Grupo de Sistemas Complejos

Escuela Técnica Superior de Ingeniería
Agronómica, Alimentaria y de Biosistemas

**MODELS OF SOCIAL INTERACTION AND DATA ANALYSIS OF
ONLINE USER BEHAVIOR DURING POLARIZATION PROCESSES**

Tesis doctoral

SAMUEL MARTÍN GUTIÉRREZ
Máster en Física de Sistemas Complejos
Licenciado en Física

DIRIGIDA POR:
ROSA MARÍA BENITO ZAFRILLA
Doctora en Ciencias Químicas

2020

ABSTRACT

The main goal of this thesis is to improve our understanding of collective human behavior both at the level of the elementary social interactions and at the level of the emergent properties that arise from these interactions, placing a special emphasis on the study of social polarization, which seems to be pervasive in the current zeitgeist. To this end, we analyze large quantities of data with both existing and new techniques and develop mathematical models and simulations.

We have developed three models of social interaction that relate individual activity (the number of actions of an individual - A) with the collective response of social systems (the number of reactions triggered on the other members of the system - R). That relationship is characterized by means of the probability distribution of the efficiency metric, defined as $\eta = R/A$. Generalizing previous results, we show that the efficiency distribution presents a universal structure in three systems of different nature: Twitter, Wikipedia and the scientific citations network. The models explain that universality and provide a description of the underlying social dynamics of each system.

Once the elementary social dynamics have been characterized, we address the study of polarization, a process of social division that results in the majority of the population holding extreme and opposing opinions with few individuals remaining neutral. Empirical data is at the core of this work and Twitter is our main data source, so we begin with a user behavior study of electoral contexts, since these are some of the most common scenarios where polarization may emerge. The objective is to find regularities in the behavioral and communication patterns among the users by analyzing the evolution of their activity and interaction networks. This study contributes to verify the stability of interaction dynamics and establish baselines that may allow the detection of fluctuations and irregularities, which could indicate changes in the social landscape.

In order to estimate the opinions of the users with Sentiment Analysis techniques, we develop a methodology to semi-automatically build a training set of tweets in a polarized context. First, we compile lists of Twitter users with known affiliation to one of the poles of a given scenario (a political party in an electoral context, one of the teams in a football match, a certain brand in a commercial competition, etc.). Then, every tweet sent by a user associated to a given pole is labeled as positive if it mentions users of the same pole and as negative if it mentions users from different poles. The methodology is tested by comparing the automatically built training sets with manually labeled reference datasets.

On the other hand, we have explored an alternative paradigm of opinion inference that is based not on the content of the messages but on the network of social links woven by the interactions. We have also taken into consideration that the behavior of many social contexts can not be fully understood by simplifying them as bipolar systems (for example, multi-party elections), and they must be treated as multipolar systems. Therefore, we have developed a general opinion inference technique that is not only valid for bipolar systems, but for systems with an arbitrary number of opinion poles. Additionally, we have developed methodological tools to measure and characterize the polarization patterns embedded in the multipolar opinion

distributions. This methodology has been applied to five real-world scenarios with two, three, four and five poles, finding clear connections between the opinion distributions and the underlying sociological contexts.

The multidimensional nature of opinion is also manifested in the fact that attitudes are often interlinked with other social dimensions. For example, in territorial conflicts language tends to play a critical role and is used in different ways by the different poles. We have studied this phenomenon in a Twitter conversation around the independence of the Spanish region of Catalonia. In particular, we have analyzed the relationship between ideology and language by combining an opinion index computed using the methodology described above with a language index calculated by analyzing the relative use of Catalan and Spanish in the tweets.

Finally, we present two network metrics that quantify the polarization of the topology of a social network, although they have potential applications in many other fields: the Network Variance and the Network Covariance. These measures generalize the usual notion of variance and covariance, elementary statistical tools applied in Euclidean spaces, to arbitrary metric spaces. In particular, to networks, because a metric space can be built from a set of nodes using an appropriate distance measure between them. We illustrate the usefulness of the Network (Co)variance by characterizing the relationship between two networks of mathematical knowledge: the *functional* network (that encodes how are mathematical ideas used in scientific papers) and the *structural* network (that encodes how are ideas conceptually related).

RESUMEN

El objetivo último de esta tesis es ampliar el conocimiento del comportamiento humano colectivo tanto a nivel de las interacciones sociales elementales, como a nivel de las propiedades emergentes que surgen de estas interacciones, poniendo especial énfasis en la polarización social, que parece omnipresente en el contexto actual. Con este fin, analizamos grandes cantidades de datos mediante técnicas tanto existentes como originales y desarrollamos modelos matemáticos y simulaciones.

Hemos desarrollado tres modelos de interacción social que relacionan la actividad individual (cantidad de acciones de un individuo - A) con la respuesta colectiva de un sistema social (cantidad de reacciones provocadas en otros miembros del sistema - R). Esta relación se caracteriza mediante la distribución de probabilidad de la eficiencia, una métrica definida como $\eta = R/A$. Generalizando resultados anteriores demostramos que la distribución de eficiencia presenta una estructura universal en tres sistemas de distinta naturaleza: Twitter, Wikipedia y la red de citas científicas. Los modelos nos permiten explicar dicha universalidad y, además, proporcionan una descripción de las dinámicas sociales subyacentes de cada sistema.

Una vez caracterizadas las dinámicas sociales elementales, abordamos el estudio de la polarización, un proceso de división social que provoca que la mayoría de la población adopte opiniones extremas y opuestas con pocos individuos en una posición neutral. Comenzamos con un estudio del comportamiento de los usuarios de Twitter en contextos electorales, puesto que la polarización surge con frecuencia en estos escenarios. El objetivo es encontrar regularidades en los patrones de comportamiento y comunicación de los usuarios mediante el análisis de la evolución temporal de su actividad así como de sus redes de interacciones. Este estudio contribuye a verificar la estabilidad de las dinámicas de interacción y establecer referencias que podrían permitir la detección de fluctuaciones e irregularidades de comportamiento y, a su vez, variaciones en el panorama social.

Con el fin de estimar las opiniones de los usuarios con técnicas de Análisis de Sentimiento, desarrollamos una metodología para construir conjuntos de datos de entrenamiento de manera semi-automática en contextos polarizados. En primer lugar, elaboramos listas de usuarios de Twitter cuya afiliación a alguno de los polos es conocida (los polos serían partidos políticos en un contexto electoral o los equipos en un partido de fútbol). Los tweets enviados por un usuario asociado a un polo se etiquetan como positivos si mencionan usuarios del mismo polo y como negativos si mencionan usuarios de otros polos. Hemos evaluado esta metodología comparando los conjuntos de entrenamiento contruidos de forma automática con corpus de referencia etiquetados manualmente.

