v) = \text{Max}\{u, v\}$ y $\phi(x, y, z) = \text{Min}\{x, y, z\}$ o $\phi(x, y, z) = xyz$. De esta forma, con motivos ilustrativos, se han aplicado dos medidas JDJ diferentes, una en la que se ha utilizado el producto como función de agrupación con las funciones de pertenencia referentes a la figura 5.2 - de aquí en adelante, JDJ_PRO - y otra en la que se usó el mínimo como función de agrupación con la función de pertenencia expuesta en el segundo caso siendo $a = 3$ y $b = 3$ - de aquí en adelante, JDJ_MIN -.

Así, se realiza un experimento en el que se reproduce una muestra de $N = 391315$ casos, en los que cada uno de ellos, representa todas las diferentes distribuciones de frecuencias relativas posibles a lo largo de una escala tipo *likert*, donde $X = \{1, 2, 3, 4, 5\}$, variando dicha probabilidad de un caso a otro en 0.1. De esta forma, se entiende por un caso como la distribución relativa de una población a lo largo de los diferentes valores de X . Una vez generados los casos, se aplicaron las dos medidas JDJ_PRO y JDJ_MIN mencionadas en el párrafo anterior, junto con la medida propuesta por Esteban y Ray - ER -, así como el índice de variación ordinal IOV , encontrando los siguientes resultados en la tabla 5.1:

Tabla 5.1: Estadísticos descriptivos de IOV , ER , JDJ_Pro y JDJ_MIN .

	Media	DT	Mediana	Recortada	Mad	Min	Max	Rango	Simetría	Curtosis
IOV	0.60	0.17	0.60	0.60	0.19	0.00	1.00	1.00	-0.15	-0.49
ER	0.37	0.10	0.35	0.36	0.10	0.00	1.00	1.00	0.80	1.17
JDJ_PRO	0.75	0.13	0.76	0.76	0.12	0.00	1.00	1.00	-0.85	1.23
JDJ_MIN	0.28	0.18	0.27	0.27	0.20	0.00	1.00	1.00	0.41	-0.37

Así mismo, como los siguientes histogramas sugieren - figura 5.3 -, se encuentra una asimetría opuesta entre las puntuaciones provenientes de la medida ER frente a los valores ofrecidos por JDJ_PRO . Estas diferencias subyacen a aspectos centrales de ambas medidas que difieren entre sí, desde un punto de vista conceptual como métrico. Por otro lado, gracias a los histogramas, se puede observar el comportamiento de la función de pertenencia de JDJ_MIN , encontrando un comportamiento distinto al del resto de medidas.

Por otro lado, se puede observar que existen correlaciones altas entre las puntuaciones de IOV , ER , JDJ_PRO y JDJ_MIN - ver 5.4 -, encontrando los mayores valores de correlación para las medidas JDJ_PRO y

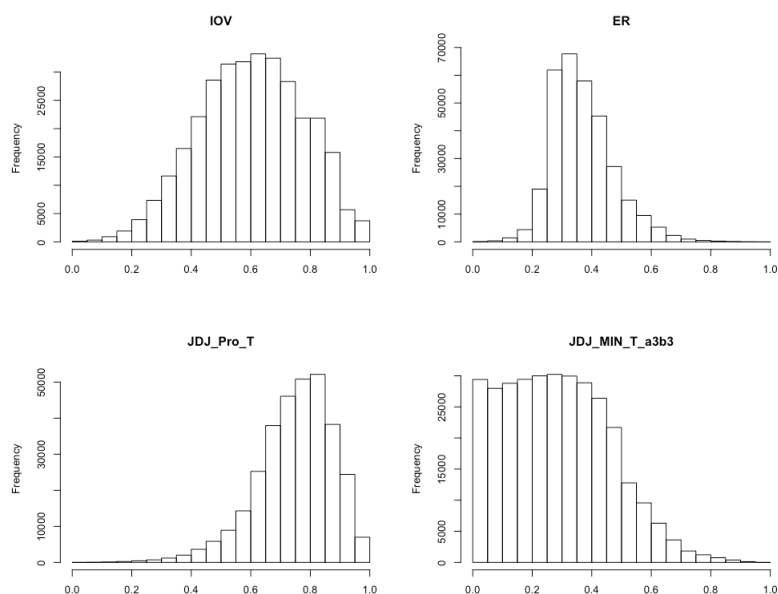


Figura 5.3: Histogramas para las puntuaciones bidimensionales provenientes de *IOV*, *ER*, *JDJ_PRO* y *JDJ_MIN*.

JDJ_MIN con *IOV*, mientras que los valores más bajos de correlación se dan con *JDJ_PRO* y *ER*.

En adición, en la figura 5.5 se observa la relación observada entre cada medida de polarización con los valores del índice de *IOV* agrupados por deciles. Se observa una tendencia natural a encontrar mayores niveles de polarización con altos valores de *IOV*, hecho que guarda estrecha relación con los valores de correlación encontrados en la figura 5.4. Resulta interesante señalar la tendencia encontrada por parte de la medida propuesta por Esteban y Ray, encontrando una porción significativa de valores medios para el primer decil de *IOV*, siendo el mismo un signo de sesgo por parte de la medida en tales escenarios.

Por último, se ha segmentado toda la población en submuestras para ver cómo responde cada medida de polarización a diferentes distribuciones. Esta segmentación se realizó de acuerdo con un criterio de evolución, esto es, desde la selección del 90% de la población agrupada en torno a un polo, a distribuirse a lo largo de la variable X y, finalmente, agruparse en dos grupos opuestos en torno a los polos -, dejando el resto del 10% de la población, para cada caso, repartida en el resto de categorías no referentes al criterio de evolución. La figura 5.6 muestra los valores medios de polarización para cada caso donde la población se agrupa en un polo extremo y se mueve

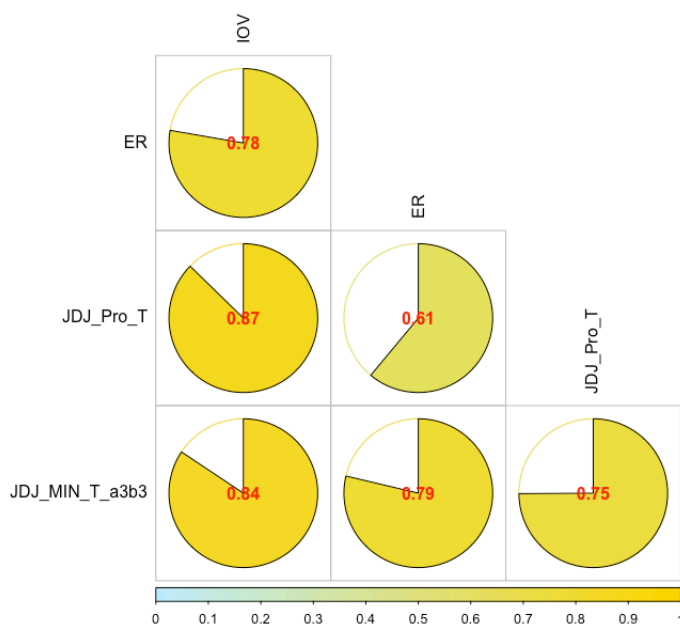


Figura 5.4: Correlaciones entre IOV , ER , JDJ_PRO y JDJ_MIN .

progresivamente hacia el centro y, finalmente, hacia ambos polos extremos. De acuerdo con el gráfico, encontramos una tendencia de las tres medidas de polarización a tomar valores más bajos cuando los individuos se distribuyen en torno a un único polo - caso 1, Figura 5.6 -.

Por otro lado, el caso 5 parece incorporar importantes conceptos clave sobre la polarización, representando el caso de máxima fragmentación, situación relacionada con altos niveles de IOV . Este caso presenta un mayor riesgo de polarización de acuerdo con la medida JDJ_PRO , especialmente en comparación con el caso 8, donde se aprecia la existencia de dos grupos diferentes agrupados alrededor del centro de la variable. A pesar de ser dos grupos diferentes, no tienen por qué estar en conflicto entre sí por su cercanía al centro, donde la comunicación no estará en riesgo, reflejando dichos aspectos en las medidas JDJ , mientras que ER se mantiene similar. Este aspecto parece de especial relevancia, en tanto que dos situaciones completamente diferentes deberían de mostrar valores diferentes. Por otro lado, el riesgo de polarización de acuerdo con la evolución de la distribución a lo largo de los valores de X , encontrando un aumento de la polarización según la formación de dos grupos diferentes se acercan, de forma opuesta y contraria, a los polos (casos 7 a 9) obteniendo los valores más altos en el caso de 9 para todas las medidas de polarización.

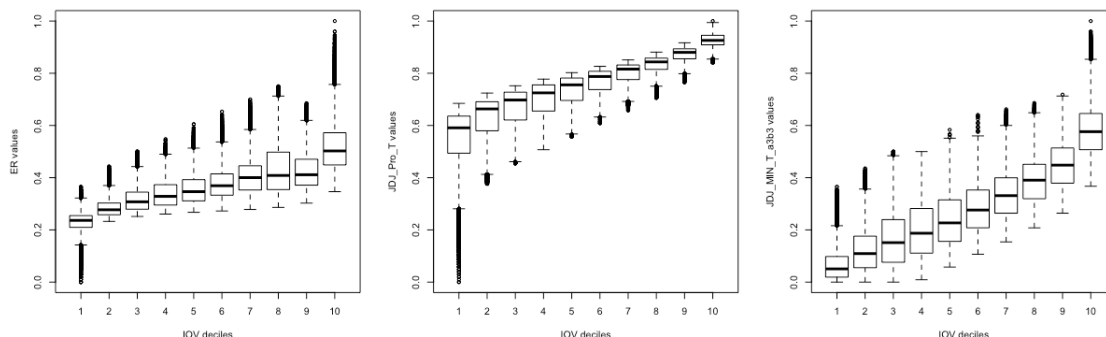


Figura 5.5: Relación entre IOV - en deciles - y las puntuaciones observadas de ER , JDJ_PRO y JDJ_MIN .

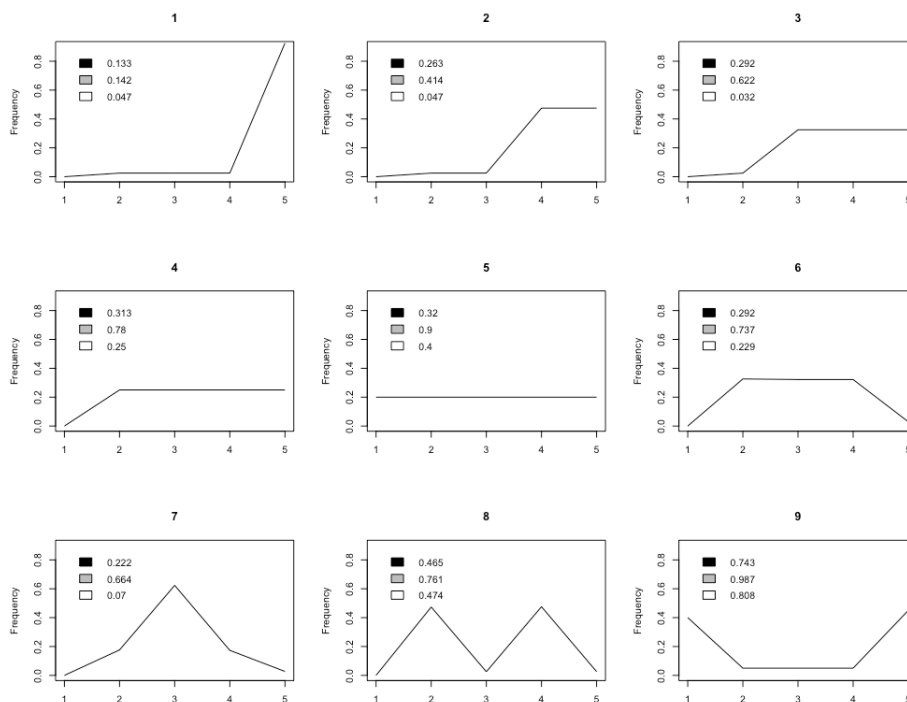


Figura 5.6: Valores medios para cada medida de polarización (ER -negro-, JDJ_PRO -gris- and JDJ_MIN -blanco-) para diferentes distribuciones de X . (1) 90% de los individuos están agrupados en torno al valor 5. (2) 45% de los individuos están agrupados en torno al valor 5 y el 45% el valor 4. (3) 30% de los individuos están agrupados en torno al valor 5, el 30% en el valor 4 y el 30% en el valor 3. (4) 22.5% de los individuos están agrupados en torno a cada uno de los valores del 2 al 5. (5) 18% de los individuos están agrupados en torno cada uno de los valores de X . (6) 30% de los individuos están agrupados en torno al valor 2, 30% en el valor 3 y el 30% en el valor 4. (7) 15% de los individuos están agrupados en torno al valor 2, 60% en el valor 3 y el 15% en el valor 4. (8) 45% de los individuos están agrupados en torno al valor 2 y 45% en el valor 4. (9) 45% de los individuos están agrupados en torno al valor 1 y 45% en el valor 5.

5.1.4. Conclusiones.

El concepto de polarización es rico y complejo, y se encuentra una necesidad de encontrar una aproximación que incluya en su modelo matemático las propiedades métricas y conceptuales adecuadas. Además, de acuerdo con lo anterior, para aquellos casos en los que no se es posible contar con toda la información referente a la polarización - e.g. comunicación entre grupos -, en este trabajo se propone hablar de riesgo de polarización.

En esta investigación, se propone una medida de riesgo de polarización basada en los conjuntos borrosos. Los resultados analizados arrojan información de vital interés, como la alta correlación encontrada entre las medidas de polarización y la variación. Si bien, a lo largo de la literatura, se ha entendido en ocasiones la polarización como sinónimo de variabilidad o fragmentación, es importante resaltar la necesidad de separar ambos conceptos, tal y como se ha podido comprobar, a pesar de su estrecha relación.

En este trabajo, se ha podido comprobar el comportamiento de la medida de riesgo de polarización JDJ . Se observa que JDJ muestra sus valores más altos de polarización en aquellos escenarios en los que el 50 % de la población presenta $\mu_{X_A} = 1$ y $\mu_{X_B} = 0$ a la vez que el otro 50 % de la población presenta funciones de pertenencia opuestas y extremas al anterior grupo. Por el contrario, es interesante mencionar que el caso de polarización nula se da en aquellos escenarios en los que el 100 % de la población, no sólo presenta exactamente los mismos niveles de grados de pertenencia hacia los polos, sino que estos grados de pertenencia son grados extremos, y no moderados. De esta forma, se hace una distinción no encontrada anteriormente en la literatura de la medición de la polarización. Se considera que el riesgo de polarización para una población es más alto para aquellos casos en el que toda la población presenta el mismo valor en la variable X , pero esta posición refleja una moderación, frente aquellos casos en los que la población entera está radicalizada en torno a un polo. Tal y como se puede observar en la figura 5.7, la distancia que debe de recorrer la población para llegar a la situación de máxima polarización - dos grupos opuestos y antagónicos, de mismo tamaño, situados en torno a los polos - es menor en el caso A frente al caso B. Además, ésta perspectiva, no sólo es más realista desde un punto de vista métrico, sino conceptual y actitudinal, siendo más factible la división de una población unida que presenta valores moderados o indecisos - el centro de la variable - que la polarización de una población radicalizada en torno a una posición extrema.

Por otro lado, en esta investigación se ha propuesto una nueva metodología a través de la cual medir la polarización. Los conjuntos borrosos proporcionan los recursos necesarios para afrontar la medición de la polari-

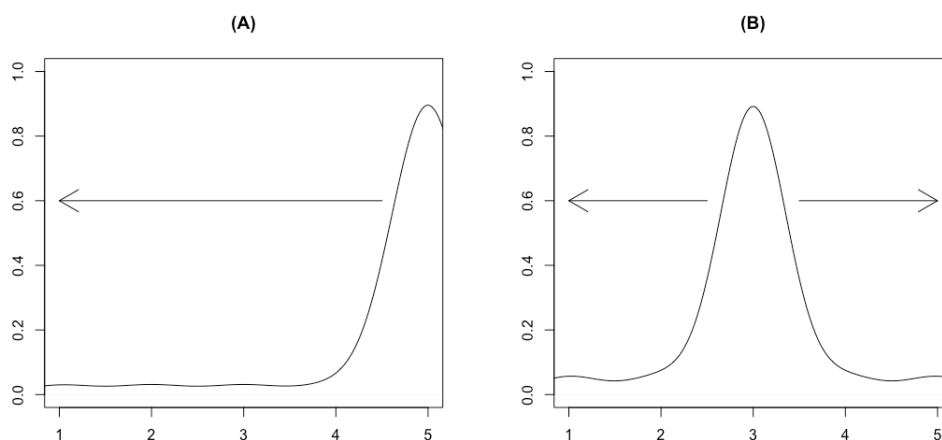


Figura 5.7: Distancia a recorrer por una distribución para llegar a un escenario de polarización máxima.

zación desde una perspectiva novedosa y con alto potencial. Así, a diario, las personas no se identifican con un único grupo político o actitudinal, sino que sienten diferentes grados de pertenencia a diversos grupos, evitando la dicotomía forzada encontrada tradicionalmente en la literatura, perdiendo gran cantidad de matices. Gracias a la naturaleza de los conjuntos borrosos, estos matices pueden ser recogidos por una medida de polarización con el fin de reflejar, de una forma más acertada y ajustada a la realidad, las actitudes de las personas, permitiendo la medición de la polarización o discrepancia entre un determinado par de individuos de una forma más ajustada a la realidad. Gracias al uso de operadores de agregación y funciones de pertenencia, los conjuntos borrosos permiten un nuevo camino para perseguir esta filosofía.

Como conclusión extraída de los resultados expuestos, se puede concluir la gran importancia, y protagonismo, que representan las funciones de pertenencia en el comportamiento de la medida JDJ . Como consecuencia, es importante enfatizar la necesidad de profundizar en el estudio del rendimiento proporcionado por diferentes funciones de pertenencia, así como encontrar las funciones de pertenencia adecuadas para determinados casos.

Para finalizar, se sugieren algunas direcciones para investigaciones futuras. En relación con las funciones de pertenencia, se considera de vital importancia profundizar en el estudio de las mismas. Así mismo, un siguiente paso natural al camino iniciado en esta investigación es el de plantear la medición de la polarización desde una perspectiva dinámica a través del tiempo, así como una conceptualización multidimensional de la polarización.

5.2. Aplicación en un caso real: Cuando la negatividad es el combustible. Bots y Polarización Política en el debate del COVID-19.

Robles, J., Guevara, J.A., Casas-Mas, B., & Gómez, D. (2022). When negativity is the fuel. Bots and Political Polarization in the COVID-19 debate. *Comunicar*, 71. <https://doi.org/10.3916/C71-2022-05>

Resumen.

La inesperada crisis del coronavirus llevó al gobierno de España a implementar un estado de alarma en el país. Estos hechos, llevados a cabo por un gobierno de coalición recién formado, consecuente de una fuerte división de opiniones de voto entre los españoles, supusieron el inicio de un fuerte debate en la red social Twitter sobre la gestión del nuevo gobierno. Los objetivos de este estudio son los siguientes: (1) detectar y medir la polarización a lo largo del debate online en la red social Twitter en torno a la opinión hacia el gobierno de España; (2) identificar los *bots* en la red en función a su afinidad política y conocer su papel en la polarización del debate (3) analizar los temas de debate y su influencia en la polarización, así como su presencia en el discurso de los *bots*. Para ello, se descargaron los datos de la API de Twitter y se aplicaron algoritmos de aprendizaje automático supervisado para, tras una codificación previa, etiquetar los mensajes como a favor en contra o neutrales hacia el gobierno. Para alcanzar estos objetivos, se aplicó la medida de polarización JDJ (Guevara et al., 2020) [50], el algoritmo de detección de *bots* tweetbotornot (Keraney, 2020) [64] y algoritmos de aprendizaje automático supervisado para detectar el estilo de comunicación de los partidos políticos en *bots*. Finalmente, se aplicó el algoritmo de detección de topics *Latent Dirichlet Allocation (LDA)* para detectar los temas de debate en el discurso producido en Twitter. Se encontraron altos niveles de polarización, siendo mayores los niveles en los usuarios *bots*. Se detectaron tres temáticas en el debate - política, economía y sanidad -, siendo la política y la economía las que mostraron mayores niveles de polarización. Así mismo, se observó que el debate generado por los usuarios era principalmente negativo.

Abstract.

The unexpected COVID-19 crisis forced the Spanish government to decree the state of alarm. This announcement, which was made by a newly formed coalition government consequence of a strong division of voting opinion, supposed the beginning of a strong debate on Twitter about the management

of this new government. The aims of this study are the following: (1) to detect and measure the political polarization during the online debate on Twitter towards the Spanish government; (2) to identify bots according to their political affinity and to know their role in the development of this debate and polarization; (3) to detect and analyse the hot topics during the debate as well as their influence in the phenomenon of polarization. To accomplish that, we downloaded our dataset from the Twitter API and machine learning algorithms were applied in order to label all tweets as supporters, neutral or detractors towards the Spanish government. We applied the polarization measure JDJ (Guevara et al., 2020) [50], the bot detection algorithm tweetbotornot (Kearny, 2020) [64] and machine learning algorithms to detect the political communication style of those tweets produced by bots. Furthermore, to detect the hot-topics we applied the Latent Dirichlet Allocation (LDA) algorithm. Preliminary results show the presence of political polarization stressed by the role of bots, during the debate on Twitter towards a few important topics such as economy, crisis or health. We found high levels of polarization, specifically in those users categorized as bots. Also, we detected three different topics into the digital debate - healthcare, economics and politics -, finding higher levels of polarization in politics and economics. Finally, we found out that the digital debate generate by the users was negative in nature.

5.2.1. Metodología.

A continuación, se presenta un esquema de la metodología llevada a cabo en este estudio.

- Planteamiento de los objetivos de estudio.
- Extracción de datos y limpieza.
- Identificación de *bots*.
- Filtración de la Base de Datos.
- Aplicación de algoritmos de aprendizaje automático para la creación de categorías y construcción de las funciones de pertenencia para la medida de polarización.
- Análisis de sentimiento.
- Detección de *topics*.
- Medición de la polarización.

5.2.1.1. Objetivos de estudio.

- O_1 : Conocer los niveles de polarización presentes en el debate digital en Twitter durante el estado de alarma.
- O_2 : Detectar cuentas automatizadas, o *bots*, así como determinar su papel en el debate digital.
- O_3 : Conocer las diferentes temáticas que tuvieron en el debate digital y su relación con los diferentes tipos de cuentas - *bots* o no *bots* -, así como su relación con los niveles de polarización encontrados.

5.2.1.2. Caso de estudio.

El caso de estudio se centró en la gestión económica, sanitaria y política de la pandemia por la COVID-19 por parte del gobierno central de España. El 15 de marzo de 2020 se decretó el estado de alarma en España que iría acompañada por un estricto confinamiento en los hogares que se iría desvaneciendo a través del denominado “plan de desescalada”. Durante este período, España, y el mundo, se enfrentaron a una situación socio-sanitaria como no se había visto en los últimos años. A éste suceso de cuarentena en los hogares se sumó el gran incremento de uso de las redes sociales online por parte de la población. Como consecuencia, el debate político, sanitario, económico y social se trasladó de la esfera a tradicional a la digital, donde la población, a través de las redes sociales debatieron sobre la gestión de la pandemia por parte del gobierno.

5.2.1.3. Fuente de datos y limpieza.

Para la descarga de datos se eligió la red social Twitter por representar una de los espacios digitales que permiten una mayor facilidad de consumir contenido digital, así como de publicarlo. Para ello, se hizo uso de la API de Twitter a través de R-Studio a través de la librería *rtweet* propuesta por Kearney (2019) [63]. Para proceder con la descarga, se diseñó una bolsa de palabras con el fin de acceder a la mayor parte de las publicaciones en la red que pertenezciesen a nuestro objeto de estudio. Para ello, se seleccionaron todas las cuentas de los principales partidos políticos de España, recogiendo así sus propias publicaciones, como las menciones a los mismos. Además, se incluyeron sus principales representantes y los términos “estado de alarma”, “coronavirus” y “COVID”. La fase de descarga se realizó a razón de cinco tandas con el fin de cubrir todo el marco temporal del suceso, cada una realizada durante la primera semana del cambio de fase referente a la cuarentena. Tras el proceso de descarga, se obtuvieron un total de 4895747 tweets.

Como consecuencia de las características de la fase de descarga, es común la recolección de mensajes que no pertenecen a la temática de estudio como consecuencia del posible doble significado de determinados términos, como por ejemplo, “podemos” y “ciudadanos”. Para eliminar datos que produjesen ruido en los resultados, se decidió entrenar algoritmos de aprendizaje automático supervisado con el fin de entrenar a los mismos a que detectasen qué contenido textual no se refiere a los objetivos de estudio de la presente investigación. Para ello, se entrenaron tantos modelos como diferentes tandas de descarga se encontraron con el fin de evitar sesgos en el aprendizaje como consecuencia del dinamismo encontrado en el debate digital a lo largo de las fases de desescalada. Así, se realizó una muestra aleatoria simple de 1500 mensajes por tanda que fueron codificados por un experto manualmente, a razón de “pertenece” o “no pertenece” a nuestro objetivo de estudio. Por consiguiente, se procedió a el entrenamiento de diferentes algoritmos - e.g.: naive bayes, redes neuronales o máquinas de soporte vectorial, entre otros -, siendo las máquinas de soporte vectorial las que mostraron un mejor rendimiento en la clasificación de texto - ver tabla 5.2 -, encontrando una medida F media de 0.768 entre tandas, así como una precisión media de 0.8.