Por otro lado, hemos explorado un paradigma alternativo de inferencia de opinión que no se basa en el contenido de los mensajes intercambiados entre los usuarios sino en la red de contactos sociales entretejida por medio de estas interacciones. También hemos tenido en consideración que el comportamiento de muchos contextos sociales no se puede comprender reduciéndolos a sistemas bipolares (por ejemplo, elecciones con más de dos partidos) y es necesario tratarlos como sistemas multipolares. Por lo tanto hemos desarrollado una técnica general de inferencia de opinión válida no solo para sistemas bipolares, sino para sistemas con un número

arbitrario de polos. Esta metodología se ha aplicado a cinco escenarios reales con dos, tres, cuatro y cinco polos, encontrando conexiones claras entre las distribuciones de opinión y el contexto sociológico subyacente.

El carácter multidimensional de la opinión también se manifiesta en el hecho de que las opiniones en muchos casos están fuertemente conectadas con otras dimensiones sociales. Por ejemplo, en conflictos territoriales el idioma tiende a jugar un papel crítico, y cada polo suele presentar un uso del idioma bien diferenciado. Hemos estudiado este fenómeno en una conversación de Twitter sobre la independencia de Cataluña. Para ello, hemos analizado la relación entre ideología y uso del idioma combinando un índice de opinión calculado mediante la metodología descrita anteriormente con un índice de lenguaje que recoge el uso relativo del catalán y el español en los tweets.

Por último, presentamos dos métricas que permiten cuantificar la polarización de la topología de una red social, aunque su aplicabilidad abarca muchos más ámbitos de la Ciencia de Redes: la Varianza de Red y la Covarianza de Red. Estas medidas generalizan las nociones de varianza y covarianza, herramientas estadísticas elementales en espacios euclídeos, a espacios métricos arbitrarios. En particular, a redes, puesto que se puede definir un espacio métrico sobre los nodos de una red utilizando una medida apropiada de distancia entre los nodos. Ilustramos la utilidad de la (Co)varianza de Red caracterizando la relación entre dos redes que estructuran el conocimiento matemático: la red *funcional* (que codifica el uso de conceptos matemáticos en artículos científicos) y la red *estructural* (que codifica las relaciones conceptuales).

PUBLICATIONS

- [AB+19] J. Atienza-Barthelemy, S. Martin-Gutierrez, J. C. Losada, and R. M. Benito. “Relationship between ideology and language in the Catalan independence context.” In: *Scientific Reports* 9.1 (2019), p. 17148. ISSN: 2045-2322. DOI: [10.1038/s41598-019-53404-x](https://doi.org/10.1038/s41598-019-53404-x).
- [DMGL20] K. Devriendt, S. Martin-Gutierrez, and R. Lambiotte. *Variance and covariance of distributions on graphs*. Submitted to SIAM Review. 2020. arXiv: [2008.09155](https://arxiv.org/abs/2008.09155) [[physics.soc-ph](#)].
- [MGLB18] S. Martin-Gutierrez, 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.” In: *Complexity* 2018 (2018), p. 2413481. ISSN: 1076-2787. DOI: [10.1155/2018/2413481](https://doi.org/10.1155/2018/2413481).
- [MLB18] S. Martin-Gutierrez, J. C. Losada, and R. M. Benito. “Semi-automatic training set construction for supervised sentiment analysis in political contexts.” In: *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. 2018, pp. 715–720. DOI: [10.1109/ASONAM.2018.8508386](https://doi.org/10.1109/ASONAM.2018.8508386).
- [MGLB20a] S. Martin-Gutierrez, J. C. Losada, and R. M. Benito. “Semi-automatic training set construction for supervised sentiment analysis in polarized contexts.” In: *Putting Social Media and Networking Data in Practice for Education, Planning, Prediction and Recommendation*. Ed. by M. Kaya, Ş. Birinci, J. Kawash, and R. Alhaji. Cham: Springer International Publishing, 2020, pp. 177–197. ISBN: 978-3-030-33698-1. DOI: [10.1007/978-3-030-33698-1_10](https://doi.org/10.1007/978-3-030-33698-1_10).
- [MGLB20b] S. Martin-Gutierrez, J. C. Losada, and R. M. Benito. “Impact of individual actions on the collective response of social systems.” In: *Scientific Reports* 10.1 (2020), p. 12126. ISSN: 2045-2322. DOI: [10.1038/s41598-020-69005-y](https://doi.org/10.1038/s41598-020-69005-y).

*No os he mostrado a la mitad de vosotros
ni la mitad de gratitud de lo que querría,
y lo que yo querría es menos de la mitad
de lo que la mitad de vosotros merece.*

— J. R. R. Tolkien - *El Señor de los Anillos: La Comunidad del Anillo*
(versionado libremente)

AGRADECIMIENTOS

Hay muchas personas a las que me gustaría mostrar gratitud por lo que han aportado a esta tesis, y más aún a las que agradecer lo que me han aportado a mi, como persona y como científico.

Agradezco en primer lugar a mi directora, Rosa, su paciencia y apoyo, incluso en situaciones de sobrecarga de trabajo en las que no es sencillo compatibilizar responsabilidades. También a Juan Carlos Losada, sin cuyas innumerables contribuciones esta tesis no sería lo que es. A Fabio, científico ejemplar y buen amigo. A Alfredo Morales, que me guio en los primeros pasos del viaje y fue el granito de arena que desencadenó una avalancha de proyectos apasionantes. En general, a los profesores y profesoras del Grupo de Sistemas Complejos: Ramón Alonso, Miguel Ángel Porras, Ana Tarquis, Javier Arranz, Javier Galeano y Juan Manuel Pastor. Y, por supuesto, a todos los compañeros y compañeras del doctorado: Pedro, Julia, Leti, Benja, Marta, Javi(s), Raúl, Iván, José Luis, Gastón, Alfonso, Luciano... Y muchos más que han pasado por el grupo dejando su huella.

Gracias a Renaud Lambiotte, Karel, Rodrigo y Yu, por recibirme en Oxford y por una estancia de investigación fascinante.

Mencionar también a mis compañeros y amigos de la carrera de Física, difícilmente se podría encontrar mejor compañía para empezar una aventura científica.

Gracias mis padres y mi hermana, por su apoyo incondicional y confianza, sin injerencias ni presiones; cosas que, por hacer la vida más sencilla, pueden muchas veces no apreciarse en su justa medida. Gracias a mi familia y amigos de toda la vida, muchos de ellos lejos a causa de la pandemia pero, por lo mismo, más cerca que nunca.

Gracias a Carmen, al otro lado de esta mesa y a mi lado todos los días.

Por último, quiero recordar con cariño a cuatro personas que estaban con nosotros cuando empecé el doctorado, pero que ya no lo están. Tío Alberto, abuelo Guillermo, abuela Manuela y abuelo Raimundo.

A todos vosotros, gracias, que los momentos que compartimos en el día a día, más que un humilde prefacio, transmitan esta gratitud.

En Madrid, a 11 de noviembre de 2020

CONTENTS

1	INTRODUCTION	1
2	MODELS OF SOCIAL INTERACTION	5
2.1	The Efficiency metric and its universal distribution	5
2.2	Models of social interaction	6
2.2.1	Methods to obtain the efficiency distribution	7
2.2.2	Independent Variables: explaining the universality of the efficiency distribution	11
2.2.3	Identical Actors: introducing a parsimonious dependency between R and A	15
2.2.4	Distinguishable Actors: letting the system discriminate between individuals	20
2.3	Application to Wikipedia and the Scientific Citations network .	26
2.3.1	Wikipedia	26
2.3.2	Scientific citations network	27
2.4	Discussion	30
3	USER BEHAVIOR DURING ELECTORAL CAMPAIGNS	33
3.1	Twitter and electoral campaigns	33
3.2	Characterization of the datasets	34
3.3	Building the networks of mentions and retweets	35
3.4	Temporal evolution of user activity and interactions	37
3.4.1	Temporal evolution of user activity	37
3.4.2	Temporal evolution of mention and retweet networks .	41
3.5	Influence and communication patterns of politicians and regular users	48
3.5.1	Communication efficiency	48
3.5.2	Communication among the elites	49
3.6	Discussion	54
4	SENTIMENT ANALYSIS IN POLARIZED CONTEXTS	57
4.1	Building training sets automatically	57
4.2	Method to build and apply the training set	59
4.2.1	The Twitter datasets	59
4.2.2	Reference datasets	61
4.2.3	Construction of the training set	62
4.2.4	Sentiment classification algorithm	63
4.2.5	Choosing a classification threshold for unknown base rates	64
4.3	Testing the training set	67
4.4	Tuning the classification threshold	71
4.5	Discussion	73
5	MULTIPOLAR OPINION INFERENCE AND MEASUREMENTS	75
5.1	Opinion inference	76
5.1.1	The multidimensional opinion model	77
5.1.2	Pole vectors and the opinion simplex	78
5.1.3	Opinion projections	80
5.1.4	Selecting the opinions to project	82

5.2	Polarization measures for opinion distributions	84
5.2.1	One-dimensional measures of polarization	84
5.2.2	Multidimensional measures of polarization	88
5.3	Application to Twitter	89
5.3.1	Building the interaction networks	89
5.3.2	Elite selection	90
5.3.3	Listeners selection	91
5.4	Discussion	91
6	EMPIRICAL STUDY OF MULTIPOLAR SYSTEMS	93
6.1	Bipolar system: The Catalan independence issue	93
6.2	Tripolar system: The Catalan independence issue	96
6.3	Quadripolar system: Spanish General Elections of 2015 and 2016	101
6.4	Pentapolar system: Spanish General Elections of 2019	105
6.5	Discussion	108
7	OPINION-LANGUAGE INTERPLAY IN A TERRITORIAL CON- FLICT	111
7.1	Daily evolution of the opinion distribution	111
7.2	Language distribution	114
7.3	Joint distribution of language and opinion	115
7.4	Active, Influential and engaged users	116
7.5	Discussion	118
8	NETWORK (CO)VARIANCE AND ITS APPLICATIONS	121
8.1	Networks of knowledge	121
8.2	Pairwise distance between co-occurring concepts	123
8.3	Network (co)variance	127
8.4	(Co)variance of a network of knowledge	132
8.5	Other potential applications	136
8.6	Discussion	136
9	CONCLUSIONS AND FUTURE WORKS	139
9.1	Conclusions	139
9.2	Future works	144

Appendices

A	RANDOM VARIABLES AND PROBABILITY DISTRIBUTIONS	149
A.1	Random variables	149
A.1.1	Discrete random variables	149
A.1.2	Continuous random variables	151
A.2	Real-world probability distributions	153
A.3	Random vectors and joint probability distributions	154
B	DESCRIPTION OF THE DATASETS	157
B.1	Twitter datasets	157
B.2	Scientific citations datasets	159
B.3	Wikipedia editions dataset	159
C	TWITTER COMMUNICATION MECHANISMS AND TWITTER API	161
C.1	How does Twitter work	161
C.2	Twitter API	161
D	EFFICIENCY DISTRIBUTIONS OF TWITTER AND SCIENTIFIC CITATIONS	163

E	INTERMEDIATE FITS TO GET THE EFFICIENCY DISTRIBUTION	167
E.1	Hybrid fitting methodology for the $p(r)$ and $p(F)$ distributions	167
E.2	Independent variables model	168
E.2.1	Scientific citations	168
E.2.2	Twitter	168
E.2.3	Wikipedia	171
E.3	Identical actors model	171
E.3.1	Scientific citations	171
E.3.2	Twitter	174
E.3.3	Wikipedia	174
E.3.4	The error associated to the power-law fit of $p(r)$ affects the quality of the IdA model's analytical approximation	174
E.4	Distinguishable actors model: Twitter fits	179
F	SOCIOPOLITICAL CONTEXTS	187
F.1	The Spanish General Elections of 2015 and 2016	187
F.2	The Catalan independence issue	188
G	THE PageRank CENTRALITY MEASURE	191
H	THE RECEIVER-OPERATING CHARACTERISTIC (ROC) CURVE	193
I	BUILDING THE ELITE SETS TO APPLY MULTIPOLAR INFERENCE	195
	BIBLIOGRAPHY	203

LIST OF FIGURES

Figure 2.1	Example of efficiency distribution where the two distinct behaviors that are manifested to each side of the point $\eta = 1$ can be appreciated. The data corresponds to the Twitter conversation around the 2015 Spanish General Elections.	7
Figure 2.2	Diagram that summarizes the main characteristics of the three models: Independent Variables (InV), Identical Actors (IdA) and Distinguishable Actors (DiA). . . .	8
Figure 2.3	Diagram showing the regions of integration defined to each side of the line $\eta_o = \frac{r_m}{a_m}$ to obtain the cumulative distribution function of the efficiency.	10
Figure 2.4	Efficiency distributions corresponding to the InV model applied to the Twitter datasets. The plots show the empirical efficiency distribution (blue dots), the Monte-Carlo simulation (orange squares) and the analytical expression of equation (2.12) (green line). See Appendix D for the results of the other 10 datasets.	13
Figure 2.5	The discrepancy of the InV model with respect to the Twitter empirical data (Δ_{InV} , computed with equation (2.15)) is related to the Spearman rank correlation (ρ_e) between A and R. The figure shows a monotonous increase of Δ_{InV} with ρ_e , which has been characterized by means of a linear regression.	14
Figure 2.6	Efficiency distributions corresponding to the IdA model applied to the Twitter datasets. The plots show the empirical efficiency distribution (blue dots), the Monte-Carlo simulation (orange squares), the discrete computation (dashed red line) and the analytical expression of equation (2.21) (green line). See Appendix D for the results of the other 10 datasets.	18
Figure 2.7	Linear regression of the correlation induced by the IdA model (blue dots) with respect to the empirical correlation computed for the Twitter datasets. Each point corresponds to a different conversation. For comparison the values corresponding to the InV model, where R and A are not correlated, have also been plotted (orange squares).	19
Figure 2.8	Histogram of the Spearman's rank correlation between A and R for the Twitter datasets computed for individuals with low efficiency ($\eta < 1$) and high efficiency ($\eta \geq 1$). The small vertical lines in the bottom of the horizontal axis are a rug plot and represent the value of $\rho_e - \rho_{InV}$ for each individual dataset.	20