Tabla 5.2: Resultados del clasificador SVM para limpieza de datos.

Tanda	Precisión	Sensibilidad	Especificidad	Kappa	F-Score	AUC
1	0.8017	0.9322	0.3846	0.3670	0.8620	0.6583
2	0.8167	0.5476	0.9213	0.5077	0.6556	0.7344
3	0.8267	0.7027	0.8995	0.6187	0.7596	0.8010
4	0.7867	0.7090	0.8494	0.564	0.7457	0.7791
5	0.7659	0.8758	0.6377	0.5216	0.8171	0.7567

Finalmente, tras la aplicación de los algoritmos entrenados, codificaron todos los mensajes a razón de “pertenece” o “no pertenece” a nuestro objetivo de estudio y se eliminaron aquellos que no eran deseados. Como consecuencia, se obtuvo una base de datos final con un total de 1208631 tweets publicados por 469616 usuarios.

5.2.1.4. Detección de *bots*.

Con el fin de poder detectar todas aquellas cuentas automatizadas y controlar las mismas en los resultados, se procedió a la aplicación de algoritmos de detección de *bots*, especialmente el propuesto por Kearney (2020) [64], incluido en la librería de R denominada *tweetbotornot2*. En concreto, y debido a la gran cantidad de usuarios, se decidió optar por la opción del algoritmo

Fast-gradient boosted. Este algoritmo se basa en diferentes variables proporcionadas por la red social Twitter, como actividad de la cuenta o últimas publicaciones como criterio de detección de *bots*. Por último, se considera de vital importancia optar por un criterio conservador en la fase de detección de cuentas como *bots* o no. De esta forma, únicamente se seleccionaron como cuentas *bots* los perfiles que presentaron una probabilidad situada en el mayor cuartil de la probabilidad.

5.2.1.5. Medición de la polarización.

Para la medición de la polarización se ha aplicado la medida presentada en este capítulo y que fue publicada en 2020 [50]. Esta medida se puede interpretar como el riesgo o la probabilidad de polarización máxima para todos y cada uno de los pares de individuos presentes en la población del estudio. Para facilitar su interpretación, se presenta la modificación de la medida para su representación como índice, encontrando el valor mínimo en 0 y el máximo en 1, ya presentado en 5.1.2.

$$IJDJ = \frac{JDJ}{NC} * 2$$

5.2.1.6. Detección de *topics*.

Para cubrir uno de los objetivos propuestos en esta investigación, se aplicaron algoritmos de detección de *topics*, los cuales permiten extraer, inherentes al texto, las diferentes temáticas presentes en el discurso de acuerdo con el grado de cercanía que presentan las palabras que lo componen. En definitiva, estos algoritmos son modelos de agrupación. En concreto, se aplicó el algoritmo denominado como *Latent Dirichlet Allocation (LDA)*, disponible en la librería de R llamada *topicmodels*, propuesta por Grün y Hornik (2011) [58]. La aplicación del algoritmo recién citado supone la selección *a priori*, del número de *topics* a detectar. Para la correcta selección de este parámetro, se llevaron a cabo una medida de coherencia semántica que contrastan, para diferentes soluciones de acuerdo con la selección de diferentes números de temáticas, el grado de similitud semántica entre todas aquellas palabras que son agrupadas en torno a una determinada temática. Dado que dos palabras pertenecientes a un mismo *topic* no tienen por qué presentar similitud semántica - e.g.: sanidad y hospital -, se seleccionó un experto que examinó cada una de las diferentes soluciones que, de acuerdo a su criterio y las medidas de concordancia, se seleccionaría un número de *topics* definitivo.

5.2.1.7. Análisis de sentimiento.

Se optó por la realización de análisis de sentimiento basados en la aplicación de diccionarios con el fin de conocer el tono emocional del debate digital en Twitter. Para ello, se optó por el diccionario *afinn*, propuesto por Hansen et al. (2011) [56] el cual se basa en la detección de contenido positivo y negativo asociado a un texto determinado. Este diccionario está compuesto por un total de 2477 términos que presentan una determinada puntuación de 1 a 5, de acuerdo la fuerza de asociación a un determinado sentimiento, ya sea positivo o negativo.

5.2.2. Análisis y resultados.

5.2.2.1. Clasificación de mensajes como *a favor* y *en contra* hacia la gestión del gobierno.

Para llevar a cabo este objetivo, a la base de datos final se aplicó el mismo procedimiento llevado a cabo para el filtro de la base de datos, creando una nueva variable categórica de dos niveles: *a favor* y *en contra* hacia el gobierno. Para ello, se procedió en primer lugar a la *tokenización* del texto es decir, se dividió el contenido textual en los diferentes términos que lo componen. A continuación, se eliminaron las *stopwords*, aquellas palabras que carecen de información de utilidad, como determinantes, artículos, preposiciones, etc. Finalmente, con el fin de crear una base de datos apta para los algoritmos de aprendizaje automático, se transformó el texto *tokenizado* en una matriz de frecuencia/término, denominadas matrices *dfm* - *document feature matrix* -, en especial, se utilizaron matrices *tf-idf*, las cuales ponderan la aparición de un término por la frecuencia inversa de su aparición en todo el documento. Tras aplicar los algoritmos de aprendizaje automático, tal y como ocurrió en la fase de limpieza, las máquinas de soporte vectorial mostraron el mejor rendimiento, considerando el mismo satisfactorio para una correcta clasificación - ver tabla 5.3 -.

Tabla 5.3: Resultados del clasificador SVM para la codificación *a favor* y *en contra* hacia la gestión del gobierno.

Tanda	Precisión	Sensibilidad	Kappa	F-Score	AUC
1	0.8492	0.9854	0.4816	0.9122	0.6950
2	0.8960	0.9619	0.7761	0.9277	0.8780
3	0.8392	0.8488	0.6675	0.8439	0.8366
4	0.9133	0.9048	0.8225	0.9090	0.9121
5	0.8318	0.8600	0.6638	0.8456	0.8335

5.2.2.2. Detección de *bots*.

A continuación, se procedió a la aplicación de algoritmo de detección de *bots* *Fast - gradient boosted* de acuerdo con el criterio conservador de clasificar un determinado usuario como *bots* si su probabilidad de serlo se sitúa por encima del mayor cuartil de la probabilidad, siendo esta > 0.98754517674 . De esta forma, del total de 469616 usuarios, 69033 fueron identificados como cuentas automatizadas. En otras palabras, el 15 % de los usuarios de la base de datos resultaron ser *bots*, los cuales publicaron un total de 172704 mensajes del total - 1208631 -, suponiendo un 14.28 % del contenido publicado durante el caso de estudio.

5.2.2.3. Medición de la polarización.

Como se mencionó en este capítulo en la presentación y propuesta de la medida difusa de polarización *JDJ*, un elemento indispensable para el cálculo de la misma es la disponibilidad de los grados de pertenencia a una determinada categoría. En este caso, se medirá la polarización de acuerdo con la posición de los usuarios hacia la gestión del gobierno, por lo que se contemplarán dos polos - *a favor* y *en contra* - para el cálculo de la polarización. Bajo esta perspectiva, cada usuario presentará dos grados de pertenencia, uno a estar *a favor* y otro a estar *en contra*. Como grados de pertenencia se utilizarán las probabilidades que ofrece el algoritmo de aprendizaje automático a que un determinado objeto pertenezca a una categoría u otra. Sin embargo, en tanto que se quiere calcular la polarización de los usuarios en Twitter, y los clasificadores automáticos codifican tweets, y no usuarios, se calculó, para un determinado usuario, la media de las probabilidades de sus mensajes publicados a pertenecer a una u otra categoría. De esta forma, para un determinado usuario i , únicamente se presentan dos valores, $\mu_{favor}(i)$ y $\mu_{contra}(i)$.

Así mismo, para el cálculo de la polarización, y debido a los costes computacionales de hacer comparaciones por pares para 469616 usuarios, se calculó el valor del índice *JDJ* como la media de 1500 iteraciones para el cálculo de *JDJ* a una muestra aleatoria simple de $N = 200$ por cada iteración. De esta forma, se procedió a la computación del índice de *JDJ* para la muestra en general, sin discriminar por tipo de usuario - *bot* o no *bot* -, encontrando un nivel de polarización de $JDJ_{mean} = 0.76$, $sd = 0.027$, donde $JDJ_{mean} \rightarrow [0, 1]$. En la figura 5.8 se puede observar la función de densidad de las funciones de pertenencia que presenta la población a estar *a favor* y *en contra*, donde valores > 0 se asocian a la posición de estar *a favor* del gobierno μ_{contra} , mientras que por el contrario, valores < 0 indican los grados de pertenencia hacia la posición *en contra*, representadas como $-\mu_{favor}$ para facilitar la

visualización de ambas posturas contrarias.

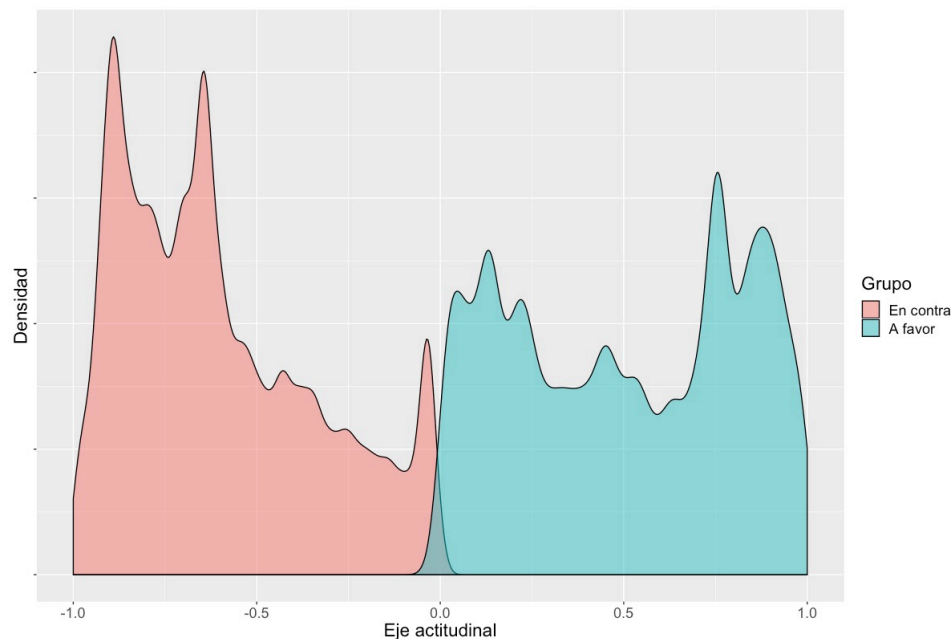


Figura 5.8: Polarización.

Finalmente, para conocer si existen, o no, diferentes niveles de polarización de acuerdo con el tipo de cuenta, se procedió a calcular la polarización para los dos grupos presentes en la población, encontrando niveles de $JDJ_{mean}(\text{no bots}) = 0.761$, $sd = 0.026$ y $JDJ_{mean}(\text{bots}) = 0.765$, $sd = 0.026$. Para comparar ambos resultados, se aplicó la prueba de Mann-Witney para dos variables independientes, encontrando un estadístico $U = 1021715$, con un nivel de significación $p < 0.00$. Así, cabe concluir que existen diferencias estadísticamente significativas entre los niveles medios de polarización para ambos grupos, encontrando mayores niveles de polarización para el grupo de *bots*.

5.2.2.4. Detección de *topics* y polarización.

Recuérdese que para la aplicación del algoritmo *LDA*, es necesario la indicación de número de temáticas presentes en el texto *a priori*. Para ello, se aplicó el algoritmo 10 veces, cada una de ellas cambiando el parámetro “número de grupos” de 0 a 10, calculando posteriormente la coherencia encontrada para cada una de las soluciones. Tras la evaluación por parte de un experto de cada uno de los diferentes resultados encontrados por el algoritmo, se determinó la opción de considerar tres temáticas, con una congruencia de 0.42. En la

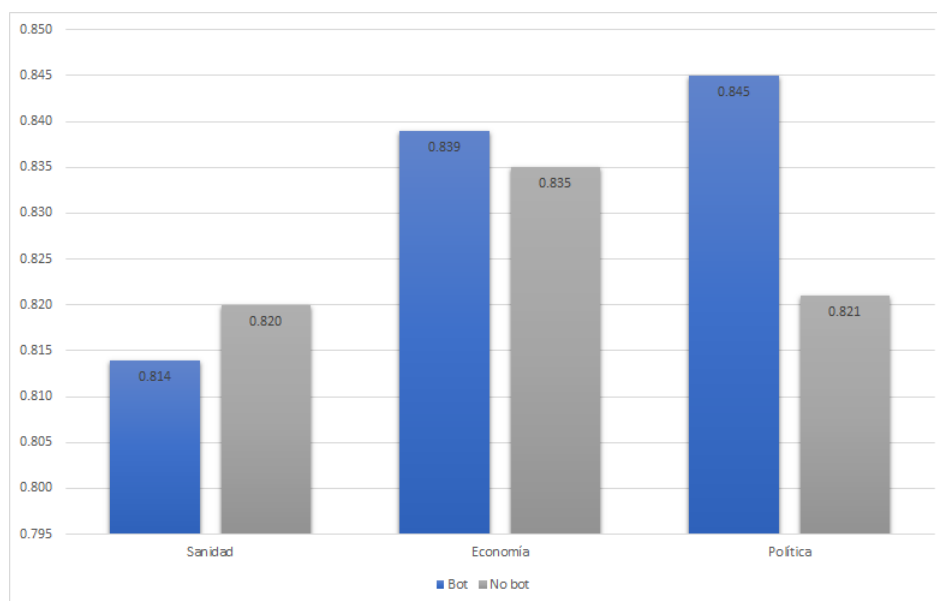


Figura 5.10: Polarización por temática y tipo de cuenta.

Para determinar si los diferentes niveles de polarización presentan diferencias estadísticamente significativas entre ellos se aplicó un ANOVA de 2 factores - (*bot* y *no bot*) x *temática* (“sanidad”, “economía” y “política”). Se encontraron diferencias estadísticamente significativas para ambos factores, así como para su interacción - ver tabla 5.4 -, concluyendo que el tipo de usuario y la temática afectan a los niveles de polarización, así como que los niveles de polarización encontrados para cada nivel de *temática* están condicionados por el tipo de cuenta de los publica.

Tabla 5.4: Pruebas de efectos inter-sujetos.

Variable dependiente: Polarización						
Origen	SC III	gl	MC	F	Sig.	
Modelo corregido	1,123a	5	,225	327,641	,000	
Intersección	6194,468	1	6194,468	9032376,674	,000	
USER	,121	1	,121	176,075	,000	
TOPIC	,668	2	,334	486,967	,000	
USER * TOPIC	,335	2	,167	244,100	,000	
Error	6,168	8994	,001			
Total	6201,759	9000				
Total corregido	7,292	8999				

a. R al cuadrado = ,154 (R al cuadrado ajustada = ,154)

Por último, y con el fin de determinar qué escenario de los posibles presenta mayores niveles de polarización respecto a los otros, se creó una nueva variable categoría de 6 niveles fruto de la combinación de tipo de cuenta y temática. De esta forma, se procedió a la aplicación de un ANOVA de 1 factor, encontrando un estadístico $F_5 = 327.641$, $p < 0.000$. Posteriormente, se realizó la prueba de Tukey de comparaciones múltiples, concluyendo que existen diferencias estadísticamente significativas entre todos los niveles de la variable a excepción de las comparaciones entre “no *bot* - sanidad” y “no *bot* - política”.

5.2.2.5. Análisis de sentimiento y *topics*.

Por último, se hizo uso del diccionario de sentimiento *afinn* para las publicaciones presentes en cada una de las temáticas. Tal y como se puede observar en la figura 5.11, se encuentra una predominación de contenido negativo a lo largo del debate digital. En concreto, se encontraron los mayores niveles de negatividad en las publicaciones generadas por cuentas no *bots* en el *topic* “economía” - 0.704% -, seguido por las publicaciones provenientes de las cuentas automatizadas cuando hablan de “política” - 0.687% - y seguidas por la temática “política” por cuentas no *bots* - 0.672% -.

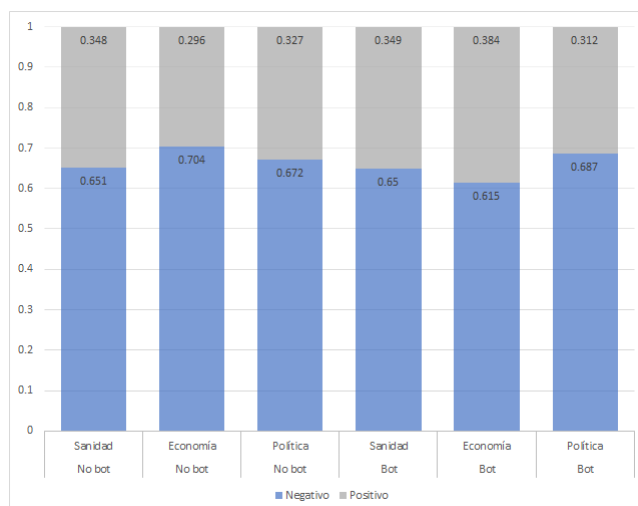


Figura 5.11: Proporción de sentimientos por tipo de cuenta y temática.

5.2.2.6. Conclusiones.

En la última década, las redes sociales se han convertido en espacios de comunicación digital donde miles de usuarios sustentan la oportunidad de

intercambiar información y opiniones de forma instantánea. Las características que ofrecen estos espacios digitales facilitan la aparición y propagación de fenómenos que pueden suponer una vulnerabilidad para el bienestar social. Además, escenarios de crisis como el presentado por el COVID-19 han planteado situaciones de riesgo donde los agentes políticos tienden a extremar su posicionamiento en la red a través de diversas estrategias, como el uso de *bots*. Estos actos no quedan exentos de repercusión, convirtiéndose en agentes de polarización política entre los usuarios de las redes sociales.

A lo largo de esta investigación se ha observado como la existencia de dos polos opuestos y contrarios a lo largo de un eje actitudinal facilitan la aparición de fenómenos como el de polarización, creando una ruptura comunicacional en la población. Tal ruptura se caracteriza por la fuerte división o polarización del debate digital, donde la actuación de cuentas automatizadas supone una fuerte influencia, actuando como guía de debate público sujetos a los intereses de quienes las diseñan. Así, se ha contrastado que los *bots* presentan una mayor tendencia a dividir la opinión pública que los usuarios corrientes. Así mismo, se ha observado como los *bots* centran el debate en aquellas temáticas que presentan una mayor dificultad para la justificación objetiva de argumentos, como es el caso de la política, ocupando prácticamente el 50 % del debate de los mismos, frente a temáticas que presentan una mayor cercanía a criterios científicos, como lo puede ser la economía o la sanidad. Bajo esta estrategia, las cuentas automatizadas sesgan el debate digital con el fin de crear ruptura en la población, creando altos niveles de polarización.

Con ello, se puede concluir que el uso de *bots* no está sujeto a la información de la sociedad (Al-Rawi y Shukla, 2020) [2] sobre la pandemia, sino centrado en la movilización negativa que propicia la ruptura de la opinión pública.

En este estudio se ha podido contrastar como, efectivamente, escenarios de crisis presentan una co-ocurrencia con la aparición de polarización política en las redes. Sin embargo, se puede concluir que el origen de este hecho no está ligado únicamente a las características del escenario y las redes sociales, sino a la intervención de agentes políticos tradicionales que, haciendo usos de campañas de marketing digital, radicalizan las actitudes de los usuarios de internet a través de la creación de cuentas automatizadas.

5.3. Otras aplicaciones: Medida de polarización *JDJ* y la mejora de algoritmos de detección de comunidades

Gutiérrez, I., Guevara, J. A., Gómez, D., Castro, J., & Espínola, R. (2021). Community Detection Problem Based on Polarization Measures: An Application to Twitter: The COVID-19 Case in Spain. *Mathematics*, 9(4), 443.

Resumen.

En esta investigación se aborda una de las problemáticas más importantes en el campo del análisis de redes sociales: los problemas de detección de comunidades. Esa información adicional es modelada por una medida borrosa que representa el riesgo de polarización. Particularmente, se tendrá en cuenta la polarización que se da entre nodos para detectar las comunidades de una red. Incluir este tipo de información a los problemas de detección de comunidades hace más realistas las soluciones encontradas, en tanto que una comunidad presentará bajos niveles de polarización y los elementos necesarios para mantener un diálogo pacífico. La polarización es modelada por una medida borrosa basada en la medida *JDJ* bi-polar. Además, se presenta un algoritmo eficiente para encontrar grupos o comunidades cuyos elementos no estén polarizados. Finalmente, se aplica a un caso real obtenido de Twitter relativo a la posición política de los nodos contra el gobierno de España. Para ello, se analiza como las particiones obtenidas cambian cuando la información adicional es añadida al problema.

Abstract.

In this research, we will discuss one of the most important topics in the field of social network analysis: the problem of community identification. This additional information is modeled by a fuzzy measure that represents the risk of polarization. In particular, we are interested in addressing the consideration of node polarization in the community detection problem. Adding such information to the community detection problem makes it more realistic because communities are more likely to be defined when its elements are willing to maintain a peaceful dialogue. The polarization capacity is modeled by a fuzzy metric based on the bipolar measure called *JDJ*. We also propose an efficient algorithm to find groups whose elements are not polarized. Below we work on a real case. This is a network obtained from Twitter involving several influential users against the Spanish government. We analyze how the obtained partitions change when some additional information about the

polarization of this society is added to the question.

5.3.1. Detección de comunidades a través de una nueva perspectiva.

El campo del Análisis de Redes Sociales - ARS o SNA (*social network analysis*) - abarca un amplio abanico de procesos relacionados con las estructuras sociales modeladas por grafos. Estas estructuras presentan la virtud de reflejar las relaciones presentes en un sistema complejo de entidades. Dentro de los diferentes procesos que incluye el ARS, uno de los más importantes es el denominado problemas de detección de comunidades. El objetivo principal de aplicar algoritmos de detección de comunidades en redes sociales es el de agrupar individuos, representados por nodos, en comunidades, con la intención de conocer la estructura interna de una sociedad determinada. En este sentido, la detección comunitaria y la polarización social están estrechamente relacionadas.

Como se ha comentado en capítulos anteriores, la polarización puede entenderse, a rasgos generales, como la división de una determinada población en dos grupos antagónicos, ambos de tamaño significativo y similar. En este sentido, en tanto que la polarización implica conceptos como el de homofilia o heterofilia, se entiende que los grupos que se forman en torno al proceso de polarización deben de compartir determinadas características entre los individuos que los componen, diferenciándolos, así, del resto de individuos pertenecientes al resto de grupos o comunidades. Así, se encuentra una estrecha relación entre los problemas de detección de comunidades y la polarización. En específico, las medidas de polarización ofrecen información de gran interés para la detección de comunidades desde un punto de vista realista. Así, el grado de polarización entre individuos aportará información sobre el grado de similitud o cercanía entre dos individuos, así como la alienación que presentan los mismos.

Tradicionalmente, la única información considerada para la definición de grupos es el conocimiento representado por un grafo, sin considerar ningún dato adicional. Yendo un paso más allá, varios autores coinciden en la idea de añadir información adicional al grafo, ya sea en un contexto de Teoría de Juegos (Gómez et al., 2003 [43]; Gómez et al., 2008 [42]), considerando conjuntos borrosos (Devarajan et al., 2019) [24] o grafos borrosos (Nair et al., 2007) [80]. Por nuestra parte, consideramos la inclusión de conocimientos sobre la polarización de los elementos de un grafo al problema de detección de comunidades. Esta información aportada por medidas de polarización permi-

te la inclusión de información exógena al grafo que, debidamente agregada, permitirá el desarrollo de un algoritmo que agrupo los nodos en comunidades más realistas teniendo en cuenta, no únicamente la información relativa a las relaciones entre los nodos, sino también de acuerdo a la polarización entre los mismos de acuerdo con su ideología social, política, religiosa, etc.

El objetivo de esta investigación es el de construir un modelo consistente en una red en combinación con una medida difusa de polarización JDJ cuya estructura refleje adecuadamente la realidad. Es el gráfico difuso extendido de polarización, que tiene en cuenta tanto la actitud de las personas como las características de la estructura de la red social. A partir de este modelo definimos un método de detección de comunidades, que al disponer de un grafo y conocer el grado de pertenencia de cada individuo a dos polos, proporciona particiones realistas de la realidad.

5.3.2. Medida de bipolarización difusa JDJ en grafos.

A continuación, se representa la medida JDJ propuesta en el presente capítulo adaptada al caso presente de grafos. Esto presenta dos implicaciones principales. En primer lugar, se definirá la medida de acuerdo con su aplicación a un conjunto de nodos V , adaptándola al caso de grafos. En segundo lugar, y con el fin de evitar conflictos de notación entre las medidas borrosas recién planteadas y las funciones de pertenencia de JDJ , ambas con notación μ , por generalidad, se mantendrá, únicamente para este apartado, la notación μ referente a la medida borrosa y se adaptará la notación referente a los grados de pertenencia de JDJ . Así, se presenta la adaptación de la medida de polarización difusa a continuación:

Definición 5.3.1 (JDJ_{pol} Medida de polarización difusa [50]). Sea V un conjunto finito, y sea η_{X_A} y η_{X_B} las funciones de pertenencia de los elementos de V a las posiciones extremas X_A y X_B . Sea $\varphi : [0, 1]^2 \rightarrow [0, 1]$ un operador de agregación y $\phi : [0, 1]^2 \rightarrow [0, 1]$ una función de *overlapping*. Entonces, JDJ_{pol} se define como:

$$JDJ_{pol}(V, \eta_{X_A}, \eta_{X_B}, \varphi, \phi) = \sum_{\substack{i, j \in V \\ i \leq j}} \varphi(\phi(\eta_{X_A}(i), \eta_{X_B}(j)), \phi(\eta_{X_B}(i), \eta_{X_A}(j))) \quad (5.6)$$

5.3.2.1. Redes con información adicional: El grafo borroso extendido de polarización.

De la medida presentada en la anterior sección, se presenta una nueva medida borrosa de polarización basada en JDJ . Para ello, se deben de asumir dos cuestiones primordiales. En primer lugar, se asume la existencia de un grafo nítido $G = (V, E)$. Segundo, se asume la posibilidad de obtener cierta información o conocimiento sobre la actitud, opinión u otras características de los elementos de V . Así, se define la siguiente medida difusa de polarización basada en las relaciones existentes entre todos los pares de elementos pertenecientes a V .