Figure 2.9	Efficiency distributions corresponding to the DiA model applied to the Twitter datasets. The plots show the empirical efficiency distribution (blue dots), the Monte-Carlo simulation (orange squares) and the discrete computation (dashed red line). See Appendix D for the results of the other 10 datasets.	25
Figure 2.10	Linear regression of the correlation induced by the DiA model with respect to the empirical correlation computed for the Twitter datasets. Each point corresponds to a different conversation. The reference values obtained for the InV model are also displayed.	26
Figure 2.11	Same as Figure 2.4 for the InV model applied to the Wikipedia dataset.	27
Figure 2.12	Same as Figure 2.6 for the IdA model applied to the Wikipedia dataset.	28
Figure 2.13	Same as Figure 2.4 for the InV model applied to the scientific citations datasets. See Appendix D for the results of the other 10 datasets.	28
Figure 2.14	Same as Figure 2.6 for the IdA model applied to the scientific citations datasets. See Appendix D for the results of the other 10 datasets.	29
Figure 3.1	Scheme that shows the method used to build the networks of retweets. In panel 1), user i posts a message starting an information flow that reaches user j , who decides to retweet it. In panel 2), the retweeted message posted by j is read by k who in turn retweets it again. In panel 3) the three users are information transmitters and in the last panel the resulting retweet network is shown.	36
Figure 3.2	Left panels: time series of aggregated user activity per day for the 2015 (top) and 2016 (bottom) elections. The shadowed region corresponds to the days of electoral campaign. Right panel: Linear regression between the activity time series of both elections. Each dot corresponds to one day and the lines correspond to the linear fits of the data including the day of the elections (red dashed line) and excluding it (black continuous line).	38
Figure 3.3	Power law relationships of the total number of tweets, retweets and mentions per day as a function of the number of unique users that participated in the conversation each day for the 2015 campaign (top) and the 2016 campaign (bottom). Note that the data corresponding to the day of the elections (marked with a circle) were not included in the fit.	39

Figure 3.4	Temporal evolution of the distribution of activity per day in both electoral campaigns (2015 in the top panels and 2016 in the bottom panels). Left panels: probability mass functions (PMF) of the distribution of activity for each day in color code. Right panels: daily evolution of the γ exponent of the power law fit of the activity distributions. The error bars correspond to 2σ	40
Figure 3.5	Left panels: strongly connected component of the aggregated mention networks for the 2015 and 2016 electoral campaign. Right panels: strongly connected component of the aggregated retweet network for the 2015 and 2016 electoral campaigns. Top panels: 2015 networks. Bottom panels: 2016 networks. Colors correspond to the communities computed with the Louvain algorithm [Blo+08]. We have indicated the most probable affiliations of the nodes of each community by visually inspecting which users correspond to the most central nodes. Every well defined group seems to correspond to a political party. The size of the nodes is proportional to $\log(\text{PageRank})$ (see Appendix G for more details).	42
Figure 3.6	Daily evolution of the γ exponent of the in and out degree distributions for mentions (@) and retweets (RT) networks for the 2015 (top) and 2016 (bottom) electoral campaigns.	44
Figure 3.7	Daily evolution of the directed degree assortativities for mentions (@) and retweets (RT) networks for the 2015 (top) and 2016 (bottom) electoral campaigns.	47
Figure 3.8	Probability distributions of efficiency for Twitter accounts associated to each party and for the whole set of users. Top: 2015 electoral campaign. Bottom: 2016 electoral campaign. The value $\eta = 1$ is marked with a dashed line.	49
Figure 3.9	Probability differences of having an efficiency that falls within a given bin ($\eta_i < \eta < \eta_{i+1}$) between political users with respect to the total of users. The blue dots and shadowed area corresponds to the average and standard deviation of the differences for 100 samples of 1000 randomly chosen users. Political accounts tend to exhibit higher probabilities of having high efficiencies and lower probabilities of having low efficiencies.	50
Figure 3.10	Adjacency matrices of C-networks where each supernode is the aggregation of political accounts belonging to a given party. Left panels correspond to mention networks and right panels to retweet networks. Top panels correspond to results of 2015 and bottom panels to results of 2016. Color is related to the proportion of mentions (retweets) directed from party i (row) to party j (column) by the following formula: $c_{ij} = \frac{A_{ij}}{\sum_j A_{ij}}$	52

Figure 3.11	Daily evolution of the assortative mixing by party and coalition for the mention (@) and retweet (RT) networks in 2015 (top) and 2016 (bottom) electoral campaigns. .	53
Figure 4.1	Diagram that shows the process adopted to retrieve tweets associated to the conversations around the Spanish elections and the football match. The subsequent filtering to extract the training set for sentiment analysis and the text processing to standardize and tokenize it is also shown.	60
Figure 4.2	Word clouds for the most used terms in each of the datasets: 2015 Spanish elections (left), 2016 Spanish elections (middle) and football match (right). The size is proportional to the frequency of each hashtag.	60
Figure 4.3	Diagram that illustrates the idea that in a polarized context, members of a pole will post positive messages about other partners of their own pole and negative messages about members of different poles.	62
Figure 4.4	Visualization of a confusion matrix in a binary classification task. This figure is a modified version of [Wal]. .	65
Figure 4.5	Dependence of the precision p with respect to the proportion of positive samples P and the positive likelihood ratio LR_+ . Although LR_+ is not bounded, the empirical results suggest that it is reasonable to expect that a classifier will not exceed a value of $LR_+ = 20$	66
Figure 4.6	Precision-recall curves for the test sets and the validation sets of the dataset corresponding to $C1 = TASS + STOMPOL + SPTGE$ (left panel) and $C2 = SPTFM + TASSf$ (right panel). The estimated precision curve applying equation (4.5) to the results for the test set and taking into account the base rates is also shown. The curves have been computed averaging 300 realizations. The shadowed areas correspond to 1σ	72
Figure 5.1	Diagram of the opinion spaces of dimension 1 (left), 2 (middle) and 3 (right). While in 1D the poles can be intuitively labeled as against (-1) or in favor ($+1$) of something, in higher dimensions they are labeled with numbers from 0 to $n - 1$ where n is the number of poles. The neutral point (the barycenter of the simplex's vertices) is marked with a black square.	78
Figure 5.2	A-D: Examples of the opinion position of a listener node for three different weighted networks of a tripolar system. Top: network of the tripolar system with one listener node in gray (L) connected to three elite nodes corresponding to the three poles. The numbers on the links are their weights. Bottom: the corresponding opinion simplex with the opinion position of the listener node.	80