Definición 5.3.2 (Medida de polarización borrosa μ_{P^-}). Dada una variable unidimensional, sea V un conjunto de n individuos, sobre los que se conoce su grado de pertenencia a ambos polos de la variable X , X_A y X_B , representados por las funciones de pertenencia η_{X_A} y η_{X_B} , respectivamente. Sean las funciones $\phi : [0, 1]^2 \rightarrow [0, 1]$ y $\varphi : [0, 1]^2 \rightarrow [0, 1]$ un operador de agrupación y un operador de solapamiento [44], respectivamente. Sea S un subconjunto de V y sea $JDJ_{pol}(\{i, j\}, \eta_{X_A}, \eta_{X_B}, \varphi, \phi) = \varphi(\phi(\eta_{X_A}(i), \eta_{X_B}(j)), \phi(\eta_{X_A}(j), \eta_{X_B}(i)))$. Así, se define la siguiente medida de polarización μ_{P^-} como:

$$\mu_{P^-}(S) = \frac{JDJ_{pol}(S, \eta_{X_A}, \eta_{X_B}, \varphi, \phi)}{JDJ_{pol}(V, \eta_{X_A}, \eta_{X_B}, \varphi, \phi)} \quad (5.7)$$

Proposición 5.3.3. *La función μ_{P^-} es una medida borrosa.*

Demostración. Para demostrar esta afirmación, se muestra que las propiedades explicadas en 2.4.6 se aplican a μ_{P^-} .

- $\mu_{P^-}(\emptyset) = 0$. Trivial.
- $\mu_{P^-}(V) = 1$. μ_{P^-} es 1-normalizada por definición.
- Sea $A, B \subseteq V$ tal que $A \subseteq B$. Por lo que, $\mu_{P^-}(A) \leq \mu_{P^-}(B)$. Por definición, JDJ_{pol} es una medida monótona.

□

Observación 4. Nótese que la medida μ_{P^-} puede ser reformulada como la suma de todos los pares de elementos:

$$\mu_{P^-}(S) = \sum_{i,j \in S} P_{i,j}^- \quad (5.8)$$

Donde,

$$P_{i,j}^- = \frac{\varphi(\phi(\eta_{X_A}(i), \eta_{X_B}(j)), \phi(\eta_{X_B}(i), \eta_{X_A}(j)))}{JDJ_{pol}(V, \eta_{X_A}, \eta_{X_B}, \varphi, \phi)} \quad (5.9)$$

Debido a las propiedades de μ_{P^-} , P^- es simétrica, no negativa normalizada y su diagonal es nula.

Debido a la naturaleza y diseño de la medida de polarización borrosa ya comentada, se puede conocer el grado de conflicto o distanciamiento entre un determinado par de nodos. Sin embargo, en tanto que el objetivo de este estudio es el de agrupar a diferentes nodos o individuos de acuerdo a sus similitudes, se presenta la siguiente modificación, a través de la negación de la medida.

Observación 5. Se define la negación de la medida de riesgo de polarización JDJ entre un par de individuos para, en su contrario, conocer el grado de cercanía o capacidad de diálogo que dos individuos muestran entre sí. Sea $N : [0, 1] \rightarrow [0, 1]$ un operador de negación, se define:

$$\widetilde{JDJ}(\{i, j\}, \eta_{X_A}, \eta_{X_B}, \varphi, \phi) = N(\varphi(\phi(\eta_{X_A}(i), \eta_{X_B}(j)), \phi(\eta_{X_A}(j), \eta_{X_B}(i)))) \quad (5.10)$$

Así, se define la matriz P^+ como:

$$P_{ij}^+ = \frac{\widetilde{JDJ}(\{i, j\}, \eta_{X_A}, \eta_{X_B}, \varphi, \phi)}{\sum_{r,s \in V} \widetilde{JDJ}(\{r, s\}, \eta_{X_A}, \eta_{X_B}, \varphi, \phi)}, \quad i, j \in V \quad (5.11)$$

Definición 5.3.4 (Medida borrosa de no polarización μ_{P^+}). Sea V un conjunto finito, φ una función de agrupación, ϕ una función de conjunción, N un operador de negación, $\eta_{X_A}, \eta_{X_B} : V \rightarrow [0, 1]$ dos funciones de pertenencia y sea P^+ la matriz caracterizada en la ecuación 5.11. Entonces, a partir de la matriz P^+ , se puede definir una medida que proporcione información sobre la capacidad que dos individuos presentan entre sí para dialogar e interactuar como consecuencia de su similitud en torno a un criterio actitudinal o aptitud, por ejemplo.

$$\mu_{P^+}(S) = \sum_{i,j \in S} P_{i,j}^+ \quad (5.12)$$

Observación 6. μ_{P^+} es una medida borrosa.

Para ilustrar así el funcionamiento de esta medida, véase el siguiente ejemplo trivial:

Ejemplo 5.3.1. Se calcula μ_{P+} de acuerdo con un conjunto de nodos V de 4 elementos. Se consideran las funciones de pertenencia η_{X_A} y η_{X_B} expuestas en la tabla 5.5. Se consideran, además, las funciones $\varphi = \text{máx}$, $\phi = \text{product}$, y $N(x) = 1 - x$.

Tabla 5.5: Grados de pertenencia de cada elemento de V hacia los polos X_A y X_B .

Elemento	η_{X_A}	η_{X_B}
1	1	0
2	0	1
3	1	0
4	0	1

Nótese que $\widetilde{JDJ}(V) = 2$ es la cantidad de diálogos pacíficos entre los 4 elementos. Estas relaciones o diálogos provienen de $\{1, 3\}$ y $\{2, 4\}$. Así, los resultados de ambas medidas - μ_{P-} y μ_{P+} - son mostrados en la siguiente tabla, ilustrando los valores correspondientes a la medida μ_{P-} así como a su negación μ_{P+} .

Tabla 5.6: Valores de las medidas difusas μ_{P-} y μ_{P+} .

S	$\{1, 2\}$	$\{1, 3\}$	$\{1, 4\}$	$\{2, 3\}$	$\{2, 4\}$	$\{3, 4\}$	$\{1, 2, 3\}$	$\{1, 2, 4\}$	$\{1, 3, 4\}$	$\{2, 3, 4\}$	$\{1, 2, 3, 4\}$
$\mu_{P-}(S)$	0.25	0	0.25	0.25	0	0.25	0.5	0.5	0.5	0.5	1
$\mu_{P+}(S)$	0	0.5	0	0	0.5	0	0.5	0.5	0.5	0.5	1

Una vez que se ha definido los dos modelos opuestos para representar la capacidad de un conjunto de elementos para argumentar/dialogar, definimos un nuevo modelo de representación: el grafo difuso extendido de no polarización. Combina la capacidad de un gráfico nítido para representar un conjunto de elementos conectados entre sí, con la representación de las sinergias entre estos elementos provenientes de la medida de no polarización μ_{P+} , independientemente de sus conexiones. Por lo tanto, a partir de un grafo nítido, dos funciones de pertenencia y dos operadores de agregación, se puede definir un grafo difuso extendido de no polarización, una herramienta que arroja luz sobre el modelado de la realidad.

Recordar que, dada la naturaleza de la medida, μ_{P+} es una medida difusa 2-aditiva. Así, grafo difuso extendido de no polarización $\tilde{G} = (V, E, \mu_{P+})$, basado en un grafo nítido y una medida de no polarización, se define como:

Definición 5.3.5 (Grafo difuso extendido de no polarización.). Sea $G = (V, E)$ un grafo nítido, cuyos nodos son representados por V y aristas por

E. Dada una variable unidimensional X que representa actitudes, opiniones, y otra información de interés exógena al grafo, sean X_A y X_B dos polos, y sea μ_{P^+} la medida de no polarización difusa definida en 5.3.4. Así, el triplete $\tilde{G} = (V, E, \mu_{P^+})$ es un grafo difuso extendido de no polarización.

Ejemplo 5.3.2. Se considera el grafo observado en la figura 5.12 con la medida de no polarización difusa μ_{P^+} . Así, se tiene un grafo $G = (V, E)$ y unas funciones de pertenencia $(\eta_{X_A}(1), \dots, \eta_{X_A}(8)) = (1, 0, 0, 1, 1, 1, 0, 0)$ y $(\eta_{X_B}(1), \dots, \eta_{X_B}(8)) = (0, 1, 1, 0, 0, 0, 1, 1)$. Se considera $\varphi = \text{máx}$ y $\phi = \text{product}$, y $N(x) = 1 - x$. Así, el grafo difuso extendido de no polarización $\tilde{G} = (V, E, \mu_{P^+})$ se muestra en la figura 5.12, en el cual se muestra un grafo nítido y la matriz P^+ referente a μ_{P^+} . Esta matriz representa los valores de la medida de no polarización difusa para cada par de individuos o nodos.

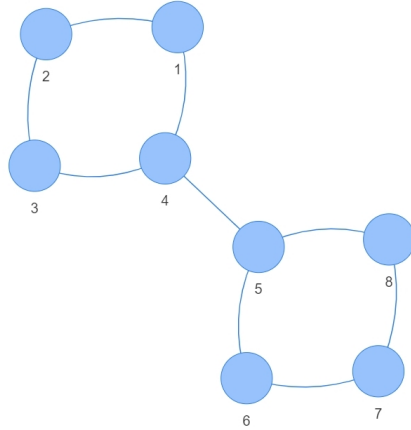


Figura 5.12: Grafo difuso extendido de no polarización. $\tilde{G} = (V, E, \mu_{P^+})$.

$$P^+ = \frac{1}{24} \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

De aquí en adelante, para simplificar la notación, se considera $\mu_P = \mu_{P^+}$ y $P = P^+$.

Finalmente, y llegado a este punto, se trabaja en la idea desarrollada inicialmente por Gutiérrez et al. (2020) [52], en cuyo trabajo los autores pro-

ponen el algoritmo Louvain con información adicional. En su trabajo, Gutiérrez et al. resaltan dos puntos fundamentales en la tarea. En primer lugar, se establecen las relaciones entre los vecinos mientras que, por otro lado, se calcula la variación de la modularidad. Para la primera de las tareas, se utiliza la matriz de adyacencia del grafo A , de tal forma que únicamente aquellos nodos que están conectados en G pueden pertenecer a la misma comunidad. Por otro lado, se sugiere considerar una combinación de los dos componentes del grafo borroso extendido de no polarización \tilde{G} como base para calcular la variación de la modularidad, con el fin de incorporar la información adicional. Seguidamente, teniendo un grafo nítido G , las funciones de pertenencia η_{X_A} y η_{X_B} , los operadores φ (agrupación), y ϕ (solapamiento) y considerando la función de negación N , o lo que es lo mismo, un grafo difuso extendido de no polarización $\tilde{G} = (V, E, \mu_P)$, proponemos una nueva metodología que se resume a continuación.

1. Se obtiene la medida difusa de no polarización μ_P con un set de nodos V con los parámetros η_{X_A} , η_{X_B} , φ , ϕ y N , de acuerdo con la ecuación (5.12).
2. Se resumen μ_P en una matriz, F .
3. Se define la matriz $M = \theta(A, F)$, donde $\theta : \Pi(n)^2 \rightarrow \Pi(n)$ combina dos matrices en una.
4. Se aplica el algoritmo de Louvain distinguiendo los roles de la matriz A , usado para la distinción entre vecinos asociado a la estructura del grafo y M , usada para calcular la variación de la modularidad.

Observación 7. En esta investigación se sugiere usar el operador de agregación de matrices θ . Sin embargo, cualquier otro operador de estas características puede ser usado.

Así, para llevar a cabo los pasos indicados anteriormente, es necesario definir la matriz F . Para ello, en esta propuesta se aborda dicha tarea a través del cálculo de un grafo ponderado asociado con μ_p . Esta matriz es una herramienta muy recomendada para la manipulación y visualización de medidas difusas, que resume el conocimiento sobre la capacidad de los elementos en un conjunto de datos de n^2 . La definición del grafo se basa en el valor de Shapley [105], específicamente en su representación en relación a la medida difusa [47].

Definición 5.3.6 (Grafo ponderado asociado con una medida borrosa G_μ [51, 52]). Sea $\mu : 2^V \rightarrow [0, 1]$ una medida borrosa definida sobre un conjunto

finito V , y sea $\xi : [-1, 1]^2 \rightarrow [0, 1]$ un operador de agregación bivariado. Se considera $Sh_i(\mu)$, el valor de Shapley de un individuo $i \in V$ en coalición con todos los elementos de V de acuerdo con su relación con μ ; análogamente, $Sh_i^j(\mu)$ denota el valor de Shapley de un individuo i en coalición con todos los elementos de $V \setminus \{j\}$, de acuerdo con μ . Así, el grafo ponderado asociado con una medida borrosa μ , definido por G_μ , es aquel cuya matriz de adyacencia es representada por la matriz F , donde:

$$F_{ij} = \xi (Sh_i(\mu) - Sh_i^j(\mu), Sh_j(\mu) - Sh_j^i(\mu)) \quad (5.13)$$

Así, dado que μ_P es la medida difusa de no polarización correspondiente a JDJ_{pol} , se supone que la cercanía de dos individuos con respecto a su actitud hacia un problema puede determinarse mediante el grafo ponderado asociado con la medida borrosa μ_P , cuya la matriz de adyacencia correspondiente es G_{μ_P} , que se calcula como:

$$F_{ij} = \xi (Sh_i(\mu_P) - Sh_i^j(\mu_P), Sh_j(\mu_P) - Sh_j^i(\mu_P)) = \xi (P_{ij}, P_{ji})$$

Donde ξ es un operador de agregación, $Sh_i(\mu_P)$ y $Sh_j^i(\mu_P)$ son los valores de Shapley de i cuando dicho nodo está en coalición con todos los elementos de V y $V \setminus \{j\}$ respectivamente.

Observación 8. Téngase en cuenta que debido a que P es simétrica, si los operadores de agregación ϕ son del tipo *max*, *min*, *average* entre otros, entonces $F_{ij} = P_{ij}, \forall i, j \in V$.

En esta propuesta, se sugiere resumir la medida difusa de no polarización μ_P en la matriz F , con la adyacencia de su grafo ponderado asociado que define el algoritmo de polarización de Louvain para detectar comunidades. Así, se combinan las matrices A y F mediante una combinación lineal $\theta(A, F) = \gamma A + (1 - \gamma) F$, usando el parámetro γ para asignar el peso o la importancia de cada componente de \tilde{G} . De esta forma, es importante resaltar que en aquellos casos en los que $\gamma = 1$ la información adicional no es considerada, correspondiente el algoritmo al de Louvain tradicional. Finalmente, se propone formalmente el siguiente algoritmo denominado *Polarization Louvain*:

Algoritmo 1 *Polarization Louvain*

```

1: Input:  $(A, \eta_{X_A}, \eta_{X_B}, \varphi, \phi)$ ;
2: Output:  $P$ ;
3: Preliminary
4:  $\mu_P \leftarrow (\eta_{X_A}, \eta_{X_B}, \varphi, \phi, N)$ ;
5:  $F_{ij} = \xi \left( Sh_i(\mu_P) - Sh_i^j(\mu_P), Sh_j(\mu_P) - Sh_j^i(\mu_P) \right)$ , for all  $i, j \in V$ ;
6:  $M \leftarrow \theta(A, F)$ ;
7:  $C_i \leftarrow \{i\}$ ,  $\forall i \in V$  (each node is an isolated community);
8:  $P \leftarrow (1, 2, \dots, n)$  (initial partition);
9: end Preliminary
10: Phase 1
11:  $(o^1, \dots, o^i, \dots, o^n) \leftarrow perm(V)$ ;
12:  $stop \leftarrow 0$ ;
13: while ( $stop == 0$ ) do
14:    $stop \leftarrow 1$ 
15:   for ( $i = 1$ ) to ( $n$ ) do
16:      $H(o^i) \leftarrow (e_1, \dots, e_h)$  (find all the neighbours of  $o^i$  in  $A$ );
17:     for ( $j = 1$ ) to ( $h$ ) do
18:       Compute  $\Delta Q_{o^i}(e_j)$  in  $M$ ;
19:     end for
20:      $j^* \leftarrow \left\{ e_\ell \mid \Delta Q_{o^i}(j^*) = \max_{\ell \in \{1, \dots, h\}} \{ \Delta Q_{o^i}(e_\ell) \} \right\}$ ;
21:     if ( $\Delta Q_{o^i}(j^*) > 0$ ) then
22:        $C_{P(o^i)} \leftarrow C_{P(o^i)} \setminus \{o^i\}$ ;
23:        $C_{P(j^*)} \leftarrow C_{P(j^*)} \cup \{o^i\}$ ;
24:        $P(o^i) \leftarrow P(j^*)$ ;
25:        $stop \leftarrow 0$ ;
26:     end if
27:   end for
28: end while
29: end Phase 1
30: Phase 2
31: Aggregate  $A^*$  from  $A$  (nodes of  $A^*$  are the communities found in Phase 1);
32: Aggregate  $M^*$  from  $M$  (nodes of  $M^*$  are the communities found in Phase 1);
33: if ( $A^* \neq A$ ) then
34:    $A \leftarrow A^*$ ;
35:    $M \leftarrow M^*$ ;
36:   Compute Phase 1 and Phase 2;
37: end if
38: end Phase 2
39: return( $P$ );

```

5.3.3. Aplicación a un caso real.

5.3.3.1. Fuentes de datos y metodología.

Los nodos y aristas que se utilizan en este trabajo provienen de la red social Twitter, particularmente de todos aquellos mensajes que fueron publicados desde el 16 de marzo al 29 de junio de 2020. Todos los mensajes fueron descargados de acuerdo a un conjunto de palabras clave relacionadas con la COVID-19 y la situación política en España, donde cada nodo representa una cuenta influyente y verificada.

Para este estudio, se ha usado la red de retweets - RT -. RT es una publicación que se origina en el acto de cualquier usuario escoge un mensaje de otro usuario para difundir ese contenido con sus seguidores. En la literatura, las redes RT a menudo se usan como redes dirigidas [93], por lo que si el tweet original es retweeteado por el usuario i y luego por j , entonces hay conexiones direccionales, que representan las acciones de cada usuario. En nuestro caso, sin embargo, lo entendemos como una red no direccional: conocer la direccionalidad de la conexión (quién tuitea y quién retuitea) no es una cuestión relevante para el estudio, sino del contenido. En este caso, una vez que un usuario determinado j ha retuiteado el tuit de i , lo que nos importa es la intención de j de transmitir y difundir ese contenido. En un sentido más amplio, podemos asumir que j está de acuerdo con el contenido y corre la voz para que el tweet sea visible e influyente para las personas de su grupo. El objetivo es comprender las actitudes políticas de los usuarios hacia el gobierno español y medir cómo se reflejan en sus tuits. Por lo tanto, no se requiere direccionalidad.

Todos los datos fueron descargados a lo largo de 5 tandas con el fin de acceder a todo el discurso digital proveniente de Twitter en torno a las palabras clave que actuaron como criterio de descarga. Para ello, se usó el paquete estadístico de R-Studio *rtweet* [63]. La bolsa de palabras utilizada fue la siguiente:

psoe OR pp OR vox OR ciudadanos OR gobierno OR podemos OR españa OR sanchezcastejon OR vox_es OR pabloiglesias OR pablocasado_ OR santi_abascal OR inesarrimadas OR CiudadanosCs OR populares OR estadodealarma

Así, se descargaron un total de 4895747 mensajes, de los cuales fueron codificados manualmente 1500 por cada una de las tandas, a razón de si pertenecen o no a la temática de estudio - TOPIC - y, por otro lado, la posición política que muestran, siendo *a favor*, *en contra* y *neutral* hacia el gobierno de España. Se adaptaron los datos a una estructura que posibilite la

aplicación de los algoritmos de aprendizaje automático a través de un proceso de *tokenización*, eliminación de *stopwords*, y la conversión de los datos a matrices *dfm*, específicamente, *tf-idf*. Los resultados de la clasificación para ambas tareas se observan en la tabla 5.7, los cuales fueron aportados por las máquinas de soporte vectorial, las cuales mostraron el mejor rendimiento con esta base de datos de entre todos los clasificadores probados.

Tabla 5.7: Resultados de las máquinas de soporte vectorial (SVM) para las variables “TOPIC” y “POSITION”.

Tanda	Variable	Precisión	Recall	Kappa	F-Score	AUC
1	TOPIC	0.8017	0.9322	0.3670	0.8620	0.6583
2	TOPIC	0.8167	0.5476	0.5077	0.6556	0.7344
3	TOPIC	0.8267	0.7027	0.6187	0.7596	0.8010
4	TOPIC	0.7867	0.7090	0.564	0.7457	0.7791
5	TOPIC	0.7659	0.8758	0.5216	0.8171	0.7567
1	POSITION	0.8492	0.9854	0.4816	0.9122	0.6950
2	POSITION	0.8960	0.9619	0.7761	0.9277	0.8780
3	POSITION	0.8392	0.8488	0.6675	0.8439	0.8366
4	POSITION	0.9133	0.9048	0.8225	0.9090	0.9121
5	POSITION	0.8318	0.8600	0.6638	0.8456	0.8335

Es importante mencionar que, en este caso, la categoría *neutral* para la variable “POSITION” fue eliminada con el fin de entrenar al clasificador de acuerdo con dos extremos - *a favor* y *en contra* -. Sin embargo, estas dos categorías, que formarán los polos actitudinales de los mensajes, son extraídas, no en valores nítidos, sino en las probabilidades que ofrece el clasificador a pertenecer a cada una de estas categorías, por lo que serán utilizadas como las funciones de pertenencia η_{X_A} y η_{X_B} . En este sentido, probabilidades en torno al 0.5 son entendidas como posiciones *neutrales*, no perdiendo esta información en caso de necesidad.

Finalmente, una vez que se aplicó el filtro a la base de datos de acuerdo con la variable TOPIC, se tuvo una base de datos final de 1208631 tweets que fueron publicados por un total de 469616 usuarios. Por último, para detectar las cuentas más influyentes, se filtraron los mensajes publicados de acuerdo con los siguientes criterios:

- Tweets con alta repercusión en Twitter, se consideraron cuentas cuyos tweets con un número de RT se ubican por encima del percentil 50 - $n \geq 317$ -.

- Cuentas verificadas. Se seleccionaron aquellas cuentas que presentaban verificación oficial por parte de la plataforma.
- Finalmente, se seleccionaron las cuentas con un mayor número de seguidores, seleccionando aquellas cuyo número se sitúa por encima del percentil 50, siendo $n \geq 21779$.

Así, únicamente restaron 406 usuarios, siendo los mismos mayormente políticos, cuentas de instituciones y reporteros. A continuación, y para asegurar una red cerrada, se filtraron los usuarios en función a aquellos que habían producido un mensaje que ha sido retuiteado o, bien, retuitearon algunos de estos mensajes. De esta forma, la base de datos se redujo a un total de 295 usuarios. Sin embargo, en tanto que alguno de estos usuarios han retuiteado mensajes de otras cuentas que no están originalmente en la base de datos, se cuenta, finalmente, con una red compuesta por 454 usuarios con 657 interacciones entre ellos. Con esta información, se construye la red con la que se trabaja en esta investigación $G = (V, E)$, siendo los 454 usuarios el conjunto de nodos V y las 657 interacciones entre ellos la lista de aristas E . Además, se tendrá en cuenta el número de veces que los usuarios interactúan entre ellos, por lo que se trabaja con un grafo ponderado, cuyo parámetro de ponderación w_{ij} viene determinado por la cantidad de interacciones dadas entre un par de individuos i, j .

Por último, cabe señalar que tanto el grafo, como la polarización, se construyen a través de los usuarios, mientras que los algoritmos de aprendizaje automático han sido entrenados de acuerdo con mensajes o publicaciones. En tanto que se necesita una probabilidad de posicionarse *a favor* y *en contra* del gobierno para cada usuario con el fin de conocer su grado de pertenencia a ambos polos, se calcula, para cada usuario, las probabilidades medias de todos sus mensajes, obteniendo, así, para cada usuario i , únicamente dos probabilidades. Así, se podrá calcular la medida de polarización JDJ , que permitirá la computación de la medida de no polarización μ_P , que a su vez, permitirá la aplicación del grafo borroso extendido no polarizado $\tilde{G} = (V, E, \mu_P)$ con el que se aplicará el algoritmo de detección de comunidades.