Figure 5.3	Opinion projections in different multidimensional spaces: A. One-dimensional projection of tripolar (2D) opinions to the $(0, 1)$ edge; B. One-dimensional projection of quadripolar (3D) opinions to the $(0, 1)$ edge and C. Two-dimensional projection of quadripolar (3D) opinions to the $(1, 2, 3)$ face.	81
Figure 5.4	Diagram of all the <i>adversarial</i> directions in a tripolar opinion space.	81
Figure 5.5	Tripolar opinion distribution projected onto the $(1, 2)$ edge when all the listeners' opinions are equally taken into account (left) and when only the opinions of the ones closest to poles 1 and 2 are considered (right). The resulting 1D distribution is represented as a gray histogram with its maximum rescaled to have a fixed height to facilitate comparison between projections.	83
Figure 5.6	A: Voronoi cells of a cloud of points in the Euclidean plane. Dashed lines are those with an end point in infinity. B: Voronoi cells of the vertices (the poles) of an equilateral triangle (a tripolar opinion space).	83
Figure 5.7	Examples of distributions with different kurtosis.	86
Figure 5.8	Diagram of a network showing the criterion to select the listeners once the elite has been chosen. Listener users (blue) are those which have a directed path to an elite user (orange). The gray users are those that are neither elite, nor have a directed path to any elite user. Two examples of directed paths from two circled listeners to two elite users are marked in white. Links are colored according to their target node.	91
Figure 6.1	Retweets received vs. percentage of participation days of users in the Catalan independence Twitter conversation. Colored rectangles represent three subsets of possible elite users according to different criteria thresholds. We will work with the pink set, although the outcome is practically the same with the other two sets.	95
Figure 6.2	Left: opinion distribution of all the users and of the users with an activity corresponding to more than 10 tweets in the Catalan independence conversation. Right: distribution of user activity. The shadowed area shows the percentage of users that have posted more than 10 tweets in the whole period.	95

Figure 6.3	Heat maps and contour plots of the opinion distribution for the Catalan independence Twitter conversation analyzed as a tripolar system. The white square indicates the position of the center of mass of the tripolar distribution. The two-headed arrow centered at the barycenter of the simplex (marked with an empty circle) shows the direction of maximum polarization. The 1D projections onto each edge of the simplex are shown on the sides of the triangles. In the left panel, the 1D projections include the opinions of all the listeners while in the right panel only the opinions of the listeners closer to the two poles of each edge are considered. The labels correspond to the against-independence pole (0), the pro-independence pole (1) and Podemos (2).	97
Figure 6.4	Comparison of the opinions associated to the Catalan independence inferred with the bipolar opinion inference technique and the tripolar one. In the left panel we plot the linear regression between the two opinions (each point is a user's opinion) and in the right panel we plot their corresponding opinion distributions.	98
Figure 6.5	1D adversarial projections of the 2D opinion distribution shown in Figure 6.3, which corresponds to the Catalan independence case study.	99
Figure 6.6	Projection of the Catalan independence tripolar opinion distribution onto the direction of maximum polarization. The projected positions of the poles are shown as vertical dashed lines.	100
Figure 6.7	Process followed to compute the 1D projection of a tripolar (2D) opinion distribution onto the direction of maximum polarization (first PC). A: The cloud of opinion points with the barycenter of the simplex marked as an empty circle and a double-headed arrow indicating the direction of maximum polarization. B: Same as A plus the projections of the opinion points (white squares) and the poles (red stars) onto the direction of maximum polarization. C: The corresponding projected opinion distribution.	100
Figure 6.8	Quadripolar opinion distributions of the 2015 (left) and 2016 (right) Spanish elections. The distributions are shown as heat maps and contour plots of the 2D projections onto the faces of the simplex. The centers of mass of the projected opinion distributions are represented as white squares and the projection of the direction of maximum polarization (first PC), as a double headed arrow. The 1D projections onto each edge of the simplex are shown on the sides of the triangles. The poles are PP (0), PSOE (1), Podemos (2) and Cs (3).	102

Figure 6.9	Directions of the PCs for the 2015 (left) and 2016 (right) Spanish elections represented as double arrows centered at $\vec{O}$. Their length is proportional to their explained variance. The center of mass of the opinion distribution is represented as a red square, which in this case is between pole 2 and $\vec{O}$ but closer to the latter.	103
Figure 6.10	Projection of the 2015 (left) and 2016 (right) elections opinion distribution onto the direction of maximum polarization. The projected positions of the poles are shown as vertical dashed lines.	104
Figure 6.11	Process followed to compute the 1D projection of a quadripolar (3D) opinion distribution onto the direction of maximum polarization (first PC). A: the cloud of opinion points with the barycenter of the simplex marked as an empty circle and a double-headed arrow indicating the direction of maximum polarization. B: same as A plus the projections of the opinion points (white squares) and the poles (red stars) onto the direction of maximum polarization. C: the corresponding projected opinion distribution.	104
Figure 6.12	Contour plots corresponding to the 2D projections of the opinion distribution of the 28A and 10N Spanish elections onto the faces of the simplex. The centers of mass of the projected opinion distributions are represented as white squares and the projection of maximum polarization as a double-headed arrow. The 1D projections onto each edge of the simplex are shown on the sides of the triangles. The pole labels correspond to PP (0), PSOE (1), Podemos (2), Cs (3) and Vox (4).	106
Figure 6.13	Projection of the 28A (left) and 10N (right) opinion distributions onto the adversarial direction (PSOE,Podemos) vs (PP,Cs,Vox) that defines the left-wing / right-wing axis.	107
Figure 6.14	Projection of the quadripolar (3D) opinion distribution of the 28A (left) and 10N (right) elections onto the direction of maximum polarization. The projected positions of the poles are shown as vertical dashed lines.	108
Figure 6.15	Bubble chart where the sizes of the bubbles are proportional to the number of listeners closer to the corresponding pole (that is, within its Voronoi cell) and the height of the position of the bubble indicates the distance from the center of mass of those listeners to the neutral point, i.e., the barycenter of the simplex, which is opinion $\vec{O}$. Left panel shows the chart for the 28A election and right panel for the 10N election.	108