5.3.3.2. Resultados.

A pesar de contar con una red de 454 nodos y 657 aristas, para el proceso de agrupamiento, se escogerá en su componente débil, que contiene 261 nodos y 484 aristas ponderadas. La red obtenida, $G = (V, E)$, con matriz de adyacencia, A , se muestra en la Figura 5.13 - tomando en consideración el peso de cada arista, se representa el grado de los nodos por su tamaño en la imagen, de modo que los nodos más grandes representarán a los usuarios con la mayor

cantidad de interacciones -. Entonces, considerando los grados de pertenencia de cada nodo a los polos X_A - estando “en contra” del gobierno español - y X_B - estando “a favor” del gobierno español -, representados por η_{X_A} y η_{X_B} , respectivamente, se puede calcular la medida de Polarización JDJ_{pol} a partir de la cual se define la matriz P , según la ecuación (5.11) de la cual se deriva la medida difusa de no polarización μ_P que es uno de los componentes de $\tilde{G} = (V, E, \mu_P)$. Tenga en cuenta que la información proporcionada por el grafo difuso extendido de no polarización va más allá que la proporcionada por un grafo nítido. También incluye el conocimiento sobre la posición de los nodos de G con respecto a un eje actitudinal, información que no puede ser modelada por herramientas clásicas. A continuación, se muestra la estructura de la red sin discriminar por comunidades.

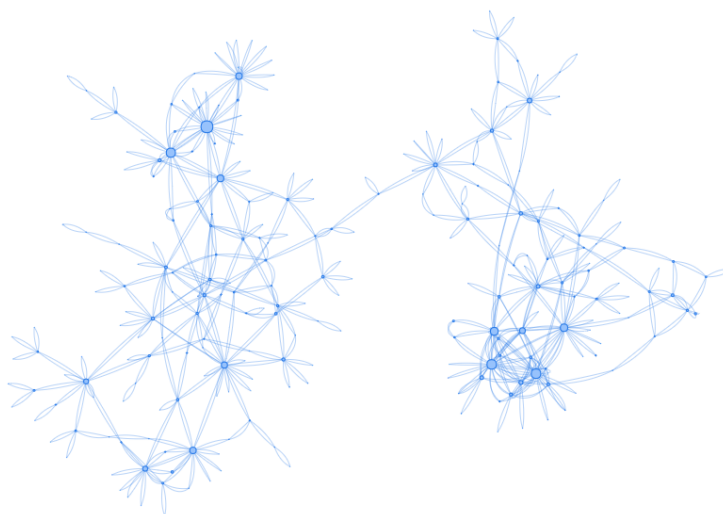


Figura 5.13: Grafo $G = (V, E)$.

Por el contrario, la figura 5.14 muestra la red con las comunidades detectadas por el algoritmo de Louvain tradicional.

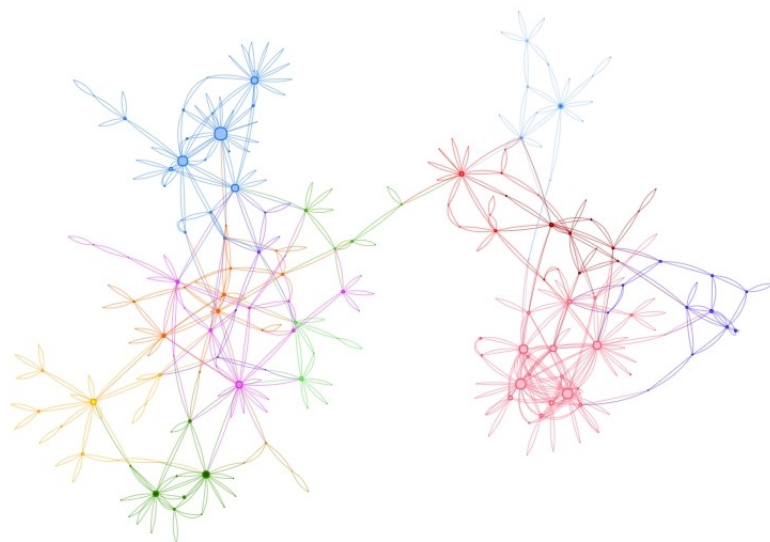


Figura 5.14: Partición obtenida por el algoritmo de Louvain del grafo $G = (V, E)$.

La medida μ_P depende de la selección de un operador de negación, N , y dos tipos diferentes de operadores de agregación: una función de agrupación φ y un operador *overlapping* ϕ . Como operador de negación, usamos $N(x) = (1 - x)$. En cuanto a los operadores de agregación, utilizamos algunos de los operadores más importantes en este campo, teniendo dos escenarios diferentes para la agregación de los grados de pertenencia: (a) $\phi = \min$ y $\varphi = \max$, y (b) $\phi = \text{producto}$ y $\varphi = \max$.

Debido a la caracterización de P , y siendo μ_P una medida difusa definida en la ecuación 5.12, P puede verse también como la matriz de adyacencia de G_{μ_P} , F , por lo que podemos considerar indistintamente ambas herramientas.

A continuación, se aplica el algoritmo *Polarization Louvain* para encontrar comunidades en el grafo difuso extendido de no polarización $\tilde{G} = (V, E, \mu_P)$. Nótese que las comunidades obtenidas serán cohesivas con todo el conocimiento modelado por ella, la estructura del grafo así como la información adicional modelada por μ_P . La noción de *qué es una comunidad* estará estrechamente relacionada con el operador de agregación θ elegido, así como con el operador de agrupación φ y el operador de solapamiento ϕ . Poder considerar la información adicional a la hora de encontrar comunidades permite obtener comunidades realistas mucho más cohesionadas con la situación abordada, que las que dan otros métodos que no pueden analizar más información aparte de la estructura del grafo.

Para combinar las dos componentes de \tilde{G} , trabajamos con combinaciones lineales de las matrices A y P asignándoles una importancia mediante el parámetro de balanceo $\gamma \in [0, 1]$, es decir, consideramos la matriz $M = \theta(A, P) = \gamma A + (1 - \gamma)P$.

La influencia de cada componente de \tilde{G} varía según el valor de γ . Para valores de γ cercanos a 1, el componente estructural cobra importancia, por lo que los grupos contienen nodos estrechamente conectados en A . Por el contrario, cuando γ se acerca a 0, la información adicional modelada por μ_P se vuelve determinante en la definición de las comunidades, por lo que, si es posible respecto a la estructura de A , los grupos contienen nodos con bajo nivel de polarización, es decir, nodos cuyo grado de pertenencia a cada polo es similar. En este caso, aquellos usuarios sobre los que podemos asumir un punto de vista político similar, estarán en el mismo grupo.

Aplicamos el algoritmo de *Polarization Louvain* para los dos escenarios de funciones de agrupación/*overlapping* antes mencionadas, y considerando la matriz $M = \gamma A + (1 - \gamma)P$, para varios valores del parámetro de importancia, $\gamma = 0.5, 0.4, 0.3, 0.2, 0.1, 0$. También calculamos el algoritmo de Louvain con la matriz A , en cuyo resultado, mostrado en la figura 5.14, se basa nuestro análisis de comparación. Tenga en cuenta que el rendimiento del algoritmo de Louvain coincide con el algoritmo de *Polarization Louvain* cuando $M = A$, ($\gamma = 1$).

Aquí, se muestra cómo la organización de los grupos sigue cambiando dependiendo de la importancia de cada componente de \tilde{G} en el proceso de agrupación. En particular, para los casos extremos, Louvain - en el que no hay información adicional - y $\gamma = 0$ - en el que la información adicional cobra toda la importancia -, considerando los dos escenarios mencionados anteriormente sobre los operadores de agregación utilizados - producto o mínimo -. Los resultados se muestran en las Figuras 5.15 y 5.16.

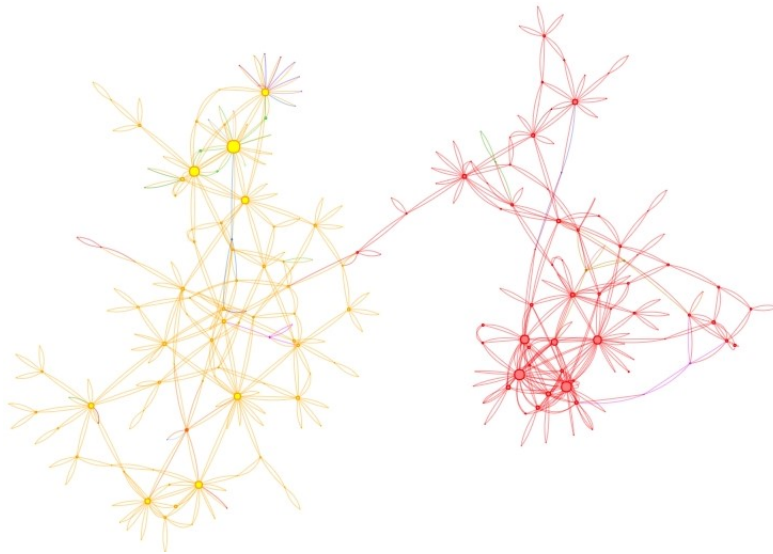


Figura 5.15: Particiones obtenidas con el algoritmos de *Polarization Louvain* con el grafo borroso extendido de no polarización $\tilde{G} = (V, E, \mu_P)$. $\gamma = 0$; $\varphi = \max$; $\phi = \min$.

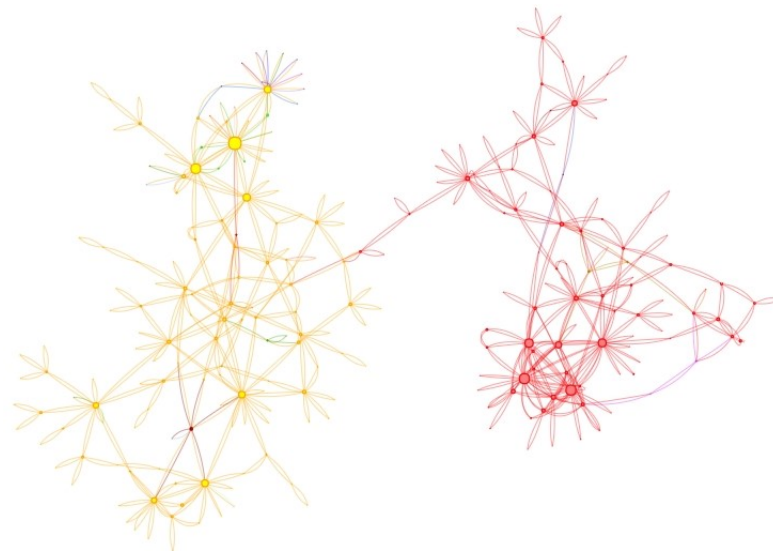


Figura 5.16: Particiones obtenidas con el algoritmos de *Polarization Louvain* con el grafo borroso extendido de no polarización $\tilde{G} = (V, E, \mu_P)$. $\gamma = 0$; $\varphi = \max$; $\phi = \text{prod}$.

Nótese cómo, cuando sólo se considera el punto de vista político de los usuarios, el gráfico se divide en dos comunidades principales, de modo que

se puede diferenciar fácilmente entre los detractores y los partidarios del gobierno español.

Para medir la bondad de las particiones obtenidas, se hace referencia a la medida JDJ_{pol} . Se asume que un grupo cohesionado debe estar compuesto por usuarios conectados con puntos de vista similares. En este sentido, se puede decir que un grupo está tan cohesionado como bajo es su valor de JDJ_{pol} correspondiente.

Nótese que las particiones obtenidas al considerar varios valores de γ varían en el número de comunidades. Luego, para compararlos, se considera el promedio ponderado del valor $JDJ_{pol}(C_i)$ de todas sus comunidades. Por lo tanto, se calcula el valor de polarización de la partición $P = \{C_1, \dots, C_s\}$ como:

$$pol(P) = \left[\frac{\sum_{i=1}^s JDJ_{pol}(C_i) * |C_i|}{\sum_{i=1}^s |C_i|} \right]_{|C_i| > 1} \quad (5.14)$$

En las tablas 5.8 y 5.9, se muestra el valor de JDJ_{pol} de cada comunidad en las particiones obtenidas - solo comunidades no aisladas -, así como la correspondiente $pol(P)$. Para cada partición, mostramos el vector $(JDJ_{pol}(C_1), \dots, JDJ_{pol}(C_s))$, de modo que la i ésima componente se corresponda con $JDJ_{pol}(C_i)$.

Tabla 5.8: Comparación entre las particiones obtenidas. $\varphi = max$ and $\phi = min$.

$\varphi = max$ $\phi = min$	# Comunidades $ C_i > 1$	$(JDJ(C_1), \dots, JDJ(C_s))$	$pol(P)$
<i>Louvain</i>	14	(0.256, 0.514, 0.253, 0.301, 0.458, 0.377, 0.302, 0.4403, 0.459, 0.349, 0.190, 0.475, 0.108, 0.415)	0.359
$\gamma = 0.5$	11	(0.239, 0.259, 0.513, 0.297, 0.377, 0.440, 0.514, 0.257, 0.459, 0.415, 0.455)	0.341
$\gamma = 0.4$	8	(0.254, 0.332, 0.259, 0.513, 0.450, 0.514, 0.459, 0.455)	0.348
$\gamma = 0.3$	7	(0.304, 0.300, 0.253, 0.513, 0.512, 0.526, 0.246)	0.343
$\gamma = 0.2$	8	(0.334, 0.267, 0.444, 0.512, 0.462, 0.440, 0.528, 0.246)	0.330
$\gamma = 0.1$	7	(0.323, 0.273, 0.418, 0.482, 0.440, 0.462, 0.246)	0.319
$\gamma = 0$	8	(0.302, 0.263, 0.439, 0.463, 0.277, 0.440, 0.462, 0.246)	0.292

Tabla 5.9: Comparación entre las particiones obtenidas. $\varphi = max$ and $\phi = prod$.

$\varphi = max$ $\phi = prod$	# Comunidades $ C_i > 1$	$(JDJ(C_1), \dots, JDJ(C_s))$	$pol(P)$
<i>Louvain</i>	14	(0.218, 0.454, 0.228, 0.261, 0.378, 0.296, 0.261p, 0.359, 0.389, 0.306, 0.168, 0.392, 0.102, 0.258)	0.306
$\gamma = 0.5$	11	(0.220, 0.453, 0.190, 0.260, 0.261, 0.296, 0.359, 0.326, 0.382, 0.389, 0.258)	0.299
$\gamma = 0.4$	9	(0.214, 0.281, 0.220, 0.453, 0.363, 0.389, 0.369, 0.258, 0.343)	0.292
$\gamma = 0.3$	7	(0.257, 0.251, 0.220, 0.453, 0.369, 0.417, 0.343)	0.292
$\gamma = 0.2$	7	(0.259, 0.228, 0.453, 0.369, 0.417, 0.249, 0.186)	0.289
$\gamma = 0.1$	7	(0.274, 0.228, 0.393, 0.417, 0.369, 0.249, 0.186)	0.277
$\gamma = 0$	8	(0.256, 0.224, 0.316, 0.376, 0.199, 0.243, 0.249, 0.186)	0.244

Como se puede ver en las tablas anteriores, así como en la figura 5.17, el valor de $pol(P)$ relacionado con aquellas particiones obtenidas con el algoritmo *Polarization Louvain* es menor que el relacionado con la partición proporcionada por el algoritmo de Louvain. Así, se puede afirmar que este método proporciona estructuras comunitarias más cohesionadas de acuerdo con la realidad modelada.

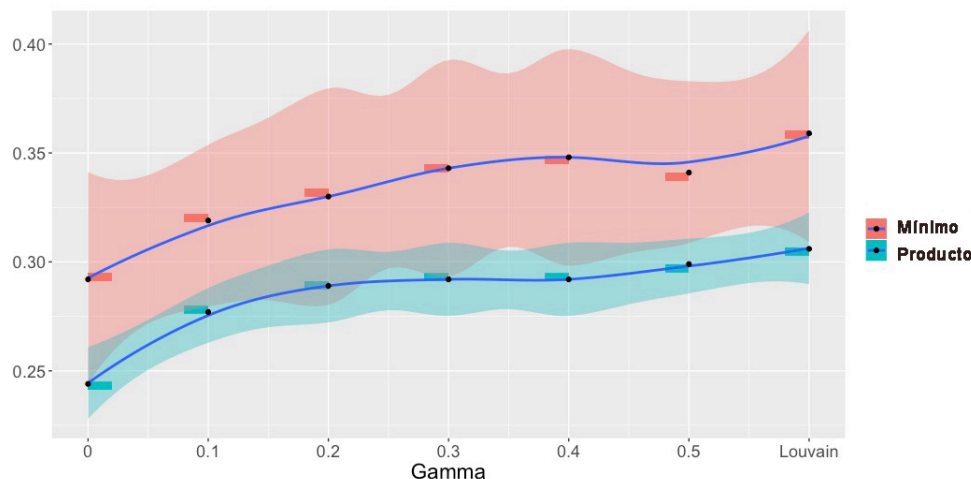


Figura 5.17: Valores de Polarización para la partición $P = \{C_1, \dots, C_s\}$ por operadores de *overlapping*.

Para ilustrar esto, en las siguientes figuras se muestra un ejemplo de cómo dos pares de nodos que deberían pertenecer a las mismas comunidades, respectivamente, se dividen en cuatro comunidades diferentes con el algoritmo de Louvain. Por un lado, tenemos los nodos 38 y 115, ambos partidos políticos de izquierda que se unieron en marzo de 2019. Por otro lado, se tienen los nodos 76, un partido político de derechas, y 203 miembro de este grupo político. Después de aplicar el algoritmo de *Polarization Louvain*, estos pares de nodos se agrupan en las mismas comunidades - consulte las figuras 5.19 y 5.18 -. Nótese que las imágenes mencionadas son un zoom sobre toda la red, por lo que no se muestran todas las aristas que inciden en estos nodos. Aunque pueda parecer que algunos nodos agrupados en las mismas comunidades no están conectados por aristas - por ejemplo, los nodos 76 y 203 en la imagen de la figura 5.19 -, todos ellos están correctamente conectados en la red.

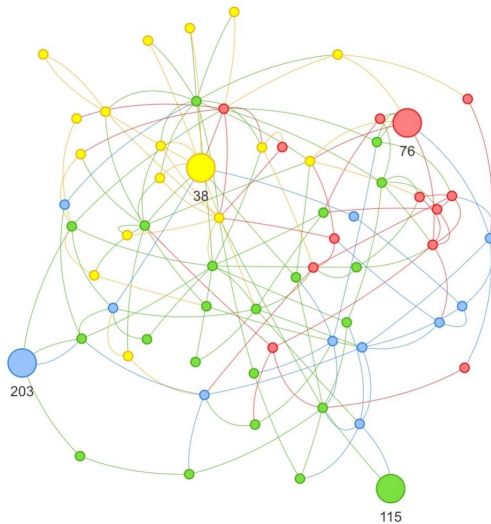


Figura 5.18: Nodos 38, 76, 115 y 203 agrupados por Louvain.

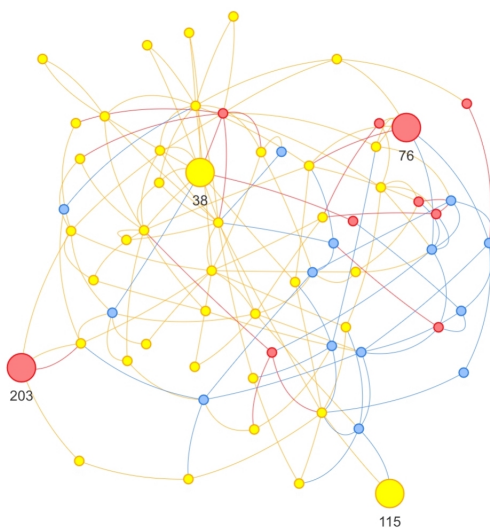


Figura 5.19: Nodos 38, 76, 115 y 203 agrupados por *Polarization Louvain algorithm*.

Capítulo 6

Medidas de Polarización en grafos y redes

Resumen: En este capítulo se amplía la definición de polarización *JDJ* propuesta anteriormente al caso de grafos, definiendo una nueva medida en grafos borrosos. Este capítulo se centrará en el caso de dos comunidades en un contexto borroso. Se presenta un problema bien conocido en las redes sociales de la vida real para comparar nuestros resultados con el caso de grafos nítidos, así como con las medidas de polarización aplicadas a los mismos. Los resultados muestran mejoras en la detección de polarización enmascarada en un contexto nítido.

Simón de Blas, C., Guevara, J.A., Morillo, J., Gómez González, D. (2022). Polarization Measures in Bi-partition Networks Based on Fuzzy Graphs. Communications in Computer and Information Science, vol 1601. Springer, Cham.

Abstract.

In this paper, we extend the definition of polarization in networks by defining new metrics in fuzzy graphs. We will focus on the case of two communities in a fuzzy context. We ask a known problem on a real social network to compare our results with crisp cases. The results show improvements in detecting polarization that are masked in crisp environments.

El análisis de redes sociales se ha convertido en una de las disciplinas más candentes del análisis de datos debido a su aplicación natural en áreas tan diversas como la biología, la medicina, el transporte, comunicación, sociología, marketing, matemáticas o informática (ver, por ejemplo Fortunato, 2010

[29]; Gómez et al., 2003 [43]; Jackson, 2010 [61] o Simón et al., 2018 [22]; entre muchos otros).

Por otro lado, la modelización de la incertidumbre es uno de los principales temas tratados por los científicos en una amplia gama de áreas disciplinarias. Dependiendo de la fuente de la falta de conocimiento que genera la incertidumbre en el área de estudio, se han desarrollado diferentes técnicas para completar la ausencia de información. Desde que se desarrolló el primer lenguaje humano, basado en un pequeño conjunto de fonemas, comienzan a surgir vaguedades en el desarrollo del lenguaje, lo que dificulta deducir la intención completa e interpretar correctamente un mensaje dado. Inspirándose en los aspectos imprecisos del conocimiento humano, Zadeh (1965) [122] introduce conjuntos borrosos como un método para tratar problemas cuando la fuente de imprecisión es la ausencia de criterios bien definidos de pertenencia a una clase.

La inclusión de la lógica difusa en los problemas de las redes sociales es claramente una mejora significativa para hacer frente a problemas reales, ya que es más natural considerar las relaciones borrosas que algo nítido. Un claro ejemplo de este hecho sería el problema de detección de comunidades en redes. Dado que el agrupamiento fue uno de los principales problemas en los que el enfoque difuso parece más natural, las soluciones de detección de comunidades son más realistas cuando se presentan de forma difusa, donde los nodos no siempre se identifican con un grupo o comunidad en su totalidad, sino en grados de pertenencia (Biswas y Biswas, 2018 ([10]; Gómez et al, 2015 [46]; Gómez et al, 2016 [45]).

Este capítulo se centra en una propiedad estructural muy importante de las redes complejas: la capacidad de medir el grado de polarización de la red. La principal contribución de esta investigación es la extensión de la medida de polarización JDJ propuesta en el capítulo anterior 5.6 (Guevara et al., 2020) [50] al caso de los grafos, donde las relaciones entre los nodos pueden ser nítidas o borrosas, y los nodos que pertenecen a varias comunidades pueden no estar totalmente involucrados con todas ellas. Para ello, se considera el caso en que cada vértice pertenece a cada grupo en distinta medida o gradualidad.

Con esta idea, se pretende abordar alguna de las deficiencias encontradas, y mencionadas, anteriormente en el capítulo 3 sobre las medidas de polarización en redes.

6.0.1. Construyendo las funciones de pertenencia a los polos.

En esta sección se propone una medida que permita adaptar la medida de polarización JDJ propuesta en la sección 5.1 al caso de redes, donde la información de partida sea un grafo con una partición de dos clases del conjunto de sus nodos.

Para llevar a cabo esta tarea, y teniendo en cuenta la existencia de dos comunidades nítidas C_A y C_B cuya unión forma el conjunto de nodos, se hará uso de un marco borroso donde los elementos pertenecen a las clases con diferentes intensidades. De la misma manera a como un nodo que se encuentre en el corazón de una comunidad se podría pensar que está ciertamente identificado con ella - $\mu_A(i) = 1$ - se pueden tener otros nodos que se encuentren en la frontera entre ambas comunidades y su grado de identificación no sea el mismo.

Cómo explicar la estructura de comunidades en un grafo es uno de los problemas más importantes en el análisis de redes sociales (ver Fortunato, 2010 [29] Gómez et al, 2015 [46] o Gómez et al, 2016 [45]). En términos generales, una comunidad en una red es un subgrafo conectado cuyos nodos están densamente conectados dentro de sí mismo pero escasamente conectados con el resto de la red. La mayoría de los enfoques asumen que las comunidades forman una partición del conjunto de nodos. En este sentido, dado un grafo $G = (V, E)$, se dice que $P = \{C_1, \dots, C_k\}$ es una solución del problema clásico de detección de comunidades si P es una partición del conjunto de V y los elementos de P están conectados en el grafo. Sin embargo, las comunidades reales en redes complejas, a menudo presentan solapamiento, tal que cada vértice puede ocurrir en más de una comunidad. Los problemas de detección de comunidades con comunidades solapadas - *overlapping communities* - también han sido estudiados en la literatura con diferentes propuestas. Por un lado, la detección de nodos clave que pertenecen a más de una comunidad es una tarea fundamental, mientras que, por otro lado, la detección e identificación de nodos que pertenecen a más de una comunidad - generalmente denominados nodos superpuestos - es otra tarea relacionada con la anterior.

Como señala Gregory (2011) [48], dos tipos diferentes de solapamiento podrían ser posibles en un grafo dado: el caso nítido - donde cada nodo pertenece completamente a cada comunidad de la que es miembro - y el caso borroso - donde cada nodo pertenece a cada comunidad en diferente medida -. En el caso de la detección de comunidades difusas (Gómez et al, 2016) [45], existen dos desafíos principales: el desarrollo de algoritmos que produzcan una agrupación difusa de los nodos en la red y la cuantificación

de la calidad del rendimiento de los algoritmos. Para concluir esta sección, vamos a referirnos al enfoque desarrollado en Gregory (2011) [48], en el que se obtiene una partición difusa a partir de una solución de detección de comunidades nítida mediante el algoritmo *Make Fuzzy*. Formalmente, dado un grafo $G = (V, E)$ con una partición nítida $P = \{C_1, \dots, C_k\}$, se construye la siguiente partición difusa $\widetilde{P}_{Gre} = \{\mu_1, \dots, \mu_k\}$, donde $\mu_i : V \rightarrow [0, 1]$ se define como:

$$\mu_{C_i}(v) = \frac{|N(v) \cap C_i|}{|N(v)|}$$

donde $N(v) = \{u \in V \text{ con } (u, v) \in E\}$.

Ejemplo 6.0.1. Sea $G = (V, E)$ una red con 8 nodos $V = \{1, 2, 3, 4, 5, 6, 7, 8\}$ y 10 aristas $E = \{(1, 2), (1, 3), (2, 4), (3, 4), (4, 5), (4, 6), (5, 6), (5, 7), (6, 8), (7, 8)\}$, y sea $P = \{C_1 = \{1, 2, 3, 4\}, C_2 = \{5, 6, 7, 8\}\}$ la partición de la red en dos comunidades. Así, la partición borrosa $\widetilde{P}_{gre} = \{\mu_{C_1}, \mu_{C_2}\}$ viene dada por:

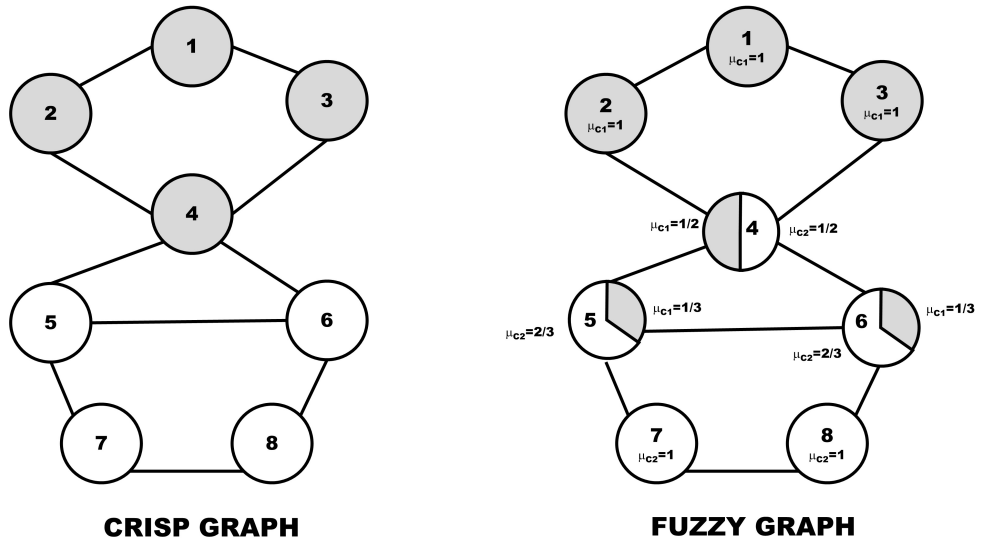


Figura 6.1: En gris, nodos pertenecientes a la comunidad C_1 , en blanco, los referentes a C_2 .

6.0.2. Medida de Polarización difusa JDJ aplicada a grafos

Una de las afecciones más comunes del término polarización se asocia cuando existen dos grupos opuestos que rompen el diálogo y la comunicación. En cuanto a las redes, se puede pensar en el hecho de que existen pocas relaciones entre ellas. En una red con dos comunidades, los nodos de una comunidad que se vinculan con un nodo de la otra comunidad forman lo que se llama un límite. Entre dos comunidades altamente polarizadas el tamaño de la frontera es muy pequeño o incluso nulo (Guerra et al., 2013) [49]. Pero también, es importante señalar que el tamaño de las dos comunidades también debe verse afectado por la polarización (Esteban y Ray, 1994) [28], ya que la mayor polarización se alcanza cuando existen dos grupos sin relaciones y el tamaño de los grupos es similar. No se puede asumir los mismos niveles de polarización en el caso en el que se tienen dos grupos de tamaño $100k$ y $105k$, o por el contrario, con tamaño $195k$ y $10k$.

Siguiendo otros enfoques desarrollados en la literatura que abordan la medición de la polarización en redes (Guerra et al., 2013) [49], la medida de polarización JDJ usada en este estudio se centrará en el caso en el que se presenta una red con dos comunidades, ya sean para el caso nítido o borroso. Así, JDJ trata con el problema de medir la polarización de redes difusas $(V, E, \tilde{P}) = (V, E, \mu_1, \mu_2)$ o en un grafo nítido $(V, E, P) = (V, E, \{C_1, C_2\})$.

Para medir la polarización de una red de dos comunidades $(V, E, P) = (V, E, \{C_1, C_2\})$, los enfoques más comunes (ver Reynal et al, 2001 [92] o Wang y Tsui, 2000 [118] por ejemplo), son los de asociar el concepto de polarización al caso en que existen dos comunidades de tamaño similar. En este sentido, la polarización puede verse como una medida de dispersión sobre variables nominales, que para el caso de dos grupos esta medida equivale a la varianza de una variable aleatoria de *Bernoulli*. Este hecho refleja la situación en la que la red se divide en dos grupos con un tamaño similar, siendo uno de los factores importantes en la polarización. Sin embargo, el problema de este enfoque es considerar que los nodos de una comunidad pertenecen claramente a dicha comunidad sin diferenciar que hay nodos que, por su proximidad a la frontera o sus relaciones con los elementos de otra comunidad, podrían ser claramente considerados en ambas comunidades. Por esta razón, se considera que una partición difusa representa más adecuadamente la pertenencia a un grupo y es en este tipo de partición donde se debe medir la polarización.

Para tener en cuenta el tamaño de los grupos en cada comunidad pero también su grado de pertenencia, se propone la siguiente definición de polarización para redes con dos comunidades difusas basada en la medida de

polarización JDJ con operadores de agrupación y agregación *overlapping* o solapamiento.

Definición 6.0.1 (Medida de polarización borrosa JDJ en grafos). Sea $(V, E, \tilde{P}) = (V, E, \mu_1, \mu_2)$ un grafo con dos particiones borrosas. Sea G_G, G_O un operador de agrupación y *overlapping* ya definidos en capítulos anteriores. Así, la medida de polarización difusa en grafos $JDJ_{pol}(V, E, \tilde{P}, G_O, G_G)$ se define como:

$$JDJ_{pol}(V, E, \tilde{P}, G_O, G_G) = \frac{1}{(|N|/2)^2} \sum_{u,v \in V, u \neq v} G_G(G_O(\mu_1(u), \mu_2(v)), G_O(\mu_2(u), \mu_1(v))) \quad (6.1)$$

Observación 9. Obsérvese que el coeficiente $\frac{1}{(|N|/2)^2}$ es solo para garantizar que la medida anterior es de hecho un índice que pertenece al intervalo unitario y es igual a 1 si y solo $N = C_1 \cup C_2$, $|C_1| = |C_2|$, y $\mu_i(u) = 1$ si $u \in C_i$ y 0 en caso contrario. Esta medida se puede ver como el promedio de polarización entre cada par de elementos en V . Para hacer más interpretable esta medida, se puede obtener el promedio de esta suma cambiando el coeficiente $\frac{1}{(|N|/2)^2}$ por $\frac{2}{(|N|(|N|-1))}$. El promedio representa la polarización promedio entre cada par de nodos en la red. Se puede ver esta diferencia en el siguiente ejemplo.

Ejemplo 6.0.2. Recuérdesse nuevamente el ejemplo 6.0.1 en el que se da una partición difusa. Sean $G_O(x, y) = xy$ y $G_G(x, y) = Max(x, y)$ dos operadores de agregación. Si se quiere obtener como se polarizan los nodos 1 y 4 en la partición difusa, se tiene que calcular $Pol(1, 4) = Max\{\mu_1(1)\mu_2(4), \mu_2(1)\mu_1(4)\} = Max\{0.5, 0\} = 0.5$. En la tabla 6.1 se aprecian algunas de las 28 comparaciones por pares que deben agregarse con la suma para obtener la medida de polarización.

Tabla 6.1: Algunas comparaciones por pares y su medición de polarización de la red de comunidades borrosas (V, E, \tilde{P}_{gre})

																Media	JDJ
Nodo x	1	1	1	1	2	2	3	3	4	4	4	4	6	6	...		
Nodo y	2	4	5	7	4	5	7	8	5	6	7	8	7	8	...		
Pol.	0	0.5	0.66	1	0.5	0.66	1	1	0.5	0.5	0.5	0.5	0.33	0.33	...	0.541	0.94

De estos resultados se observa que la polarización es máxima entre los nodos internos de diferentes comunidades C_1 y C_2 $\{(1, 7), (1, 8), (2, 7), (2, 8), (3, 7), (3, 8)\}$ y mínimo entre nodos internos de la misma comunidad $\{(1, 2), (1, 3), (2, 3), (7, 8)\}$ Obsérvese que la medida

JDJ de polarización podría aplicarse también a particiones nítidas. Por ejemplo, si se considera la partición nítida $P = \{\{1, 2, 3, 4\}, \{5, 6, 7, 8\}\}$, $JDJ_{pol}(V, E, P, Prod, Max)$ se puede ver que solo cuando se mide la polarización entre dos elementos de diferentes comunidades se tiene un valor diferente de 0 - 1, por lo que es fácil comprobar que:

$$JDJ_{pol}(V, E, P, Prod, Max) = \frac{1}{(16)} \sum_{u,v \in V, u \in C_1, v \in C_2} 1 = 16/16 = 1$$

Proposición 6.0.2. *Sea $G = (V, E)$ un grafo con dos comunidades $P = \{C_1, C_2\}$, se tiene que:*

$$JDJ_{pol}(V, E, P, G_O, G_G) = 4 \frac{|C_1|}{|V|} \frac{|C_2|}{|V|} = 4Var(Ber(p)),$$

donde p es la proporción de elementos en la comunidad C_1 .

Prueba. *Es trivial que $G_G(G_O(\mu_1(u), \mu_2(v)), G_O(\mu_2(u), \mu_1(v))) = 1$, si y solo si, los nodos u, v pertenecen a diferentes comunidades y 0 en cualquier otro caso para cualquier agrupación G_G o solapamiento G_O operadores de agregación.*

A partir de esta proposición, se observa que JDJ aplicada a particiones nítidas en redes puede verse como la forma clásica de medir la polarización de dos grupos. También se puede ver que la máxima polarización se alcanza cuando se tienen dos comunidades del mismo tamaño. No obstante, téngase en cuenta que dada la partición P , la medida JDJ es independiente - como sucede con la medida de polarización clásica - de la estructura de la red. Entonces la polarización es la misma si eliminamos las relaciones entre las comunidades C_1 y C_2 :

$$\begin{aligned} JDJ_{pol}(V, E, P, Prod, Max) &= \\ &= JDJ_{pol}(V, E \setminus \{(4.5)\}, P, Prod, Max) = \\ &= JDJ_{pol}(V, E \setminus \{(4.5), (4.6)\}, P, Prod, Max) \end{aligned}$$

Es importante señalar que la polarización desde un punto de vista teórico siempre se enfoca en el hecho de que la polarización es máxima cuando los grupos rompen sus relaciones, por lo que se propone que este desempeño clásico debe mejorarse.

Para proporcionar una forma alternativa de medir la polarización en redes nítidas (V, E, P) usando la medida JDJ , se propone convertir en borrosa una partición nítida, para, posteriormente, obtener el valor de JDJ para la partición difusa. Formalmente, se propone la siguiente definición.

Definición 6.0.3. Sea $(V, E, P) = (V, E, P = \{C_1, C_2\})$ un grafo con dos particiones, y sea G_G, G_O un operador de agregación de agrupación y *overlapping* o solapamiento. Así, la medida de polarización difusa $JDJ_{pol,gre}(V, E, P, G_O, G_G)$ se define como:

$$JDJ_{pol,gre,k}(V, E, P, G_O, G_G) = JDJ_{pol}(V, E, \widetilde{P}_{gre}, G_O, G_G) \quad (6.2)$$

Donde $\widetilde{P}_{gre} = (\mu_1, \mu_2)$ se obtiene con la función *makefuzzy* 6.0.1, $\mu_{C_i}(v) = \frac{|N_k(v) \cap C_i|}{|N_k(v)|}$, donde $N_k(v) = \{u \in V \text{ con distancia}(u, v) \leq k\}$.

6.0.3. Aplicación y resultados computacionales.

En esta sección se calcula la medida JDJ de polarización en la bien conocida *karate club network*. Para mostrar cómo esta medida se ve afectada por pequeñas modificaciones de la red, se agregarán y eliminarán bordes comparando el resultado obtenido con las medidas de polarización basadas en los tamaños de grupo de Esteban y Ray (1994) [28] y la *Medida de polarización de frontera* definida por Guerra et al. (2013) [49] que se presenta a continuación.

Definición 6.0.4 (Medida de Polarización de Guerra et al. (2013)[49]). Sea $G = (V, E, P)$ un grafo con una partición nítida $P = \{C_1, C_2\}$. Entonces

$$Pol_{boundary}(V, E, P) = \sum_{v \in B} \frac{1}{|B|} 2 \left(\frac{|N(v) \cap C_{\delta(v)}|}{|N(v) \cap C_{\delta(v)}| + |N(v) \cap V \setminus \{C_{\delta(v)}\}|} - 0.5 \right),$$

Donde $B = \{u, v \in V \text{ con } (u, v) \in E \text{ y } u \in C_i, v \in C_j \text{ } i \neq j\}$ son los nodos que tienen relaciones con miembros de otro grupo, $N(v) = N_{k=1}(v)$, y $C_{\delta(v)}$ es la comunidad a la que el nodo v pertenece.

Así, considérese la red de amistad de Zachary del Club de kárate que se muestra en la figura 6.2 sin considerar los pesos de las aristas. Esta red presenta las relaciones entre 34 miembros de un club de kárate estadounidense en la década de 1970 y considérese el caso en el que el número de comunidades es de dos, calculado con el algoritmo propuesto por Girman y Newman (2002) [41].

Para comparar el desempeño de las medidas $Pol_{boundary}$, Pol_{crisp} , la función de modularidad clásica $Mod(Q)$ y nuestra propuesta $JDJ_{pol,gre,k}$, se toma el caso de la partición nítida de la red $P = \{C_1, C_2\}$ en dos grupos obtenidos usando el algoritmo de Girvan y Newman y deteniéndose en la primera partición del grafo. En la figura 6.2 y tabla 6.2 se puede ver esta

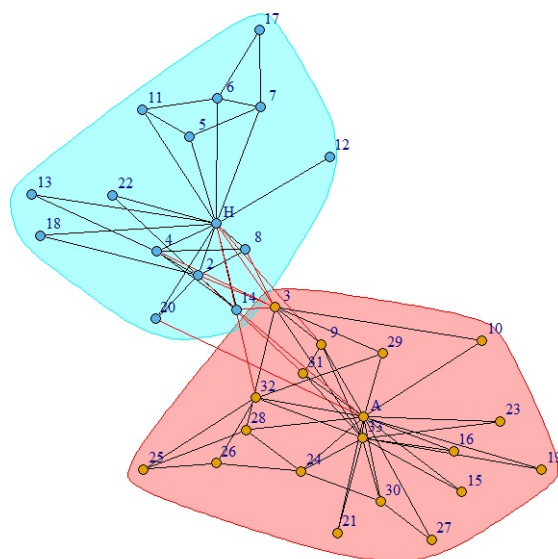


Figura 6.2: Red de Kárate con dos comunidades y modularidad $Q = 0.3599$.

partición nítida. Dado que las medidas $Pol_{boundary}$, Pol_{crisp} , $Mod(Q)$ no pueden tratar con soluciones de detección de comunidades difusas (V, E, \tilde{P}) , aquí solo realiza la comparación para el caso nítido.

Para analizar la robustez de las distintas medidas que se quieren comparar, en las siguientes figuras - 6.3 y 6.4 - se presentan las cuatro medidas anteriormente mencionadas obtenidas para una misma partición - figura 6.2 - $P = \{C_1, C_2\}$ con:

- $C_1 = \{H, 2, 4, 5, 6, 7, 8, 11, 12, 13, 14, 17, 18, 20, 22\}$
- $C_2 = \{A, 3, 9, 10, 15, 16, 19, 21, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32\}$

En la cual la red ha sido modificada añadiendo y eliminando aristas entre nodos.

Obsérvese que dado que $|C_1| = 16$ y $|C_2| = 18$, la medida $Pol_{crisp}(V, E_k, P)$ será la misma para cualquier grafo E_k y, como se mencionó anteriormente, será:

$$Pol_{crisp}(V, E_k, P) = \frac{|C_1| |C_2|}{|V| |V|} = \frac{16 \ 18}{34 \ 34}$$

Tabla 6.2: Particiones obtenidas para la red de Kárate proveniente del algoritmo *makefuzzy* [104].

Nodos	$(\mu_{C_1}(v), \mu_{C_2}(v))$	Nodos	$(\mu_{C_1}(v), \mu_{C_2}(v))$
1	(0.1875,0.8125)	18	(0,1)
2	(0.22,0.77)	19	(1,0)
3	(0.5,0.5)	20	(0.33,0.66)
4	(0.16,0.83)	21	(1,0)
5	(0,1)	22	(0,1)
6	(0,1)	23	(1,0)
7	(0,1)	24	(1,0)
8	(0.25,0.75)	25	(1,0)
9	(0.8,0.2)	26	(1,0)
10	(1,0)	27	(1,0)
11	(0,1)	28	(1,0)
12	(0,1)	29	(1,0)
13	(0,1)	30	(1,0)
14	(0.4,0.6)	31	(0.75,0.25)
15	(1,0)	32	(0.83,0.16)
16	(1,0)	33	(1,0)
17	(0,1)	34	(0.88,0.12)

Se añaden / eliminan aristas en la frontera entre comunidades sin repetir aristas que ya existentes					
Medida	Modularidad	Kleinberg	Nítida	Borrosa	Aristas
Grafo original	0.4012	0.1265	0.9861	0.9569	
Añadimos 1 aristas	0.391	0.1302	0.9861	0.9542	[15,20]
Añadimos 2 aristas	0.3987	0.1401	0.9861	0.9381	[15,20],[25,12]
Añadimos 3 aristas	0.351	0.1414	0.9861	0.9206	[27,8], [33,11], [25,20]
Añadimos 4 aristas	0.3799	0.1019	0.9861	0.859	[19,13], [29,12], [3,17], [31,4]
Añadimos 5 aristas	0.3574	0.14	0.9861	0.9245	[28,13],[25,14],[23,6],[10,14], [34,4]
Añadimos 10 aristas	0.3277	0.1039	0.9861	0.8704	[24,11],[15,14],[28,12],[23,4],[23,13]
Añadimos 15 aristas	0.3204	0.1166	0.9861	0.8427	[33,20],[19,22],[28,14],[28,1],[15,22], [34,2],[30,20],[31,22],[33,7],[24,22], [10,2],[10,7],[16,5],[33,13],[29,18]
Añadimos 20 aristas	0.28	0.1032	0.9861	0.741	[34,17], [30,12], [10,6], [27,6], [28,18], [32,13], [10,17], [29,5], [23,17], [19,20], [33,11], [31,5], [26,12], [21,7], [19,2], [25,5], [32,14], [33,13],[30,4],[15,18]
Eliminamos 1 arista	0.4092	0.1468	0.9861	0.9593	[2,3]
Eliminamos 2 aristas	0.4173	0.1234	0.9861	0.9662	[3,8],[1,32]
Eliminamos 3 aristas	0.4089	0.2114	0.9861	0.9702	[2,3],[3,4],[3,8]
Eliminamos 4 aristas	0.4368	0.1854	0.9861	0.9701	[3,8],[1,3],[1,32],[14,34]
Eliminamos 5 aristas	0.4285	0.2698	0.9861	0.9697	[3,14],[1,3],[3,4],[2,3],[2,31]
Eliminamos 10 aristas	0.4844	NaN	0.9861	0.9861	[1,3],[1,9],[1,32],[2,3],[2,31],[3,4],[3,8],[3,14],[14,34],[20,34]

Figura 6.3: Diferentes Medidas de Polarización para la red de Kárate añadiendo y eliminando aristas entre comunidades.

Se añaden / eliminan aristas en el interior de las comunidades sin repetir aristas que ya existentes					
Medida	Modularidad	Kleinberg	Nítida	Borrosa	Aristas
Grafo original	0.4012	0.1265	0.9861	0.9569	
Añadimos 1 aristas	0.3927	0.1311	0.9861	0.9563	[19,3]
Añadimos 2 aristas	0.3974	0.1265	0.9861	0.9569	[18,6],[24,29]
Añadimos 3 aristas	0.4023	0.1299	0.9861	0.9602	[14,8],[5,14],[10,9]
Añadimos 4 aristas	0.4104	0.1265	0.9861	0.9569	[17,6],[12,7],[30,23],[30,24]
Añadimos 5 aristas	0.3904	0.1289	0.9861	0.9587	[23,16],[30,32],[12,8],[13,2],[11,5]
Añadimos 10 aristas	0.3925	0.1349	0.9861	0.9606	[25,10],[33,19],[21,23],[33,31],[31,21], [17,7],[20,17],[14,18],[17,14],[14,6]
Añadimos 15 aristas	0.3854	0.1368	0.9861	0.9623	[29,33],[9,28],[10,29],[10,9],[26,3],[3,3,24],[23,16],[4,5],[5,6],[8,14],[12,4],[18,8],[12,7],[4,11],[8,6]
Añadimos 20 aristas	0.3974	0.1367	0.9861	0.9615	[15,16],[16,24],[19,24],[26,3],[23,30],[30,16],[23,24],[27,21],[21,31],[27,34], [13,12],[8,20],[5,18],[12,5],[12,6],[17,22],[13,6],[12,20],[2,22],[20,18]
Eliminamos 1 arista	0.4075	0.1265	0.9861	0.9569	[24,30]
Eliminamos 2 aristas	0.382	0.1258	0.9861	0.9569	[16,34],[5,7]
Eliminamos 3 aristas	0.4036	0.1265	0.9861	0.953	[1,20],[21,33],[1,18]
Eliminamos 4 aristas	0.4022	0.124	0.9861	0.9569	[1,10],[1,15],[1,23],[6,17]
Eliminamos 5 aristas	0.4167	0.1083	0.9861	0.9569	[24,28],[3,10],[1,15],[3,9],[1,8]
Eliminamos 10 aristas	0.395	0.1027	0.9861	0.9467	[1,15],[26,32],[1,32],[4,14],[23,33],[1,14],[2,4],[1,6],[24,26],[2,14]

Figura 6.4: Diferentes Medidas de Polarización para la red de Kárate añadiendo y eliminando aristas dentro de comunidades.

Como puede verse en las figuras 6.3 y 6.4, la medida de polarización JDJ tiene en cuenta dos aspectos fundamentales en la polarización: el tamaño de los grupos formados, que también se tiene en cuenta en las medidas de polarización tradicionales - ver Esteban y Ray -, así como la frontera y relaciones que existen entre grupos, lo cual es considerado por $Pol_{Boundary}$. De la tabla anterior se puede ver que dado que el tamaño de los grupos es similar - 18 y 16 -, la polarización nítida es alta y obviamente es la misma para todas las modificaciones de la red - el valor más alto se alcanza en el caso de dos grupos de 17 miembros -. Obsérvese el siguiente desempeño de la propuesta de este estudio:

- Cuando se agrega un enlace entre nodos internos - nodos que no pertenecen a la frontera y están en la misma comunidad como por ejemplo (6, 18) - en la red, la polarización JDJ no cambia - ya que la función de pertenencia de todos los miembros de la red no cambia -. Esto coincide con la idea teórica de que en términos de polarización nada ha cambiado.

- Cuando se agrega un enlace entre dos nodos de la misma comunidad, pero uno de ellos está en la frontera y el otro no: la polarización aumenta, ya que los nodos fronterizos se identifican más con su comunidad y crece su antagonismo con el otro grupo.
- Cuando se agrega un vínculo entre dos comunidades, entonces la polarización disminuye ya que existe más comunicación entre las comunidades.

También es importante ver el principal inconveniente de la medida propuesta por Guerra et al. (2013) [49] ya que no tiene en cuenta el tamaño de los grupos así como los nodos que se encuentran lejos de la frontera. Se puede argumentar que solo considerar el límite parece no ser suficiente para medir la polarización, ya que tener un nivel de polarización alto es invariable al tamaño de los grupos que se analizan.

6.0.4. Conclusiones.

Se considera como un asunto importante señalar las diferencias entre polarización y modularidad, ambas confundidas comúnmente en la literatura al medir la polarización en grafos. Según las figuras 6.3 y 6.4, se puede ver que la modularidad depende de la densidad de los grupos mientras que este no es el caso de la polarización. A pesar de que existe una correlación entre la modularidad y la polarización y los altos niveles de modularidad generalmente implican altos niveles de polarización, estos conceptos son significativamente diferentes. Considérese el ejemplo en el que hay dos grupos - A y B - con el mismo tamaño. En aquellos casos en los que no se modifiquen las relaciones entre el grupo A y el grupo B pero aumenten las relaciones intragrupalas se incrementarán los niveles de modularidad mientras que no ocurre lo mismo con los niveles de polarización. De lo contrario, cuando las relaciones entre A y B aumentan, la modularidad y los niveles de polarización disminuyen.