Figure 7.1	Temporal evolution of the number of tweets published in a Twitter conversation about the Catalan independence issue from 09/11/2017 to 11/04/2017. The three highest activity peaks match the off-line events of: the referendum day (1-0), the Catalonia independence declaration signed and suspended on the same day, and the Unilateral Declaration of Independence (DUI) and the enforcement of the Article 155 of the Spanish Constitution.	112
Figure 7.2	Time evolution of the daily distributions of opinion index for the Catalan independence twitter conversation. Color indicates the number of users that participate each day. The three days framed in red correspond to the three activity peaks observed in Figure 7.1.	113
Figure 7.3	Temporal evolution of the number of users and the normalized pole distance (d), relative population size (ΔA) and the polarization index μ for the Catalan independence conversation.	113
Figure 7.4	Distributions of language index (L) for (a) Elite users, where the pro-independence elite users are in blue and the against-independence in orange, (b) Listeners users and (c) Geo-located in Catalonia users.	115
Figure 7.5	Interplay between language index and opinion index of the users of the Catalan independence Twitter conversation. Color measures the number of users located in the 2D bin. The red $\times$'s correspond to Geo-located in Catalonia users.	116
Figure 7.6	Left side: Comparison of the effect of activity, influence and engagement on the users' opinion distribution. Results are filtered by activity (blue), retweets received (purple) and days of participation (pink) for three different threshold values increasing from top to bottom panels. Right side: Effect of the activity filters in the interplay between language index and opinion index. Color indicates the number of users located in each 2D bin.	117
Figure 8.1	Diagram showing how the arXiv co-occurrence network (left) and Wikipedia hyperlink network (right) are built from their respective datasets of scientific papers and Wikipedia pages.	122
Figure 8.2	Giant connected component of the Hyperlink network of Wikipedia pages corresponding to the considered set of mathematical concepts. The size of the nodes is proportional to their PageRank (see Appendix G) and the color corresponds to the communities according to the Louvain algorithm.	123

Figure 8.3	Co-occurrence network of the considered set of mathematical concepts in scientific papers. The size of the nodes is proportional to their PageRank. In this case the Louvain algorithm does not find any significant community, so we have assigned the same color to all the nodes. 124
Figure 8.4	Minimum path distance distribution between nodes in the (undirected version of the) Wikipedia network for pairs co-occurring concepts and for random pairs of nodes (which, as expected, coincides with the minimum path distance distribution for all the possible minimum paths). 125
Figure 8.5	Diagram that shows the two possible ways to form a triangle in the co-occurrence network. 126
Figure 8.6	Average minimum path distance distribution between nodes in the (undirected version of the) Wikipedia network for triads of co-occurring concepts and for random triads of nodes according to the restrictive null model. 126
Figure 8.7	Four different distributions on the same network that are increasingly spread out from A to D and therefore present increasing values of the network variance. . . . 129
Figure 8.8	Top: Three clouds of points representing a positive correlation (A), no correlation (B) and a negative correlation (C) between two variables x and y . Bottom: Their corresponding joint probability distributions. 132
Figure 8.9	Distribution of network variances of concept distributions $p^{(i)}$ in papers, calculated with respect to the geodesic distance in the Wikipedia network. In each panel, the probability density function (PDF) of variances is calculated for papers containing 3 (resp. 5,7,9) concepts present in the Wikipedia network, and is compared with the variance PDF for a collection of (null model) virtual papers with the same number of concepts. The empirical variance distributions seem to be concentrated on smaller variances compared to the null model variances. 133
Figure 8.10	Distribution of network variances of concept distributions $p^{(i)}$ for arXiv papers from different sub-fields. . . 134
Figure 8.11	Comparison between empirical covariance of the co-occurrence distribution P on the Wikipedia network of concepts and two null models. In the left panel, the covariance of P is calculated with respect to the shortest-path distance on the Wikipedia network (dashed line) and the shortest-path distance on 100 realizations of a degree-preserving randomization of the network (blue bars). In the right panel, the covariance of P (dashed line) and of 100 realizations of a marginal-preserving randomization of P (blue bars) are calculated with respect to the shortest-path distance on the Wikipedia network. . . . 135
Figure A.1	Visualization of the PMF of equation (A.5). 150

Figure A.2	Histogram of an experiment that consists in flipping a coin 4 times and counting the number of heads. . . .	151
Figure A.3	PDF (blue line) and CDF (dashed green line) of a continuous r.v. with normal distribution of mean 0 and variance 1. The probabilities of finding the continuous r.v. in the intervals $[-1.5, -1]$, $[0, 1]$ and $[2, 4]$ are shown as orange bars with height equal to said probability and span coinciding with the corresponding interval.	152
Figure A.4	PDF of the power law given by equation (A.13) in a linear plot (A) and in a log-log plot (B). Panel B also shows estimations of the PDF from 100k random samples computed with a histogram with 20 regularly spaced bins (orange dots), 5000 regularly spaced bins (green squares) and 20 logarithmically spaced bins (red crosses). . . .	154
Figure A.5	Uniform PDF in the interval $[0, 1]$ and histograms computed with 100k random samples. A: histograms with regular bins (orange dots) and logarithmically spaced bins (green squares) normalized by dividing the frequency of each bin by the total number of samples. B: histograms with regular bins (orange dots) and logarithmically spaced bins (green squares) normalized by dividing the frequency of each bin by the total number of samples and the width of the bin to obtain a proper density. . .	154
Figure A.6	Bivariate joint probability distribution $P(x_1, x_2)$ shown as a 3D surface. A: Diagram of the aggregation process to compute the marginal probability $p(x_1)$. B: Diagram of the computation of the conditional probability $p(x_1 x_2)$ from the joint probability distribution $P(x_1, x_2)$	155
Figure D.1	Efficiency distributions corresponding to the InV model applied to the Twitter datasets. The plots show the empirical efficiency distribution (blue dots), the Monte-Carlo simulation (orange squares) and the analytical expression of equation (2.12) (green line).	163
Figure D.2	Efficiency distributions corresponding to the IdA model applied to the Twitter datasets. The plots show the empirical efficiency distribution (blue dots), the Monte-Carlo simulation (orange squares), the discrete computation (dashed red line) and the analytical expression of equation (2.21) (green line).	163
Figure D.3	Efficiency distributions corresponding to the DiA model applied to the Twitter datasets. The plots show the empirical efficiency distribution (blue dots), the Monte-Carlo simulation (orange squares) and the discrete computation (dashed red line).	164