La medida de polarización que se propone en este capítulo tiene en cuenta dos aspectos principales de la medición de la polarización. En primer lugar, el tamaño de los grupos, que también se tiene en cuenta mediante las medidas tradicionales de polarización aplicadas a un grafo. En segundo lugar, esta medida también incluye cuántas relaciones existen entre grupos o comunidades. En este sentido, se considera el máximo nivel de polarización en aquellos casos en los que sólo hay dos comunidades del mismo tamaño y no hay aristas o vínculos entre ellas.

Así, se considera que esta medida mantiene la idea que proponía Guerra et al. [49], siendo que a mayor número de aristas en la frontera entre

comunidades menor polarización se da, pero también contempla el concepto tradicional de polarización basado en el tamaño de los grupos. Sin embargo, la idea propuesta por Guerra et al. [49] carece de la idea de tomar en consideración el tamaño de los grupos así como los nodos que se ubican lejos de la frontera. De acuerdo con el ejemplo que se mostró en las secciones anteriores, se puede notar que la medida de Guerra et al. [49] no es suficiente para medir la polarización del grafo, ya que solo se centra en la frontera entre grupos. Por lo tanto, no centrarse en el tamaño de la red podría traer algunos problemas a la hora de medir la polarización de manera adecuada. En este sentido, tener un nivel alto de polarización que es invariable al tamaño de los grupos implica que se está considerando dos escenarios diferentes como homónimos, es decir: el escenario que presenta dos grupos con el mismo tamaño *vs* el escenario que presenta un grupo A con 90 % de la población total y un grupo B con los otros 10 %. Además, para las medidas de polarización tradicionales basadas en grafos que solo consideran el tamaño de los grupos, los niveles de polarización son invariantes cuando se modifica el número de bordes entre comunidades, lo que también conduce a presentar niveles de polarización que no reflejan la realidad. En este último ejemplo, si se consideran dos grupos del mismo tamaño, pero que se comunican continuamente entre sí de forma civilizada, no se corresponde con un escenario de polarización. Debido a la agregación de estos dos conceptos aplicados a un grafo borroso, se considera que la medida de polarización propuesta en este estudio cubre las ineficiencias encontradas en los últimos métodos.

Capítulo 7

Una nueva propuesta: La medición de la Polarización desde una perspectiva dinámica

Resumen: En este capítulo se aborda el problema del modelado de polarización con Cadenas de Markov (PMMC - *Polarization Modeling with Markov Chains*). Se propone un modelo probabilístico que proporciona un enfoque novedoso para saber cuál es la probabilidad de que una determinada distribución actitudinal llegue a un estado de polarización social, política o afectiva, entre otras. También cuantifica cuántos pasos se necesitan para alcanzar la polarización para esa distribución. De esta forma, se puede saber el riesgo que presenta una distribución actitudinal para llegar a la polarización en un futuro cercano. Para ello, se establecen unas determinadas premisas sobre las que el modelo se ajusta a la realidad. Además, se compara esta probabilidad con la medida de polarización propuesta por Esteban y Ray y la medida de polarización difusa de Guevara et al. De esta manera, PMMC brinda la oportunidad de estudiar en profundidad cuál es el desempeño de estas medidas de polarización en condiciones específicas. Se encuentra que el modelo propuesto presenta evidencia de que, de hecho, algunas distribuciones presumiblemente mostrarán un mayor riesgo que otras, incluso cuando toda la población presenta la misma actitud. En este sentido, y según el modelo desarrollado, se encuentra que las actitudes moderadas/indecisas presentan un mayor riesgo de polarización que las actitudes extremas y no deberían ser consideradas como un mismo escenario a pesar de que toda la población presente la misma actitud.

7.1. Un nuevo enfoque a la modelización de la Polarización usando Cadenas de Markov

Guevara, J.A., Gómez, D., Castro, J., Gutiérrez, I., Robles, J.M. (2022). A New Approach to Polarization Modeling Using Markov Chains. Communications in Computer and Information Science, vol 1602. Springer, Cham.

Abstract.

In this study we approach the problem of polarization modeling with Markov Chains (PMMC). We propose a probabilistic model that provides an interesting approach to know what the probability for a specific attitudinal distribution is to get to a i.e. social, political or affective Polarization. It also quantifies how many steps are needed to reach Polarization for that distribution. In this way, we can know how risky an attitudinal distribution is for reaching polarization in the near future. To do so, we establish some premises over which our model fits reality. Furthermore, we compare this probability with the polarization measure proposed by Esteban and Ray and the fuzzy polarization measure by Guevara et al. In this way, PMMC provides the opportunity to study in deep what is the performance of these polarization measures in specific conditions. We find that our model presents evidence that in fact some distributions will presumably show higher risk than others even when the entire population holds the same attitude. In this sense, according to our model we find that moderate/indecisive attitudes present a higher risk for polarization than extreme attitudes and should not be considered as the same scenario despite the fact that the entire population maintains the same attitude.

7.1.1. Polarización como proceso dinámico.

De acuerdo con los capítulos anteriores, y con respecto a la medición de la polarización, se debe dar un paso adelante como consecuencia de la tendencia encontrada en la literatura de proponer medidas para identificar los niveles de polarización en un instante de tiempo específico. Aunque algunos estudios han llevado sus objetivos a modelar la influencia interpersonal sobre las actitudes donde podría terminar en situaciones polarizadas (Baldassarri et al., 2007) [5], aún no se ha abordado desde una perspectiva estocástica. En este sentido, las Cadenas de Markov pueden proporcionar los recursos adecuados

para abordar esta tarea. Se considera que este es un enfoque novedoso que permitirá a los investigadores saber, para condiciones específicas y dada una distribución específica de actitud en una población, cuál es el riesgo de alcanzar el nivel máximo de polarización desde un enfoque probabilístico. Las cadenas de Markov también se han utilizado para modelar fenómenos a lo largo de la literatura y en diversos campos como la genética (Nix y Vose, 1992) [84], la gestión de redes urbanas (Osorio et al., 2018) [86] o la relación con el cliente en marketing (Pfeifer y Carraway, 2000) [90].

El objetivo de este estudio es modelar el dinamismo del fenómeno de polarización utilizando Cadenas de Markov. Brindará la oportunidad de comprender en profundidad el desempeño de algunas medidas de polarización propuestas en la literatura (Esteban y Ray, 1994 [28]; Guevara et al., 2020 [50]). Para solucionarlo, se hace uso de las probabilidades de transición y el número de pasos que puede tomar un estado dado - una distribución de actitud específica - para llegar al máximo nivel de polarización. Además, se contrastan los resultados con niveles específicos de polarización proporcionados por medidas de polarización.

7.1.2. Un nuevo problema: Modelización de la Polarización usando Cadenas de Markov.

En esta sección se propone un modelo probabilístico novedoso para la Modelización de la Polarización con Cadenas de Markov (PMMC - *Polarization Modelling with Markov Chains* -). Así, se establecen algunas premisas para asegurar que el modelo probabilístico tenga un grado mínimo de concordancia sobre cómo debe comportarse la polarización según la teoría del fenómenos. Este enfoque es completamente nuevo en la literatura.

7.1.2.1. Hipótesis.

A continuación, se presenta la propuesta con el objetivo de cubrir los principales conceptos básicos del comportamiento de la polarización.

1. **Número de individuos para modelar.** Se refiere al número de individuos a considerar en el modelado de polarización, siendo $N = \{1, \dots, k, \dots, n\}$ el conjunto completo.
2. **Medición de actitud.** Este punto se refiere a la variable por la cual se miden las actitudes de los individuos por la cual se mide la polarización. Sea $Z = \{1, \dots, z\}$ una variable categórica para medir la actitud de

N elementos con z niveles. Se asume que Z tiene dos polos, 1 y z . Finalmente, sea Z_k la posición del individuo k en Z , donde $k \in N$.

3. **Independencia del comportamiento entre individuos.** Se propone especificar si el comportamiento de cada individuo es independiente - o no - con respecto al de los demás. En otras palabras, el investigador tiene que decidir si $\forall k, l \in N$ con $k \neq l$, Z_k es independiente de Z_l o no.
4. **Naturaleza de los polos.** Cuando las personas alcanzan actitudes extremas, tienden a radicalizarse, lo que conduce a una menor probabilidad de cambio en sus actitudes. Este hecho se tomará en consideración para decidir si un individuo que ocupa una posición extrema presenta probabilidad de cambiar de actitud o no. En este escenario, donde las personas están en posiciones extremas, el investigador tiene que decidir si la cadena termina - porque las actitudes extremas presumiblemente no cambiarán - o no.
5. **Grado de inmovilidad.** Todo individuo presenta una probabilidad > 0 de permanecer en la misma posición a lo largo de la variable con la que se miden sus actitudes. Asimismo, se debe determinar si la cantidad de esta probabilidad depende o no de las unidades de tiempo.
6. **Cambio de actitud.** Se debe considerar el rango del cambio de actitud por unidad de tiempo. Se entiende que un parámetro a incluir es el grado en que los individuos cambian sus actitudes. Dependiendo de este parámetro, pueden aparecer diferentes enfoques siendo (a) *gradualidad*, considerando que la actitud de los individuos no cambia abruptamente sino suavemente y (b) *brusquedad* que establece que los individuos pueden saltar abruptamente dentro de los valores de la variable actitudinal.
Sea Z una variable categórica, Z_k el valor que el individuo k tiene en Z . Entonces, tenemos que $Z_k \pm d$ denota los posibles valores en Z que podría tomar el individuo k en la siguiente unidad de tiempo, siendo d las unidades a saltar en Z .
7. **Simetría del cambio.** Esta premisa se refiere a la simetría de las probabilidades de cambio de los individuos de sus posiciones actuales a otras. Así, es importante indicar si las probabilidades de cambio son las mismas según: (a) *cambio de dirección* y (b) *cercanía a los polos*, es decir, si para un individuo $k \in N$, Z_k cambiará dependiendo de $\pm d$ y μ_{X_A}/μ_{X_B} .

7.1.2.2. Un nuevo modelo probabilístico.

Una Cadena de Markov en PMMC es un proceso estocástico en tiempo discreto en el que las actitudes medidas por la variable Z cambian a lo largo del tiempo. Al modelar la polarización, las Cadenas de Markov permiten representar las probabilidades de transición entre distribuciones de actitudes en una población. Se puede inferir la evolución de las actitudes de una población para saber cuándo, cómo y en qué condiciones se polarizan estos individuos. En este sentido, la distribución actitudinal de la población se mueve de un determinado estado a otro con una probabilidad específica.

Definición 7.1.1 (Estados en PMMC). Se denomina un estado en términos de PMMC a una distribución específica y estática de una población $N = \{1, \dots, k, \dots, n\}$ a lo largo de la característica actitudinal Z donde $Z = \{1, \dots, z\}$. Así, sea S el conjunto completo de estados con longitud $z^n = s$ donde $i \in S$ si y solo si $i = [Z_1, \dots, Z_k, \dots, Z_n]$. Asimismo, definimos como estados polarizados las distribuciones en las que una parte significativa de la población se sitúa en un polo de la variable y otra parte significativa de la población se sitúa en el otro polo.

Definición 7.1.2 (Probabilidades de transición en PMMC). Dado $i, j \in S$ se define como P_{ij} a la probabilidad de llegar a una opinión poblacional j a partir de una opinión poblacional i . Además, se dice que i es un estado adyacente a j cuando presenta una probabilidad de transición > 0 de i a j . Los valores de P_{ij} dependen directamente de las reglas establecidas en las hipótesis propuesta anteriormente.

Observación 10. Tenga en cuenta que habiendo definido previamente los estados en la definición 7.1.1 y las probabilidades de transición en la definición 7.1.2, se puede calcular la matriz de transición T con dimensiones $s \times s$, de acuerdo con las secciones anteriores.

Ejemplo 7.1.1. Sea $\alpha = 0.8$ la probabilidad de permanecer en la misma posición para un individuo dado, $n = 2$, $z = 3$, $d = 1$ y se asume simetría en el cambio de posición. Sea $S = \{[1, 1], [1, 2], [1, 3], [2, 1], [2, 2], [2, 3], [3, 1], [3, 2], [3, 3]\}$ el conjunto completo de estados. Además, se asume como posiciones absorbentes los valores extremos de Z . Entonces, se tiene que:

- $P_{[1,1][1,1]} = P_{[3,3][3,3]} = P_{[1,3][1,3]} = P_{[3,1][3,1]} = 1.$
- $P_{[1,2][1,2]} = P_{[2,1][2,1]} = P_{[2,3][2,3]} = P_{[3,2][3,2]} = 1 \cdot 0.8 = 0.8$
- $P_{[2,2][2,2]} = 0.8 \cdot 0.8 = 0.64$

- $P_{[1,2][1,1]} = P_{[1,2][1,3]} = P_{[2,1][1,1]} = P_{[2,1][3,1]} = P_{[2,3][1,3]} = P_{[2,3][3,3]} = P_{[3,2][3,1]} = P_{[3,2][3,3]} = (1 - 0.8)/2 = 0.1$
- $P_{[2,2][1,1]} = P_{[2,2][1,2]} = P_{[2,2][1,3]} = P_{[2,2][2,1]} = P_{[2,2][2,3]} = P_{[2,2][3,1]} = P_{[2,2][3,2]} = P_{[2,2][3,3]} = (1 - 0.64)/8 = 0.045$
- Finalmente, el resto de probabilidades de transición restantes presentan un valor igual a 0.

De acuerdo con esto, la matriz de transición T es la siguiente:

Tabla 7.1: Matriz de transición para $\alpha = 0.8$ y $z = 3$

	1,1	1,2	1,3	2,1	2,2	2,3	3,1	3,2	3,3
1,1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1,2	0.10	0.80	0.10	0.00	0.00	0.00	0.00	0.00	0.00
1,3	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
2,1	0.10	0.00	0.00	0.80	0.00	0.00	0.10	0.00	0.00
2,2	0.045	0.045	0.045	0.045	0.64	0.045	0.045	0.045	0.045
2,3	0.00	0.00	0.10	0.00	0.00	0.80	0.00	0.00	0.10
3,1	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00
3,2	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.80	0.10
3,3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Una vez que se define la matriz de transición, se pueden calcular todos los procesos de la Cadena de Markov, como las probabilidades de transición en t pasos o determinar los estados absorbentes. PMMC proporciona información útil: (1) probabilidad de que una población alcance finalmente un estado polarizado, (2) número de pasos necesarios para alcanzar estados polarizados, (3) distribución de una población en n etapas dada una distribución inicial o (4) número de visitas a estados recurrentes dadas ∞ etapas, entre otros. De esta forma, la agregación de esta información es una herramienta útil para la medición de la polarización.

7.1.3. Experimento y resultados.

En esta sección se lleva a cabo un experimento con un ejemplo ilustrativo de PMMC en el que asumimos valores específicos para la hipótesis propuesta en la última sección. Para ello, una vez establecidas las reglas y caracterizada la matriz de transición, utilizamos el paquete de R “*markovchain*”[107] para calcular los diferentes procesos de la Cadena de Markov.

1. **Número de individuos a modelar.** Debido a que la mayoría de las medidas de polarización se calculan mediante la comparación entre dos individuos y, por motivos de simplicidad, se aplica PMMC para el escenario más simple en el que se tiene dos individuos, siendo $n = 2$. Nótese que este ejemplo es equivalente a la situación en la que el 50 % de la población presenta el mismo valor, y los otros 50 % de la población presentan otro valor, igual para todos.
2. **Medición de actitud.** Se considera la medición actitudinal a través de una escala de categórica $Z = \{1, 2, 3, 4, 5\}$.
3. **Independencia del comportamiento entre individuos.** Se asume que no hay independencia entre el comportamiento de los individuos. En este sentido, los estados en los que no hay individuos que cambien su comportamiento en más de una categoría en Z son equiprobables.
4. **Naturaleza de los polos.** Se asumen los valores de Z , 1 y 5 como posiciones absorbentes. Así, los estados en los que todos los individuos tienen posiciones extremas se consideran estados absorbentes, siendo $[1, 1]$, $[1, 5]$, $[5, 1]$ y $[5, 5]$.
5. **Grado de inmovilidad.** Todo individuo presenta una probabilidad de permanecer en su misma posición de $\alpha > 0$. En este ejemplo, se varía el parámetro α para observar diferentes resultados. Así, se tiene $\alpha \in [0.1, 0.9]$ de 0.1 en 0.1. Evitamos valores extremos para $\alpha = 0$ y $\alpha = 1$ para no contradecir nuestra premisa sobre la *naturaleza de los polos*.
6. **Cambio de actitud.** En este ejemplo seguimos el enfoque de *gradualidad*, asumiendo que los individuos no cambian abruptamente sus actitudes. Se supone que un individuo k solo puede cambiar su opinión a valores adyacentes de la variable Z por unidad de tiempo, donde $d = 1$ siendo de Z_k a $Z_k \pm 1$ en el siguiente paso. Por ejemplo, si $Z_k = 3$, Z_k solo puede tomar los valores 2 y 4 en el siguiente paso de la Cadena de Markov.
7. **Simetría del cambio.** Se asume que las probabilidades de cambiar de posición son iguales en cualquier dirección ($\pm d$), y no se ven afectadas por la cercanía a los polos (μ_{X_A} o μ_{X_B}).

De acuerdo con lo anterior, se tiene que $S = \{[1, 1], [1, 2], [1, 3], \dots, [3, 5], [4, 5], [5, 5]\}$ es el conjunto completo de estados, con longitud $s = z^n = 5^2$. Posteriormente, se calcula tantas matrices

de transición $T = 25 \times 25$ como diferentes valores de α donde $\alpha \in [0.1, 0.9]$ de 0.1 en 0.1. Una vez que todas las matrices de transición están listas, se simulan las Cadenas de Markov para saber cuál es la probabilidad de que un estado no polarizado se polarice y cuántos pasos requiere el proceso para alcanzar estos estados específicos según cada valor de α . Para hacerlo, se ejecuta la función *rmarkovchain* desde el paquete R [107] “*markovchain*”, el cual permite generar cientos de pasos futuros de acuerdo con el proceso estocástico subyacente de T . Primero, se consideran todos los estados a partir de S menos los absorbentes ($[1, 1]$, $[1, 5]$, $[5, 1]$ y $[5, 5]$) como estados iniciales. Después, se comienza la simulación hasta que dicho estado inicial alcance un estado absorbente, de modo que se calcula la probabilidad frecuentista de que este estado alcance uno absorbente y se contabilizan cuántos pasos se requieren para que este estado los alcance. Para ello, se calculan 10000 iteraciones para cada estado inicial y valor de $\alpha \in [0.1, 0.9]$ por 0.1.

Así, se calculan los resultados medios de todas las 10000 iteraciones para cada α y para cada estado inicial como unidades de análisis. A continuación, se observan las probabilidades y los pasos que podría tomar un estado transitorio para alcanzar ambos estados polarizados $[1, 5]$ y $[5, 1]$. La desviación estándar media de cada estado para alcanzar ambos estados de polarización en términos de pasos a lo largo de la variación de α es 17.341 mientras que la *sd* para la probabilidad de alcanzarlos es 0.004. Se puede concluir, como consecuencia, que la variación de α no afecta la probabilidad de un determinado estado de alcanzar otro estado polarizado, sino los pasos requeridos. Para ilustrar esto, se muestra en la figura 7.1 los pasos medios tomados por cada estado para alcanzar ambos estados de polarización a lo largo de la variación de α .

En la figura 7.2 se pueden observar las probabilidades medias y los pasos medios a lo largo de todos los valores de α para que cada estado transitorio alcance ambos estados de polarización, donde a mayor probabilidad y menor número de pasos requeridos implica un mayor riesgo de polarización. Como se puede observar en los resultados, la figura 7.2 muestra una coherencia entre la cercanía de los estados a los estados polarizados y su probabilidad de alcanzarlos así como un menor número de pasos requeridos para alcanzarlos. Gracias a esto, se puede ver la evolución de dos grupos principales - o pareja de individuos - con igual número de individuos y misma actitud intragrupal a lo largo del futuro según su distribución de actitudes y su probabilidad de polarizarse. A partir de este ejemplo, se encuentra evidencia de que algunas distribuciones presentan un mayor riesgo de polarización en el futuro cercano que otras, incluso cuando toda la población mantiene la misma posición de

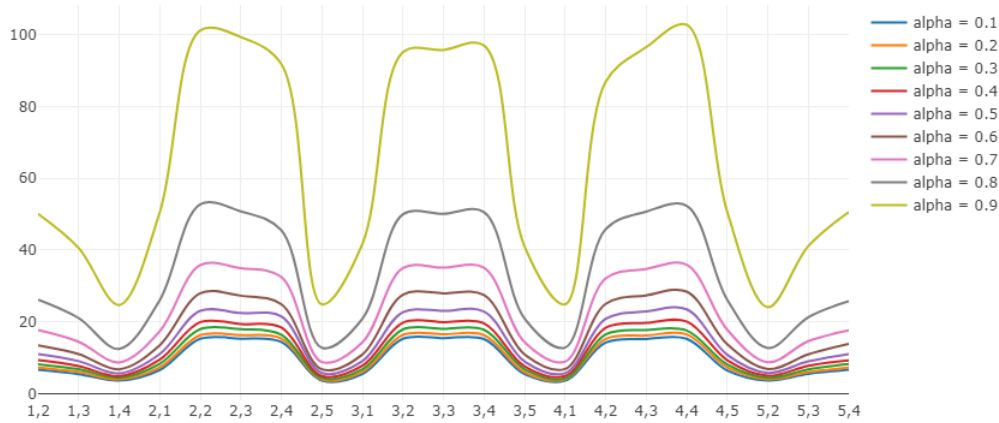


Figura 7.1: Pasos medios para alcanzar estados de polarización a través de la variación del parámetro α .

actitud, como es el caso de los estados [2, 2] o [3, 3]. Además, se observa que a pesar de que [3, 1] y [3, 2] presentan la misma probabilidad de alcanzar un estado polarizado, los pasos medios necesarios para alcanzar un estado de polarización para el estado [3, 1] son menores, presentando, así, un mayor riesgo de polarización que [3, 2].

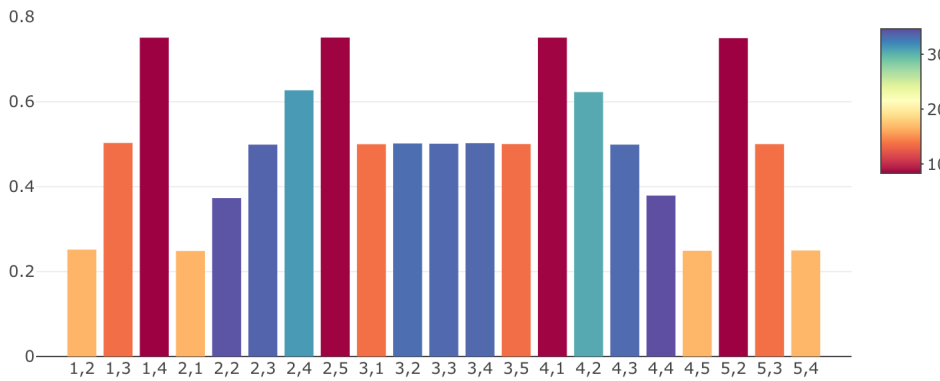


Figura 7.2: Probabilidades medias - barras - y pasos - color - para alcanzar estados polarizados.

Finalmente, se calculan los valores de polarización de dos medidas de polarización ER [28] y JDJ donde $\phi = product$ [50] a lo largo de todos los estados transitorios del ejemplo propuesto. Además, en la Figura 7.3 se

representan las probabilidades sumadas de alcanzar ambos polos para cada estado inicial. Nótese que las probabilidades de alcanzar un estado polarizado no cambian a lo largo de la variación de α , mostrándose en la Figura 7.3 las probabilidades reales de polarización según el modelo propuesto. Así, se encuentra que ambas medidas siguen la misma tendencia que las probabilidades proporcionadas por PMMC. De hecho, las correlaciones entre las medidas de polarización y las probabilidades de PMMC son altas, siendo $r_{ER} = 0.759$ y $r_{JDJ} = 0.976$. Estos resultados están ligados a los parámetros específicos impuestos al modelo, lo que es importante a la hora de analizar los resultados. Además, estos resultados arrojan luz sobre el desempeño de estas medidas según condiciones específicas, concluyendo que PMMC puede ayudar a los investigadores a saber en qué condiciones una medida de polarización puede ser más adecuada que otra.

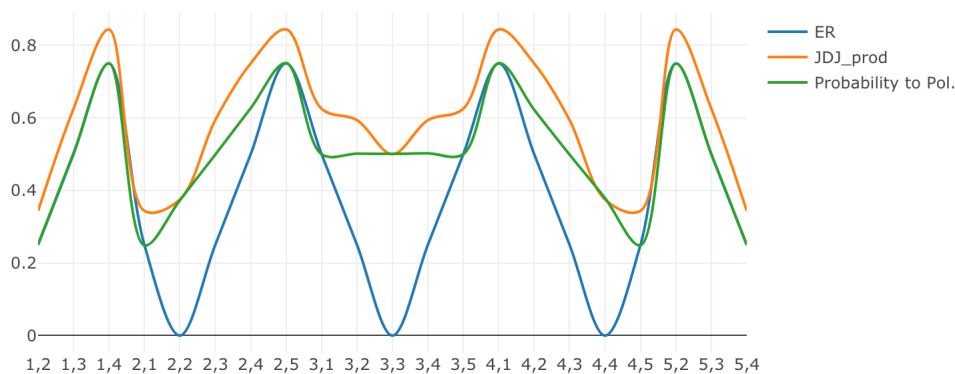


Figura 7.3: Valores de polarización y probabilidad de polarización para cada estado inicial.