Figure D.4	Efficiency distributions corresponding to the InV model applied to the scientific citations datasets. The plots show the empirical efficiency distribution (blue dots), the Monte-Carlo simulation (orange squares) and the analytical expression of equation (2.12) (green line).	164
Figure D.5	Efficiency distributions corresponding to the IdA model applied to the scientific citations datasets. The plots show the empirical efficiency distribution (blue dots), the Monte-Carlo simulation (orange squares), the discrete computation (dashed red line) and the analytical expression of equation (2.21) (green line).	165
Figure E.1	Fit of the empirical distribution of activity of the scientific citations network to a truncated power law. Blue dots correspond to the full data and orange squares to the data selected to perform the fit ($A \in [1, 30]$).	169
Figure E.2	Fit of the empirical distribution of response of the scientific citations network to a truncated power law. Blue dots correspond to the full data and orange squares to the data selected to perform the fit ($R \in [2, 300]$).	169
Figure E.3	Fit of the empirical distribution of activity of the Twitter datasets to a power law (dashed green line) and to a truncated power law (continuous orange line).	170
Figure E.4	Fit of the empirical distribution of response of the Twitter datasets to a power law (dashed green line) and to a truncated power law (continuous orange line).	170
Figure E.5	Fit of the right tail of the distribution of efficiency of the Twitter datasets to a power law.	172
Figure E.6	Fit of the empirical distribution of activity of the Wikipedia dataset to a power law (dashed green line) and to a truncated power law (continuous orange line).	172
Figure E.7	Fit of the empirical distribution of response of the Wikipedia dataset to a TPL.	173
Figure E.8	Fit of the empirical distribution of activity of the scientific citations network datasets to a PL.	173
Figure E.9	Hybrid lognormal fit of the empirical distribution of response to single actions of the scientific citations network datasets.	175
Figure E.10	Power law fit of the empirical distribution of response to single actions of the scientific citations network datasets.	175
Figure E.11	Hybrid lognormal fit of the empirical distribution of response to single actions of the Twitter datasets.	176
Figure E.12	Power law fit of the empirical distribution of response to single actions of the Twitter datasets.	176
Figure E.13	Hybrid PL fit of the empirical distribution of response to single actions of the Wikipedia dataset.	177
Figure E.14	Power law fit of the empirical distribution of response to single actions of the Wikipedia dataset.	178

Figure E.15	Relationship between the fitting error committed when $p(r)$ is modeled as a power law (ϵ_r) and the deviation of the analytical approximation with respect to the numerical computation of the IdA model (ϵ_η) in the right tail of the efficiency distribution $f(\eta)$	181
Figure E.16	Fit of the empirical distribution of followers of the Twitter datasets to the empirical values with a PL tail. . . .	181
Figure H.1	ROC curves for a random classifier, a perfect classifier and the intermediate curve that is typically obtained in empirical settings.	193
Figure I.1	Levels 3 to 7 of the hierarchical community structure for influential and engaged users of the Catalan independence system. The level used to build the elite is highlighted in light blue and the specific communities assigned to each pole are marked with squares of different colors. Yellow: In favor of independence. Red: Against independence. Purple: Podemos (third pole). .	197
Figure I.2	Levels 2 to 5 of the hierarchical community structure for influential and engaged users of the 2015 elections system. The level used to build the elite is highlighted in light blue and the specific communities assigned to each pole are marked with squares of different colors. Blue: PP. Red: PSOE. Purple: Podemos. Orange: Cs. . . .	198
Figure I.3	Levels 2 to 7 of the hierarchical community structure for influential and engaged users of the 2016 elections system. The level used to build the elite is highlighted in light blue and the specific communities assigned to each pole are marked with squares of different colors. Blue: PP. Red: PSOE. Purple: Podemos. Orange: Cs. . . .	199
Figure I.4	Levels 2 to 7 of the hierarchical community structure for influential and engaged users of the 28A elections system. The level used to build the elite is highlighted in light blue and the specific communities assigned to each pole are marked with squares of different colors. Blue: PP. Red: PSOE. Purple: Podemos. Orange: Cs. Green: Vox.	200
Figure I.5	Levels 1 to 8 of the hierarchical community structure for influential and engaged users of the 10N elections system. The level used to build the elite is highlighted in light blue and the specific communities assigned to each pole are marked with squares of different colors. Blue: PP. Red: PSOE. Purple: Podemos. Orange: Cs. Green: Vox.	201

LIST OF TABLES

Table 2.1	Cut-off values used to perform the numerical computations for the IdA model.	16
Table 3.1	Names of the four main political parties contending in the 2015 and 2016 Spanish general elections and the minority left-wing party (IU) that formed a coalition with Podemos (named UP) before the 2016 election.	35
Table 3.2	Values of the exponents (α) and correlation coefficients (r^2) of the linear regressions of the log-log plots of tweets, retweets and mentions with respect to the number of unique users per day shown in Figure 3.3 for both electoral campaigns.	39
Table 3.3	Values of the modularity (Q) for the communities of the retweet and mention network computed with the Louvain algorithm.	42
Table 3.4	General statistical properties of the retweet and mention networks aggregated for the whole period of study for both electoral campaigns. $k_{in(out)}$ corresponds to in (out) degree and $\gamma_{in(out)}$ to the exponent of the degree distribution. The confidence intervals in the power law exponents calculations correspond to 2σ	45
Table 3.5	Assortativities of the different aggregated networks.	46
Table 3.6	Z-Scores of the assortativities of the different aggregated networks with respect to 500 realizations of the directed configuration model.	46
Table 3.7	P-values of the paired t-test performed in the assortative mixing time series of parties and coalitions. The test has been carried out by taking each time series of 2015 and coupling it with its counterpart of 2016.	54
Table 4.1	Average area under the receiver-operating characteristic curve (ROC) for a stratified 3-fold cross validation for the electoral context datasets. The model is trained in a single dataset or a combination of them. Then it is tested in unseen samples of every corpus (including unseen samples from the corpus used as training set). The Combination column corresponds to the area under the ROC for a test set composed of samples of the combination of corpora used as training set (in the case of a single corpus it coincides with the value obtained for the corresponding test set of the same name). The uncertainty corresponds to 1σ	69