7.1.4. Conclusiones.

En este capítulo se aborda el problema de modelado de polarización utilizando Cadenas de Markov. Gracias a los procesos estocásticos que subyacen a las Cadenas de Markov se pueden modelar los fenómenos sociales en términos de probabilidad y eventos futuros. En este caso, se entienden por estados las distribuciones actitudinales específicas de una población y se parte de unas premisas establecidas que permiten modelar la polarización desde un punto de vista realista. Gracias al establecimiento y variación de estos parámetros se pueden simular diferentes situaciones que brindan la oportunidad de saber cómo podría evolucionar en el futuro un escenario específico en tér-

minos de probabilidad. Gracias a la simulación de PPMC se puede conocer el riesgo probabilístico de una determinada distribución actitudinal para llegar finalmente a una distribución polarizada, así como los pasos requeridos en unidades de tiempo. Se ha ilustrado un ejemplo en el que se han medido las actitudes con una escala likert de 5 categorías que bien podría ser una actitud política, por ejemplo, en la población más simple de todas, siendo $n = 2$. Este escenario puede entenderse como un par de individuos o la existencia de dos grupos con alto grado de homogeneidad dentro de grupos en los que todos sus individuos ocupan la misma posición en una escala de actitudes.

La simulación de este ejemplo ha traído algunos resultados interesantes. En primer lugar, se puede conocer qué estados presentan un mayor riesgo de polarizarse en un futuro próximo y cuántas unidades de tiempo les puede llevar alcanzar el nivel máximo de polarización. Asimismo, PPMC parece apoyar la premisa de que en aquellas situaciones en las que toda la población mantiene la misma actitud no son iguales en términos de riesgo de polarización. De esta forma, se concluye que las posiciones neutrales o centrales presentan un mayor riesgo de polarización que las posiciones extremas, siendo significativamente diferente el escenario en el que toda la población se sitúa por el centro de la variable, lo que podría entenderse como una posición indecisa con respecto al escenario en el que toda la población tiene una actitud radicalizada. Así, este último escenario apoya el hecho de que las actitudes radicales son más difíciles de cambiar que las moderadas y por lo tanto este matiz debería de ser considerado por las medidas de polarización, hecho que sí ocurre en el caso de la medida *JDJ*.

Finalmente, también se ha comparado la probabilidad de que una determinada distribución esté polarizada con dos medidas de polarización. Se ha encontrado un alto grado de correlación entre esta probabilidad y las puntuaciones de provenientes de las medidas de polarización, donde la medida difusa parece encajar mejor con el enfoque PPMC que la medida tradicional. No obstante, es importante resaltar que esta comparación se ha realizado bajo condiciones, premisas e hipótesis específicas mediante las cuales se ha modelado la polarización y diferentes ejemplos pueden mostrar resultados diferentes.

Conclusions and Future Work

Resumen: The intention of the last chapter is to reflect on the work presented, the planned tasks that have been undertaken, and the way they have been reached, in addition to pose the future research work which could stem from this dissertation. This chapter is structured into three parts. Firstly, the conclusions derived from this research work, presented from a philosophical point of view. Then, it is showed the relation between the tasks enunciated in the Introduction with the obtained results and the contributions related to them. Finally, future research lines that could be derived from this thesis are enunciated.

Conclusions

The concept of polarization turns out to be rich and complex. For years, the phenomenon of polarization has been addressed, finding a growing number of publications in the last two decades. Although relative consensus has been found in the conceptualization of polarization in the literature, understanding it as the division of society into two antagonistic groups of similar size, this has not happened with its metrics.

The measurement of polarization is wide, varied and heterogeneous, having been approached from different perspectives, each one from its own nature and distinction. However, this has led to a lack of agreement regarding the measurement of such phenomenon. The polarization measures proposed throughout the literature have been essential to develop and deepen the concept of polarization. However, this diverse approaches have led to different ways whose direction has turned out to be quite different. Thus, the different measures of polarization are different in nature, both conceptually and metrically. As a consequence, the literature is conspicuous by the lack of deepening in the measurement of polarization where, in addition, an adequate

synergy has not been found between the more technical aspects of it and the conceptual ones.

Thus, there are measures that focus on the measurement of polarization from different concepts, being, on some occasions, confused by dispersion or fragmentation. On other occasions, a lack of rigor is found in the behavior of the measure with respect to the concept of polarization, showing unrealistic values in certain situations. On the other hand, a high dependence on the information available for the polarization measurement has been found, as well as on the need for a specific type of *input*, - i.e.: numerical or categorical data -. Furthermore, the metric conceptualization of polarization varies greatly depending on the type of data to be used. As a consequence of all of the above, advances in polarization measurement have turned out to be slow and scattered.

One of the main objectives of this thesis is to propose a polarization measure that would be able to overcome the deficiencies found in the literature. The proposal of a polarization measure that addresses the problem from a realistic and coherent perspective from a metric and conceptual point of view has become of great importance, especially in recent years with the increase in scientific publications that study the phenomenon of polarization.

This proposal was addressed through fuzzy logic, opening new frontiers for the measurement of sociological phenomena. Human and social behavior is characterized by great complexity, containing a high number of nuances and variables that influence it. Fuzzy sets provide us with the necessary resources to address social aspects of reality that, in the vast majority of cases, lack an exclusively crispy nature. Thus, after the study and observation of highly polarized scenarios, the idea of including the concept of radicalization in the measurement of polarization was picked up, facing the metric of such phenomenon from a totally innovative perspective that, together with the use of fuzzy sets, it offers values consistent with its conceptualization.

Likewise, various applications of polarization measurement have been explored in other scenarios, such as the case of community detection problems in graphs. Thus, adding as extra information to the classic community detection algorithms the calculation of the polarization between pairs of nodes, together with the application of fuzzy graphs, the community detection problem was taken to the next level. From this new perspective, the new community detection algorithm offered a reality check not achieved by other models. Thus, for the first time in the literature, community detection problems are addressed from a combined perspective, including into the model both the network's relational information and its ideology.

Likewise, the fuzzy bipolarization measure JDJ was extended to the case

of graphs. In this sense, the objective is to measure polarization, not only from an attitudinal or ideological point of view with the first proposal, but to measure structural polarization from a fuzzy logic. To do this, using the fuzzy graphs as a base, the degrees of belonging of the nodes to their respective communities are obtained, allowing the measurement of polarization in networks from a novel perspective. Compared to the rest of polarization measures proposed in the literature, this new approach allows polarization to be measured considering their degrees of belonging to such communities. Applications of different measures in different scenarios showed a better adequacy of the measure proposed in this work with respect to those previously presented in the literature.

Finally, the measurement of polarization was approached from a totally new perspective in the literature through stochastic models such as Markov Chains. Thanks to the approach of a model that aims to simulate the phenomenon of polarization through various parameters, a stochastic mathematical model was generated capable of predicting the levels of polarization over time given a certain distribution of polarization in the population. From this new perspective, it is possible to know the risk of polarization that a certain society presents according to its attitudinal distribution towards an ideological axis.

With all this, this thesis has delved into the measurement of polarization from various novel perspectives, always facing the measurement of the phenomenon from a hybrid technical-theoretical perspective in order to address the problem from a realistic position. The synergy created as a result of the union of both positions, theoretical and mathematical, has characterized the work developed by a novel proposal in the literature that is expected to bring greater - and better - advances in the future.

Relation between proposed tasks and contributions

In this section, the tasks initially proposed in the introduction of this thesis will be addressed along with their resolution and the scientific contributions produced as a consequence of them.

Task 1: To study and collect the main characteristics and measures of polarization proposed in the literature in order to explore their elements, nature and create a taxonomy around them. Likewise, a study will be carried out to determine their performance in different polarization scenarios.

- **1.1.** To make a bibliographical search and compilation around the

metric and characteristics of polarization.

- **1.2.** To create a taxonomy around the polarization measures proposed in the literature according to their characteristics, especially in relation to the type of data they need to be computed.
- **1.3.** To compute the polarization measures collected for their application to different polarization scenarios and to observe their performance against certain population distributions.

Regarding the task **1.1.** was carried out as part of chapter 2 section 2.2. Thus, the tasks **1.2.** and **1.3.** were addressed through the chapter 3. This objective is considered to be of special importance as it provides the basis for understanding the measurement of polarization in the literature and achieving a broad and rich vision that allows new measures to be proposed.

Task 2: To analyze and identify polarization in the digital debate in order to detect the origins, consequences, and agents of polarization in networks.

- **2.1.** To study the role of bots in the digital debate as polarizing agents and generators of hate.
- **2.2.** To analyze the polarization around the digital debate in the cases of La Manada and #Cuéntalo.

The tasks **2.1.** and **2.2.** have been addressed throughout the chapter 4 in its two sections. The achievement of these objectives is of vital importance for understanding the behavior of society in certain scenarios of presumed polarization. Thus, the results obtained provide a crucial insight when approaching the design of a polarization measurement. The contributions associated with these objectives are the following:

Contribution 1: Robles, J. M., Atienza, J., Gómez, D., & Guevara, J. A. (2019). La polarización de La Manada El debate público en España y los riesgos de la comunicación política digital. *Tempo Social*, 31 (3), 193-216.

Contribution 4: Guevara, J. A., Atienza-Barthelemy, J., Gomez Gonzalez, D., & Robles, J. M. (2021). Polarization and incivility in digital debates on womens rights in Spain. Not just a matter of machismo. *Journal of Gender Studies*, 1-15.

Contribution 6: Martínez, A., Guevara, J.A., & Jiménez, A. (Aceptado, Enero 2023). La influencia de los bots políticos en Twitter durante la crisis del COVID-19 en España. *REIS: Revista Española de Investigaciones Sociológicas*.

Task 3: To design a measure of polarization that makes up for some of the shortcomings found in those already proposed in the literature and that addresses the key aspects of polarization detected in the previous objectives.

- **3.1.** To address the measurement of polarization through fuzzy sets in order to propose a new measure of polarization.
- **3.2.** To apply the polarization measure to a real case - the digital debate on networks about COVID-19 - in the social big data analysis context.
- **3.2.** To explore other applications of the proposed polarization measure in improving community detection algorithms in graphs.

The tasks **3.1.**, **3.2.** and **3.3.** have been solved in the chapter 5. Thus, the aim **3.1.** is one of the fundamental pillars of this thesis, building, for the first time in the literature, a measure of fuzzy polarization. The aim **3.2.** is the culmination of the design of a measure, being applied to a real case and obtaining polarization values based on the novel premises proposed in the previous tasks. Finally, the objective **3.3.** allows exploring the uses of a polarization measure beyond the simple application of the measure from a traditional perspective. In this way, it was possible to improve the community detection algorithms proposed in the literature. Including the values of the polarization measure as extra information, more realistic communities were built. The publications associated with these goals are:

Contribution 2: Guevara, J.A., Gómez, D., Robles, J.M., Montero, J. (2020). Measuring Polarization: A Fuzzy Set Theoretical Approach. *Communications in Computer and Information Science*, vol 1238. Springer, Cham. https://doi.org/10.1007/978-3-030-50143-3_40

Contribution 3: Gutiérrez, I., Guevara, J. A., Gomez, D., Castro, J., & Espínola, R. (2021). Community Detection Problem Based on Polarization Measures: An Application to Twitter: The COVID-19 Case in Spain. *Mathematics*, 9(4), 443.

Contribution 5: Robles, J. M., Guevara, J. A., Casas-Mas, B., & Gómez, D. (2022). When negativity is the fuel. Bots and political polarization in the COVID-19 debate. *Comunicar*, 30(71), 63-75.

Task 4: To adapt the proposed polarization measure to graphs.

- **4.1.** To propose a polarization measure based on fuzzy sets applicable to graphs.
- **4.2.** To compare the proposed polarization measure with the most used polarization measures in graphs.

The translation of the measurement of polarization to the case of graphs allows such a phenomenon to be measured, not only from an attitudinal or ideological perspective, but also from a structural one. The tasks **4.1.** and **4.2.** address this logic, being raised throughout the chapter 6. The scientific contribution associated with these objectives is the following:

Contribution 7: Simón de Blas, C., Guevara, J.A., Morillo, J., Gómez González, D. (2022). Polarization Measures in Bi-partition Networks Based on Fuzzy Graphs. *Communications in Computer and Information Science*, vol 1601. Springer, Cham.

Task 5: To explore and propose a model that makes it possible to address polarization from a dynamic perspective.

- **5.1.** to model the phenomenon of polarization from a dynamic perspective using Markov Chains.
- **5.2.** To run the model proposed above with a toy example to study its performance.
- **5.3.** To compare the results of the dynamic polarization model with different measures of polarization.

Finally, the approach of the tasks **5.1.**, **5.2.** and **5.3.**, addressed in the chapter 7, have made possible the approach of the measurement of polarization from a totally new perspective in the literature. Thus, a stochastic model was created that allows predicting polarization levels over time, considering this phenomenon as a dynamic process . The scientific publication associated with these objectives is the following:

Contribution 8: Guevara, J.A., Gómez, D., Castro, J., Gutiérrez, I., Robles, J.M. (2022). A New Approach to Polarization Modeling Using Markov Chains. Communications in Computer and Information Science, vol 1602. Springer, Cham.

Future research lines

Throughout this thesis, and as a consequence of achieving the tasks proposed, novel advances have been proposed that address some of the most important aspects of polarization measurement, such as the proposal of a polarization measurement with a more synchrony with the conceptual proposals, the translation of the same to the case of graphs, or even the measurement of polarization from a dynamic perspective. With this, a set of lines of research of great interest for the measurement of polarization have been opened.

Regarding the measure JDJ , some directions for future research are suggested. In relation to the membership functions, it is considered of vital importance to deepen their study. The membership functions used to measure the radicalization of individuals have turned out to play a fundamental role in the values offered by the measure. Thus, the contemplation of different functions of membership, as well as the exchange between them according to certain scenarios, represents an object of study of great interest. Regarding the measurement itself, it has been approached from a bipolar perspective - understanding a single axis with two extreme poles -, so approaching the measurement from a multidimensional conceptualization will allow polarization to be measured from novel positions in the field of fuzzy logic. Also, this will expand the number of variables that are included in the polarization measurement.

Likewise, the above considerations can be transferred to the JDJ measure applied to graphs - $JDJ_{pol,gre}(V, E, P, G_O, G_G)$ -, allowing the exploration of new ways of understanding the membership of the nodes to their communities, as well as addressing the measurement of polarization with the consideration of multiple communities in the graph. In this sense, the possible lines of research are divided into two. In the first place, to categorically order the different communities detected in the network according to an ideological axis, in order to know the two most extreme groups - poles - and the intermediate values. However, this consideration would approach the measurement of polarization in graphs from a bipolar perspective. On the other hand, extending the measure to a multipolar dimension would allow measuring the polarization in graphs with > 2 communities from a new view.

For both the case of the JDJ measurement and for $JDJ_{pol,gre}(V, E, P, G_O, G_G)$, it is vitally important to consider the risk of confusing the polarization measurement with fragmentation or segregation. In fact, this distinction does not seem to be clear in the literature, considering it essential to deepen the study of these terms to define the borders between them in order to finally propose multipolar polarization measures from a logical and coherent perspective.

Finally, regarding the measurement of polarization from a dynamic perspective, it was approached through Markov Chains, being one of the simplest mathematical models to contemplate phenomena in time. The consideration of other more sophisticated time series models will allow the possibility of facing this afternoon in a more complex way. Likewise, the parameters considered in the model presented in chapter 7 are a first approximation to the modeling of the polarization phenomenon. Studying in greater depth those aspects that influence the behavior of this phenomenon will allow the creation of models that are much more adjusted to reality.

Conclusiones y Futuras Investigaciones

Resumen: Conclusiones

Conclusiones

El concepto de polarización resulta ser rico y complejo. Desde hace años, se ha abordado el fenómeno de polarización, encontrando un creciente número de publicaciones en las últimas dos décadas. Si bien se ha encontrado relativo consenso en la conceptualización de la polarización en la literatura, entendiéndola como la división de la sociedad en dos grupos antagónicos de tamaño similar, esto no ha ocurrido así con su métrica.

La medición de la polarización es amplia, variada y heterogénea, habiendo sido abordada desde diferentes perspectivas, cada una desde su propia naturaleza y distinción. Sin embargo, esto ha llevado a una falta de acuerdo con respecto a la medición de dicho fenómeno. Las medidas de polarización propuestas a lo largo de la literatura han sido fundamentales para desarrollar y profundizar en el concepto de la polarización. Sin embargo, esta aproximación diversa ha conducido a diferentes ramas cuya dirección ha resultado ser diferente. Así, las diferentes medidas de polarización son de naturaleza distinta, tanto conceptual como métricamente. Como consecuencia, la literatura brilla por la falta de profundización en la medición de la polarización donde, en adición, no se ha encontrado una sinergia adecuada entre los aspectos más técnicos de la misma y los conceptuales.

Así, se observan medidas que enfocan la medición de la polarización desde conceptos diferentes siendo, en algunas ocasiones, confundida por dispersión o fragmentación. En otras ocasiones, se encuentra una falta de rigurosidad en el comportamiento de la medida con respecto al concepto de polarización, mostrando valores irreales en determinadas situaciones. Por otro lado, se

ha encontrado una alta dependencia sobre la información disponible para la medición de la polarización, así como de la necesidad de un específico tipo de *input*, - e.g.: numérico, categórico, etc. -. Además, la conceptualización métrica de la polarización varía mucho en función al tipo de dato que ha de usarse. Como consecuencia de todo lo anterior, los avances en la medición de la polarización han resultado ser lentos y dispersos.

Uno de los objetivos principales de la presente tesis es el de proponer una medida de polarización que fuese capaz de sobreponerse a las deficiencias encontradas en la literatura. La propuesta de una medida de polarización que abordase el problema desde una perspectiva realista y coherente desde un punto de vista métrico y conceptual se ha vuelto de gran importancia, especialmente en los últimos años con el aumento de publicaciones científicas que estudian el fenómeno de la polarización.

Esta propuesta se abordó a través de la lógica difusa, abriendo nuevas fronteras para la medición de fenómenos sociológicos. El comportamiento humano y social se caracteriza por una gran complejidad, conteniendo un alto número de matices y variables que influyen en el mismo. Los conjuntos borrosos nos aportan los recursos necesarios para abordar aspectos sociales de la realidad que, en la gran mayoría de las ocasiones, carecen de una naturaleza exclusivamente nítida. Así, tras el estudio y la observación de escenarios altamente polarizados, se recogió la idea de incluir el concepto de radicalización en la medición de la polarización, afrontando la métrica de dicho fenómeno desde una perspectiva totalmente innovadora que, junto con el uso de la lógica difusa, ofrece valores coherentes con su conceptualización.

Así mismo, se han explorado diversas aplicaciones de la medición de la polarización en otros escenarios, como lo es el caso de problemas de detección de comunidades en grafos. Así, añadiendo como información extra a los algoritmos clásicos de detección de comunidades el cálculo de la polarización entre pares de nodos, junto con la aplicación de grafos difusos, se llevó el problema de detección de comunidades a un siguiente nivel. Desde esta nueva perspectiva, el nuevo algoritmo de detección de comunidades ofrecía un ajuste con la realidad no alcanzado por otros modelos. Así, por primera vez en la literatura, se abordan los problemas de detección de comunidades desde una perspectiva mixta, incluyendo en el modelo tanto la información relacional de la red como su ideología.

Así mismo, se extendió la medida de bipolarización difusa *JDJ* al caso de los grafos. En este sentido, el objetivo es medir la polarización, no sólo desde un punto de vista actitudinal o ideológico con la primera propuesta, sino de medir la polarización estructural desde una lógica difusa. Para ello, usando los grafos difusos como pilar, se obtienen los grados de pertenencia de

los nodos a sus respectivas comunidades, permitiendo la medición de la polarización en redes desde una perspectiva novedosa. En comparación al resto de medidas de polarización propuestas en la literatura, este nuevo enfoque permite medir la polarización con la consideración de sus grados de pertenencia a tales comunidades. Aplicaciones de diferentes medidas en distintos escenarios mostraron una mejor adecuación de la medida propuesta en este trabajo con respecto a las presentadas con anterioridad en la literatura.

Finalmente, se abordó la medición de la polarización desde una perspectiva totalmente nueva en la literatura a través de modelos estocásticos como lo son las Cadenas de Markov. Gracias al planteamiento de un modelo que pretende simular el concepto de polarización a través de diversos parámetros, se generó un modelo matemático estocástico capaz de predecir los niveles de polarización a lo largo del tiempo dada una determinada distribución de polarización en la población. Desde esta nueva perspectiva, se puede conocer el riesgo de polarización que presenta una determinada sociedad de acuerdo con su distribución actitudinal hacia un eje ideológico.

Con todo ello, en esta tesis se ha profundizado en la medición de la polarización desde diversas perspectivas novedosas, afrontando la medición del fenómeno siempre desde una perspectiva híbrida técnico-teórica con el fin de abordar la problemática desde una posición realista. La sinergia creada como consecuencia de la unión de ambas posiciones, tanto teóricas así como matemáticas, ha caracterizado el trabajo desarrollado por una propuesta novedosa en la literatura que, se espera, traiga mayores - y mejores - avances en el futuro.

Relación entre objetivos propuestos y contribuciones

En esta sección, se abordarán los objetivos propuestos inicialmente en la introducción de esta tesis 1.1 junto con su resolución y las contribuciones científicas producidas como consecuencia de la misma.

Objetivo 1: Estudiar y recolectar las principales características y medidas de polarización propuestas en la literatura para conocer sus elementos, naturaleza y crear una taxonomía en torno a ellas. Así mismo, se hará un estudio para conocer su rendimiento ante diferentes escenarios de polarización.

- **1.1.** Hacer una búsqueda y recopilación bibliográfica en torno a las características métricas de polarización.
- **1.2.** Crear una taxonomía en torno a las medidas de polarización

propuestas en la literatura de acuerdo con sus características, especialmente, en relación el tipo de datos que necesitan para ser computadas.

- **1.3.** Computar las medidas de polarización recogidas para su aplicación a diferentes escenarios de polarización y observar el rendimiento de las mismas frente a determinadas distribuciones poblacionales.

Con respecto a objetivo **1.1.** se llevó a cabo como parte del capítulo 2 sección 2.2. Por su parte, los objetivos **1.2.** y **1.3.** se abordaron a través del capítulo 3. Se considera este objetivo de especial importancia en tanto que supone las bases para comprender la medición de la polarización en la literatura y conseguir una visión amplia y rica que permita proponer nuevas medidas.

Objetivo 2: Analizar e identificar la polarización en el debate digital con el fin de detectar los orígenes, consecuencias y agentes de la polarización en redes.

- **2.1.** Estudiar el papel de los *bots* en el debate digital como agentes polarizantes y generadores de odio en la red.
- **2.2.** Analizar la polarización en torno a el debate digital en los casos de La Manada y #Cuéntalo.

Los objetivos **2.1.** y **2.2.** se han abordado a lo largo del capítulo 4 en sus dos secciones. La consecución de estos objetivos resulta de vital importancia para el entendimiento del comportamiento de la sociedad en determinados escenarios de presunta polarización. Así, los resultados obtenidos proporcionan una visión crucial a la hora de abordar el diseño de una medida de polarización. Las contribuciones asociadas a estos objetivos son las siguientes:

Contribución 1: Robles, J. M., Atienza, J., Gómez, D., & Guevara, J. A. (2019). La polarización de La Manada El debate público en España y los riesgos de la comunicación política digital. *Tempo Social*, 31(3), 193-216.

Contribución 4: Guevara, J. A., Atienza-Barthelemy, J., Gomez Gonzalez, D., & Robles, J. M. (2021). Polarization and incivility in digital debates on womens rights in Spain. Not just a matter of machismo. *Journal of Gender Studies*, 1-15.

Contribución 6: Martínez, A., Guevara, J.A., & Jiménez, A. (Aceptado, Enero 2023). La influencia de los bots políticos en Twitter durante la

crisis del COVID-19 en España. *REIS: Revista Española de Investigaciones Sociológicas*.

Objetivo 3: Diseñar una medida de polarización que supla algunas de las carencias encontradas en las ya propuestas en la literatura y que aborde los aspectos clave de la polarización detectados en los objetivos anteriores.

- **3.1.** Abordar la medición de la polarización a través de los conjuntos borrosos con el fin de proponer una nueva medida de polarización.
- **3.2.** Aplicar la medida de polarización a un caso real - el debate digital en redes sobre el COVID-19 - en entornos del *social big data analysis*.
- **3.3.** Explorar otras aplicaciones de la medida de polarización propuesta en la mejora de algoritmos de detección de comunidades en grafos.

Los objetivos **3.1.**, **3.2.** y **3.3.** han sido resueltos en el capítulo 5. Por su parte, el objetivo **3.1.** resulta uno de los pilares fundamentales de la presente tesis, construyendo, por primera vez en la literatura, una medida de polarización basada en los conjuntos borrosos. El objetivo **3.2.** resulta la culminación del diseño de una medida, siendo aplicada a un caso real y obteniendo valores de polarización basadas en las premisas novedosas propuestas en los objetivos anteriores. Finalmente, el objetivo **3.3.** permite explorar los usos de una medida de polarización más allá de la simple aplicación de la medida desde una perspectiva tradicional. De esta forma, se consiguió mejorar los algoritmos de detección de comunidades propuestos en la literatura, donde incluyendo los valores de la medida de polarización como información extra, se construyeron comunidades más realistas. Las publicaciones asociadas a estos objetivos son:

Contribución 2: Guevara, J.A., Gómez, D., Robles, J.M., Montero, J. (2020). Measuring Polarization: A Fuzzy Set Theoretical Approach. *Communications in Computer and Information Science*, vol 1238. Springer, Cham. https://doi.org/10.1007/978-3-030-50143-3_40

Contribución 3: Gutiérrez, I., Guevara, J. A., Gomez, D., Castro, J., & Espínola, R. (2021). Community Detection Problem Based on Polarization Measures: An Application to Twitter: The COVID-19 Case in Spain. *Mathematics*, 9(4), 443.

Contribución 5: Robles, J. M., Guevara, J. A., Casas-Mas, B., & Gómez, D. (2022). When negativity is the fuel. Bots and political polarization

in the COVID-19 debate. *Comunicar*, 30(71), 63-75.

Objetivo 4: Adaptar la medida de polarización propuesta al caso de grafos.

- **4.1.** Proponer una medida de polarización basada en los conjuntos borrosos aplicable a grafos.
- **4.2.** Comparar la medida de polarización propuesta con las medidas de polarización más usadas en grafos.

La traslación de la medición de la polarización al caso de los grafos, permite medir tal fenómeno, no sólo desde una perspectiva actitudinal, o ideológica, sino también estructural. Los objetivos **4.1.** y **4.2.** abordan esta lógica, siendo planteados a lo largo del capítulo 6. La contribución científica asociada a éstos objetivos es la siguiente:

Contribución 7: Simón de Blas, C., Guevara, J.A., Morillo, J., Gómez González, D. (2022). Polarization Measures in Bi-partition Networks Based on Fuzzy Graphs. *Communications in Computer and Information Science*, vol 1601. Springer, Cham.

Objetivo 5: Explorar y proponer un modelo que permita abordar la polarización desde una perspectiva dinámica.

- **5.1.** Modelar el fenómeno de polarización desde una perspectiva dinámica a través de Cadenas de Markov.
- **5.2.** Ejecutar el modelo propuesto anteriormente con un *toy example* para estudiar su rendimiento.
- **5.3.** Comparar los resultados del modelo de polarización dinámico con medidas de polarización.

Finalmente, el planteamiento de los objetivos **5.1.**, **5.2.** y **5.3.**, abordados en el capítulo 7, han permitido abordar la medición de la polarización desde una perspectiva totalmente novedosa en la literatura. Así, se creó un modelo estocástico que permite predecir los niveles de polarización a lo largo del tiempo, contemplando dicho fenómeno como un proceso dinámico en el tiempo. La publicación científica asociada a estos objetivos es la siguiente:

Contribución 8: Guevara, J.A., Gómez, D., Castro, J., Gutiérrez, I., Robles, J.M. (2022). A New Approach to Polarization Modeling Using

Markov Chains. Communications in Computer and Information Science, vol 1602. Springer, Cham.

Líneas de investigación futuras

A lo largo de esta tesis, y como consecuencia de la consecución de sus objetivos, se han propuesto novedosos avances que abordan algunos de los aspectos más importantes de la medición de la polarización, tales como la propuesta de una medida de polarización con una conceptualización más realista y con un comportamiento en mayor sincronía a las propuestas conceptuales, la traslación de la misma al caso de los grafos, o incluso la medición de la polarización desde una perspectiva dinámica. Con ello, se han abierto un conjunto de líneas de investigación de gran interés para la medición de la polarización.

Con respecto a la medida JDJ , se sugieren algunas direcciones para investigaciones futuras. En relación con las funciones de pertenencia, se considera de vital importancia profundizar en el estudio de las mismas. Las funciones de pertenencia utilizadas para medir la radicalización de los individuos ha resultado jugar un papel fundamental en los valores que ofrece la medida. Así, la contemplación de diferentes funciones de pertenencia, así como el intercambio entre unas y otras de acuerdo con determinados escenarios, representa un objeto de estudio de gran interés. Con respecto a la medición en sí misma, se ha abordado desde una perspectiva bipolar - entendiendo un único eje con dos polos extremos -, por lo que abordar la medición desde una conceptualización multidimensional permitirá medir la polarización desde posiciones novedosas en el ámbito de la lógica difusa. Así mismo, esto ampliará la cantidad de variables que se incluyen en la medición de la polarización.

Así mismo, las consideraciones anteriores pueden ser trasladadas a la medida JDJ aplicada en grafos - $JDJ_{pol,gre}(V, E, P, G_O, G_G)$ -, permitiendo explorar nuevas formas de comprender la membresía de los nodos a sus comunidades, así como abordar la medición de la polarización con la consideración de múltiples comunidades en el grafo. En este sentido, las posibles líneas de investigación se dividen en dos. En primer lugar, ordenar categóricamente las diferentes comunidades detectadas en la red de acuerdo con un eje ideológico, con el fin de conocer los dos grupos más extremos - polos - y los valores intermedios. Sin embargo, esta consideración abordaría la medición de la polarización en grafos desde una perspectiva bipolar. Por otro lado, ampliar la medida a una dimensión multipolar permitiría medir la polarización en grafos con > 2 comunidades desde una nueva posición.

Tanto para el caso de la medida JDJ como para $JDJ_{pol,gre}(V, E, P, G_O, G_G)$, es de vital importancia considerar el riesgo de confundir la medición de polarización con la de fragmentación o segregación. De hecho, esta distinción no parece estar clara en la literatura, considerando fundamental profundizar en el estudio de estos términos para definir las fronteras entre ellos para, finalmente, proponer medidas de polarización multipolares desde una perspectiva lógica y coherente.

Finalmente, con respecto a la medición de la polarización desde una perspectiva dinámica, se abordó a través de las Cadenas de Markov, siendo uno de los modelos matemáticos más simples para contemplar fenómenos en el tiempo. La consideración de otros modelos de series temporales más sofisticados permitirá la posibilidad de afrontar esta tardea de forma más compleja. Así mismo, los parámetros que se consideran en el modelo presentado en el capítulo 7 son una primera aproximación a la modelización del fenómeno de la polarización. Estudiar en mayor profundidad aquellos aspectos que influyen en el comportamiento de dicho fenómeno permitirá crear modelos mucho más ajustados a la realidad.

Bibliografía

*El que lee mucho y anda mucho, ve
mucho y sabe mucho.*

Miguel de Cervantes Saavedra

- [1] A. I. Abramowitz. The disappearing center. In *The Disappearing Center*. Yale University Press, 2010.
- [2] A. Al-Rawi and V. Shukla. Bots as active news promoters: A digital analysis of covid-19 tweets. *Information*, 11(10):461, 2020.
- [3] A. d. Amo, J. Montero, G. Biging, and V. Cutello. Fuzzy classification systems. *European Journal of Operational Research*, 156(2):495–507, 2004.
- [4] B. Apouey. Measuring health polarization with selfassessed health data. *Health Econ.*, page 20, 2007.
- [5] D. Baldassarri and P. Bearman. Dynamics of political polarization. *American sociological review*, 72(5):784–811, 2007.
- [6] P. C. Bauer. Conceptualizing and measuring polarization: A review. 2019.
- [7] R. F. Baumeister. *The self*. Oxford university press, 2010.
- [8] R. Belohlavek, G. J. Klir, H. W. Lewis III, and E. C. Way. Concepts and fuzzy sets: Misunderstandings, misconceptions, and oversights. *International journal of approximate reasoning*, 51(1):23–34, 2009.
- [9] K. J. Berry and P. W. Mielke Jr. Indices of ordinal variation. *Perceptual and motor skills*, 74(2):576–578, 1992.

-
- [10] A. Biswas and B. Biswas. Fuzag: Fuzzy agglomerative community detection by exploring the notion of self-membership. *IEEE Transactions on Fuzzy Systems*, 26(5):2568–2577, 2018.
- [11] J. Blair and M. Lacy. From the sage social science collections. rights reserved. *Sociological Methods & Research*, 28(3):251–280, 2000.
- [12] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.
- [13] J. Borondo, A. J. Morales, J.-C. Losada, and R. M. Benito. Characterizing and modeling an electoral campaign in the context of twitter: 2011 spanish presidential election as a case study. *Chaos: an interdisciplinary journal of nonlinear science*, 22(2):023138, 2012.
- [14] L. Boxell, M. Gentzkow, and J. M. Shapiro. Is the internet causing political polarization. *Evidence From Demographics. On the web at: <https://www.brown.edu/Research/Shapiro/pdfs/age-polars.pdf>*, 2017.
- [15] S. Bradshaw and P. N. Howard. Online supplement to working paper 2018.1 challenging truth and trust: A global inventory of organized social media manipulation, 2018.
- [16] S. Bradshaw and P. N. Howard. The global disinformation order: 2019 global inventory of organised social media manipulation. 2019.
- [17] A. Bramson, P. Grim, D. J. Singer, S. Fisher, W. Berger, G. Sack, and C. Flocken. Disambiguation of social polarization concepts and measures. *The Journal of Mathematical Sociology*, 40(2):80–111, 2016.
- [18] H. Bustince, J. Fernández, R. Mesiar, J. Montero, and R. Orduna. Overlap index, overlap functions and migrativity. 2009.
- [19] G. Caldarelli, A. Chessa, F. Pammolli, G. Pompa, M. Puliga, M. Riccaboni, and G. Riotta. A multi-level geographical study of italian political elections from twitter data. *PloS one*, 9(5):e95809, 2014.
- [20] J. P. Cárdenas, G. Vidal, C. Urbina, G. Olivares, P. Rodrigo, and M. Fuentes. Social crises: signatures of complexity in a fast-growing economy. *Complexity*, 2018, 2018.
- [21] R. J. Dalton. The quantity and the quality of party systems: Party system polarization, its measurement, and its consequences. *Comparative Political Studies*, 41(7):899920, 2008.

- [22] C. S. de Blas, J. S. Martin, and D. G. Gonzalez. Combined social networks and data envelopment analysis for ranking. *European Journal of Operational Research*, 266(3):990–999, 2018.
- [23] M. Deutch. *Conflict and its Resolution in Conflict resolution: Contributions of the behavioral sciences*. ed. by C. G. Smith. U. Notre Dame Press, 1971.
- [24] M. Devarajan, N. S. Fatima, S. Vairavasundaram, and L. Ravi. Swarm intelligence clustering ensemble based point of interest recommendation for social cyber-physical systems. *Journal of Intelligent & Fuzzy Systems*, 36(5):4349–4360, 2019.
- [25] P. DiMaggio, J. Evans, and B. Bryson. Have american’s social attitudes become more polarized? *American journal of Sociology*, 102(3):690–755, 1996.
- [26] S. Duval and R. A. Wicklund. *A theory of objective self awareness*. Academic press, 1972.
- [27] P. Erdős, A. Rényi, et al. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.
- [28] J.-M. Esteban and D. Ray. On the measurement of polarization. *Econometrica: Journal of the Econometric Society*, pages 819–851, 1994.
- [29] S. Fortunato. Community detection in graphs. *Physics reports*, 486(3-5):75–174, 2010.
- [30] J. E. Foster and M. C. Wolfson. Polarization and the decline of the middle class: Canada and the us mimeo. *Vanderbilt University*, 31, 1992.
- [31] F. Franceschini, M. Galetto, and M. Varetto. Qualitative ordinal scales: the concept of ordinal range. *Quality Engineering*, 16(4):515–524, 2004.
- [32] L. C. Freeman. Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239, 1978.
- [33] L. C. Freeman. Spheres, cubes and boxes: graph dimensionality and network structure. *Social Networks*, 5(2):139–156, 1983.
- [34] L. C. Freeman, S. P. Borgatti, and D. R. White. Centrality in valued graphs: A measure of betweenness based on network flow. *Social networks*, 13(2):141–154, 1991.

-
- [35] L. C. Freeman, D. Roeder, and R. R. Mulholland. Centrality in social networks: Ii. experimental results. *Social networks*, 2(2):119–141, 1979.
- [36] T. Gadrich and E. Bashkansky. Ordanova: analysis of ordinal variation. *Journal of Statistical Planning and Inference*, 142(12):3174–3188, 2012.
- [37] K. Garimella, G. D. F. Morales, A. Gionis, and M. Mathioudakis. Quantifying controversy on social media. *ACM Transactions on Social Computing*, 1(1):1–27, 2018.
- [38] M. Gentzkow. Polarization in 2016. *Toulouse Network for Information Technology Whitepaper*, pages 1–23, 2016.
- [39] B. T. Gervais. Incivility online: Affective and behavioral reactions to uncivil political posts in a web-based experiment. *Journal of Information Technology & Politics*, 12(2):167–185, 2015.
- [40] C. Gini. Variabilità e mutabilità. *Reprinted in Memorie di metodologica statistica (Ed. Pizetti E, 1912.*
- [41] M. Girvan and M. E. Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.
- [42] D. Gómez, E. González-Arangüena, C. Manuel, G. Owen, M. Del Pozo, and M. Saboyá. The cohesiveness of subgroups in social networks: A view from game theory. *Annals of Operations Research*, 158(1):33–46, 2008.
- [43] D. Gomez, E. González-Arangüena, C. Manuel, G. Owen, M. del Pozo, and J. Tejada. Centrality and power in social networks: a game theoretic approach. *Mathematical Social Sciences*, 46(1):27–54, 2003.
- [44] D. Gómez, J. T. Rodríguez, J. Montero, H. Bustince, and E. Barrenechea. n-dimensional overlap functions. *Fuzzy Sets and Systems*, 287:57–75, 2016.
- [45] D. Gomez, J. T. Rodríguez, J. Yanez, and J. Montero. A new modularity measure for fuzzy community detection problems based on overlap and grouping functions. *International Journal of Approximate Reasoning*, 74:88–107, 2016.
- [46] D. Gómez, E. Zarrazola, J. Yáñez, and J. Montero. A divide-and-link algorithm for hierarchical clustering in networks. *Information Sciences*, 316:308–328, 2015.

-
- [47] M. Grabisch, H. T. Nguyen, and E. A. Walker. Fundamentals of uncertainty calculi with applications to fuzzy inference, 1998.
- [48] S. Gregory. Fuzzy overlapping communities in networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(02):P02017, 2011.
- [49] P. Guerra, W. Meira Jr, C. Cardie, and R. Kleinberg. A measure of polarization on social media networks based on community boundaries. In *Proceedings of the international AAAI conference on web and social media*, volume 7, pages 215–224, 2013.
- [50] J. A. Guevara, D. Gómez, J. M. Robles, and J. Montero. Measuring polarization: A fuzzy set theoretical approach. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 510–522. Springer, 2020.
- [51] I. Gutiérrez, D. Gómez, J. Castro, and R. Espínola. A new community detection algorithm based on fuzzy measures. In *International Conference on Intelligent and Fuzzy Systems*, pages 133–140. Springer, 2019.
- [52] I. Gutierrez, D. Gomez, J. Castro, and R. Espinola. Fuzzy measures: A solution to deal with community detection problems for networks with additional information. *Journal of Intelligent & Fuzzy Systems*, 39(5):6217–6230, 2020.
- [53] I. Gutiérrez, J. A. Guevara, D. Gomez, J. Castro, and R. Espínola. Community detection problem based on polarization measures: An application to twitter: The covid-19 case in spain. *Mathematics*, 9(4):443, 2021.
- [54] B. N. Hague and B. D. Loader. *Digital democracy: Discourse and decision making in the information age*. Routledge, 2005.
- [55] J. Haidt and M. J. Hetherington. Look how far weve come apart. *The New York Times*, 2012.
- [56] L. K. Hansen, A. Arvidsson, F. Å. Nielsen, E. Colleoni, and M. Etter. Good friends, bad news-affect and virality in twitter. In *Future information technology*, pages 34–43. Springer, 2011.
- [57] S. Herbst. *Rude democracy: Civility and incivility in American politics*. Temple University Press, 2010.

- [58] K. Hornik and B. Grün. topicmodels: An r package for fitting topic models. *Journal of statistical software*, 40(13):1–30, 2011.
- [59] P. N. Howard, S. Savage, C. F. Saviaga, C. Toxtli, and A. Monroy-Hernández. Social media, civic engagement, and the slacktivism hypothesis: Lessons from mexicos el bronco. *Journal of International Affairs*, 70(1):55–73, 2016.
- [60] T. Hwang and S. Woolley. How politicians should or shouldnt use twitter bots. *Slate.com*, 2016.
- [61] M. O. Jackson. *Social and economic networks*. Princeton university press, 2010.
- [62] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian. Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PLoS one*, 9(6):e98679, 2014.
- [63] M. W. Kearney. rtweet: Collecting and analyzing twitter data. *Journal of open source software*, 4(42):1829, 2019.
- [64] M. W. Kearney. *tweetbotornot2: Detect Twitter Bots*, 2020. R package version 0.0.1.
- [65] T. R. Keller and U. Klinger. Social bots in election campaigns: Theoretical, empirical, and methodological implications. *Political Communication*, 36(1):171–189, 2019.
- [66] G. Klir and B. Yuan. *Fuzzy sets and fuzzy logic*, volume 4. Prentice hall New Jersey, 1995.
- [67] M. Kuhn. *caret: Classification and Regression Training*, 2020. R package version 6.0-86.
- [68] P. F. Lazarsfeld, R. K. Merton, et al. Friendship as a social process: A substantive and methodological analysis. *Freedom and control in modern society*, 18(1):18–66, 1954.
- [69] Y. Lelkes. Mass polarization: Manifestations and measurements. *Public Opinion Quarterly*, 80(S1):392–410, 2016.
- [70] J. C. Losada, J. M. Robles, R. M. Benito, and R. Caballero. Love and hate during political campaigns in social networks. In *International Conference on Complex Networks and Their Applications*, pages 66–77. Springer, 2021.

- [71] C. Lutz, C. P. Hoffmann, and M. Meckel. Beyond just politics: A systematic literature review of online participation. *First Monday*, 19(7):1–36, 2014.
- [72] S. Martín-Gutiérrez, J. C. Losada, and R. M. Benito. Recurrent patterns of user behavior in different electoral campaigns: a twitter analysis of the spanish general elections of 2015 and 2016. *Complexity*, 2018, 2018.
- [73] N. Martínez, D. Gómez, P. Olaso, K. Rojas, and J. Montero. A novel ordered weighted averaging weight determination based on ordinal dispersion. *International Journal of Intelligent Systems*, 34(9):2291–2315, 2019.
- [74] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch. *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*, 2020. R package version 1.7-4.
- [75] J. Moffitt, C. King, and K. M. Carley. Hunting conspiracy theories during the covid-19 pandemic. *Social Media+ Society*, 7(3):20563051211043212, 2021.
- [76] J. G. Montalvo and M. Reynal-Querol. Religious polarization and economic development. *Economics Letters*, 80(2):201–210, 2003.
- [77] J. Montero, D. Gómez, V. López, S. Rodríguez, and B. Vitoriano. Sobre las funciones y reglas de agregación. 2010.
- [78] A. J. Morales, J. Borondo, J. C. Losada, and R. M. Benito. Measuring political polarization: Twitter shows the two sides of venezuela. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(3):033114, 2015.
- [79] J. N. Mordeson and P. S. Nair. *Fuzzy graphs and fuzzy hypergraphs*, volume 46. Physica, 2012.
- [80] P. S. Nair and S. T. Sarasamma. Data mining through fuzzy social network analysis. In *NAFIPS 2007-2007 Annual Meeting of the North American Fuzzy Information Processing Society*, pages 251–255. IEEE, 2007.
- [81] M. E. Newman. Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582, 2006.

- [82] M. E. Newman and D. J. Watts. Renormalization group analysis of the small-world network model. *Physics Letters A*, 263(4-6):341–346, 1999.
- [83] T. A. Neyazi. Digital propaganda, political bots and polarized politics in india. *Asian Journal of Communication*, 30(1):39–57, 2020.
- [84] A. E. Nix and M. D. Vose. Modeling genetic algorithms with markov chains. *Annals of mathematics and artificial intelligence*, 5(1):79–88, 1992.
- [85] J. R. Norris. *Markov chains*. Number 2. Cambridge university press, 1998.
- [86] A. Osorio-Lird, A. Chamorro, C. Videla, S. Tighe, and C. Torres-Machi. Application of markov chains and monte carlo simulations for developing pavement performance models for urban network management. *Structure and Infrastructure Engineering*, 14(9):1169–1181, 2018.
- [87] Z. Papacharissi. Democracy online: Civility, politeness, and the democratic potential of online political discussion groups. *New media & society*, 6(2):259–283, 2004.
- [88] J. Pastor-Galindo, M. Zago, P. Nespoli, S. L. Bernal, A. H. Celdrán, M. G. Pérez, J. A. Ruipérez-Valiente, G. M. Pérez, and F. G. Mármol. Spotting political social bots in twitter: A use case of the 2019 spanish general election. *IEEE Transactions on Network and Service Management*, 17(4):2156–2170, 2020.
- [89] I. Permanyer and C. DÁMBROSIO. Measuring social polarization with ordinal and categorical data. *Journal of Public Economic Theory*, 17(3):311–327, 2015.
- [90] P. E. Pfeifer and R. L. Carraway. Modeling customer relationships as markov chains. *Journal of interactive marketing*, 14(2):43–55, 2000.
- [91] M. Prior. Media and political polarization. *Annual Review of Political Science*, 16:101–127, 2013.
- [92] M. Reynal-Querol. *Ethnic and religious conflicts, political systems and growth*. Ph.D. thesis, London School of Economics and Political Science. University of London, 2001.

- [93] J. M. Robles, J. Atienza, D. Gómez, and J. A. Guevara. La polarización de la manada el debate público en españa y los riesgos de la comunicación política digital. *Tempo Social*, 31(3):193–216, 2019.
- [94] J.-M. Robles, J.-A. Guevara, B. Casas-Mas, and D. Gómez. Cuando la negatividad es el combustible. bots y polarización política en el debate sobre el covid-19. *Comunicar: Revista Científica de Comunicación y Educación*, 30(71), 2022.
- [95] J. M. Robles, J. T. Rodríguez, R. Caballero, and D. Gómez. Big data para científicos sociales. una introducción. *Revista Española de Investigaciones Sociológicas (REIS)*, 176(176):157–160, 2021.
- [96] J. M. Robles, D. Velez, S. De Marco, J. T. Rodríguez, and D. Gomez. Affective homogeneity in the spanish general election debate. a comparative analysis of social networks political agents. *Information, Communication & Society*, 23(2):216–233, 2020.
- [97] K. Rojas, D. Gómez, and J. Montero. Algunas propiedades de consistencia de las familias de operadores de agregación. 2012.
- [98] A. Rosenfeld. Fuzzy graphs. In *Fuzzy sets and their applications to cognitive and decision processes*, pages 77–95. Elsevier, 1975.
- [99] E. A. Ross. *The principles of sociology*. Century Company, 1921.
- [100] M. Roubens. Pattern classification problems and fuzzy sets. *Fuzzy sets and systems*, 1(4):239–253, 1978.
- [101] L. E. Santana and G. H. Cánepa. ¿son bots? automatización en redes sociales durante las elecciones presidenciales de chile 2017. *Cuadernos. info*, (44):61–77, 2019.
- [102] G. Sartori. *Parties and party systems: A framework for analysis*. ECPR press, 2005.
- [103] J. Schradie. The digital production gap: The digital divide and web 2.0 collide. *Poetics*, 39(2):145–168, 2011.
- [104] S.Gregory. Fuzzy overlapping communities in networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2:PO2017, 2011.
- [105] L. S. Shapley. A value for n-person games, contributions to the theory of games, 2, 307–317, 1953.

- [106] M. Sherif, O. Harvey, B. J. White, W. R. Hood, and C. W. Sherif. 1961 intergroup conflict and cooperation: The robbers cave experiment. *Norman, OK: University of Oklahoma*, 1954.
- [107] G. A. Spedicato. Discrete time markov chains with r. *The R Journal*, 07 2017. R package version 0.6.9.7.
- [108] M. Stella, E. Ferrara, and M. De Domenico. Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences*, 115(49):12435–12440, 2018.
- [109] M. Sugeno. Fuzzy measures and fuzzy integrals a survey. In *Readings in fuzzy sets for intelligent systems*, pages 251–257. Elsevier, 1993.
- [110] C. R. Sunstein. *Designing democracy: What constitutions do*. Oxford University Press, 2001.
- [111] C. R. Sunstein. *# Republic*. Princeton university press, 2018.
- [112] J. J. Sylvester. Chemistry and algebra. *Nature*, 17(432):284, 1878.
- [113] H. Tajfel. Social identity and intergroup behaviour. *Social science information*, 13(2):65–93, 1974.
- [114] J. C. Turner, I. Sachdev, and M. A. Hogg. Social categorization, interpersonal attraction and group formation. *British Journal of Social Psychology*, 22(3):227–239, 1983.
- [115] S. van Haperen, W. Nicholls, and J. Uitermark. Building protest online: engagement with the digitally networked# not1more protest campaign on twitter. *Social Movement Studies*, 17(4):408–423, 2018.
- [116] J. M. Verd, C. Lozares, I. C. Gómez, and O. Barranco. La homofilia/heterofilia en el marco de la teoría y análisis de redes sociales. orientación metodológica, medición y aplicaciones. *Metodología de encuestas*, 16(1):5–25, 2014.
- [117] M. Wagner. Affective polarization in multiparty systems. *Electoral Studies*, 69:102199, 2021.
- [118] Y.-Q. Wang and K.-Y. Tsui. Polarization orderings and new classes of polarization indices. *Journal of Public Economic Theory*, 2(3), 2000.

-
- [119] A. S. Waugh, L. Pei, J. H. Fowler, P. J. Mucha, and M. A. Porter. Party polarization in congress: A social networks approach. *arXiv preprint arXiv:0907.3509*, 3(4):69, 2009.
- [120] M. C. Wolfson. When inequalities diverge. *The American Economic Review*, 84(2):353–358, 1994.
- [121] H. Y. Yan, K.-C. Yang, F. Menczer, and J. Shanahan. Asymmetrical perceptions of partisan political bots. *New Media & Society*, 23(10):3016–3037, 2021.
- [122] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.