Table 4.2	Average area under the receiver-operating characteristic curve (ROC) for a stratified 3-fold cross validation for the football context datasets. The model is trained in a single dataset or a combination of them. Then it is tested in unseen samples of every corpus (including unseen samples from the corpus used as training set). The Combination column corresponds to the area under the ROC for a test set composed of samples of the combination of corpora used as training set (in the case of a single corpus it coincides with the value obtained for the corresponding test set of the same name). The uncertainty corresponds to 1σ	70
Table 5.1	Distance between poles for a regular simplex of different dimensions with vertices placed at distance $u = 1$ from its center.	78
Table 6.1	Minimum retweets and participation ratios of the elite sets corresponding to the three rectangles drawn in Figure 6.1.	94
Table 6.2	Values of the polarization index and its associated parameters, defined in Section 5.2.1.10, that characterize the opinion distributions of Figure 6.2.	96
Table 7.1	Precision and recall of testing the method [Shu10] to detect Catalan and Spanish languages.	114
Table 8.1	Examples of covariance and correlation of some joint probability distributions defined on pairs of nodes of the network of Figure 8.7.	131
Table 8.2	Covariance of the joint probability distribution of concepts for papers of different sub-fields.	135
Table B.1	Description of the Twitter datasets. (*) Due to the high quantity of keywords of the 28A and 10N Spanish elections of 2019, we have put them in an online repository.	158
Table B.2	Description of the datasets of scientific citations.	160
Table E.1	Parameters of each fit performed with the scientific citations datasets.	182
Table E.2	Parameters of each fit performed with the Wikipedia data.	183
Table E.3	Parameters of each fit performed with the Twitter datasets.	184
Table E.4	Values of p_{inf} estimated through MLE for the Twitter datasets. The averages and standard deviations obtained for 10 samples of 1000 instances each are presented.	185
Table F.1	Results of the Spanish general elections of 20 December 2015 and 26 June 2016. Podemos and IU are together in 2016 because they formed a coalition called UP.	188

Table I.1	Minimum retweets (RT_{min}) and minimum participation ration (PR_{min}) used to filter the influential and engaged users (<i>INITIAL ELITE</i>) for each Twitter conversation. The <i>SELECTED ELITE</i> corresponds to the number of users of the <i>INITIAL ELITE</i> that have been assigned to a pole taking into account the community structure and contextual information. The number of <i>LISTENERS</i> and the number of <i>LINKS</i> of the final elite-connected network are also shown.	196
Table I.2	Number of users assigned to each pole of the Catalan independence system.	197
Table I.3	Number of users assigned to each pole of the Spanish elections of 2015.	199
Table I.4	Number of users assigned to each pole of the Spanish elections of 2016.	200
Table I.5	Number of users assigned to each pole of the Spanish elections of 2019 28A.	200
Table I.6	Number of users assigned to each pole of the Spanish elections of 2019 10N.	201

INTRODUCTION

"And tell me, what do you think of all this. A finished work of art, is it not?"

"Definitely!"

"Wrong! It is not."

The Seldon Plan is neither complete nor correct.

Instead, it is merely the best that could be done at the time.

Over a dozen generations of men have pored over these equations, worked at them.

*Before you obtain your Speakerhood, you yourself will have to make
an original contribution."*

— Isaac Asimov - *Second Foundation*

We live in a digital era, continuously interacting with electronic devices in our everyday activities and leaving a digital footprint: phone call records, credit card purchases, GPS history data, email communication, etc. The characteristics of an individual are encoded in her digital footprint and this information is hoarded by companies and institutions to compile personal profiles that are used to tailor personalized advertisements [CGH18; Goo19; Goo15; AEK00], optimize commuting routes [MPC19; Fab19; Goo20; Goo09] or recommend movies and songs [BL07; DS20; Sch+18]. These data are so valuable because they are generated by spontaneous actions; but the real power of the digital footprint lies in aggregating the data of many individuals. When this is done we can unveil behavioral regularities of the society seen as a whole physical system rather than a collection of isolated individuals [Pen15; APCP16; MG+16; Mar+16; Qua+14].

The set of techniques and methodologies used to analyze these large quantities of data are usually referred to as Big Data and the models and theories developed from them can be framed within the field of Computational Social Science [Laz+09]. These disciplines may seem very new and sometimes considered revolutionary. However, their underlying ideas and motivations can be traced back to a very long tradition of scientists that studied social dynamics from a quantitative point of view by placing data at the center of the discussion. Early examples can be found in the 1693 *life table* elaborated by astronomer Edmond Halley [Edm93] to analyze mortality rates, the 1781 study of the ratio of the births of boys to girls by mathematician Pierre-Simon Laplace [DL95] or the London map of the 1854 cholera outbreak by the physician John Snow [Sno55]. In fact, Laplace's work was taken up by astronomer Adolphe Quetelet, who adopted the term *Social Physics* (originally coined by philosopher Auguste Comte) for this emerging discipline. Quetelet's method, however, departed radically from Comte's approach, as the latter eschewed the use of mathematics or even of numbers in the social disciplines [Por86; Bal02].

Our collective behavioral patterns arise from interactions. For this reason the field of Network Science appears as an optimal tool for the study of social systems [Bor15; Oli19]. In this discipline a system is modeled as a set of simple elements (the nodes) and a set of relationships between them (the links). A natural abstraction of a social system is to consider each person as a node and each interaction (sending an email, meeting face to face, being friends...) between a pair of individuals as a link

[Mig+17; Zha+12]. The usual modeling approach of Network Science, directly inherited from Social Physics, consists in explaining the macroscopic properties of a system while assuming the simplest microscopic interactions between the elementary components to extract the most fundamental laws [New10; Say15]. The models are usually informed by or inspired in the physical sciences [Gol+20; PSV07; Mor+15; AB02; CMPS09; Pei14; ASMG20]. In this thesis we propose models of social interaction to explain the relationship between the actions of individuals embedded in a social system and the collective response of the system.

From the simple social interactions that interweave social networks, higher order structures and macroscopic phenomena emerge, like the modular structure of the networks [New06] or the rich-gets-richer phenomenon [Pri76; BA99]. In this thesis we focus on polarization, a property of social systems where different groups of people hold opposing views on a given matter in an antagonistic manner. Polarization can be found in recent and relevant contexts like Brexit [Hob16; Bau+20; Bau+20], the Catalan independence movement in Spain [BRT17; AB+19], the 2016 USA presidential elections [Gro+19], the Turkish constitutional referendum of 2017 [YS17] and many other social issues [DEB96; BBC; BKP08; AB11; Oli+19; HK16; Pri13]. Furthermore, it has important social consequences, since it affects public debates and influences the actions of the government [Dal08; Bin00].

When investigating social phenomena it is crucial to find useful and reliable data sources. We mainly use datasets from the Twitter social networking site which currently has around 330 million monthly active users [Cle19]. The concise character of a tweet (a message posted in Twitter) turns it into a powerful tool to share breaking news and take part in dynamic debates. It is one of the most used media by journalists, politicians, and politics enthusiasts to share their views and news [Par13]. Moreover, Twitter's API allows researchers to retrieve user data, enabling us to access a portion of the user's digital footprint with relevant information for the study of multiple phenomena, including influence, sentiment, polarization, etc. After all, online user behavior is a reflection of the offline world [Mor14; CP15].

Polarization is frequently measured by building and analyzing distributions of attitudes or opinions with respect to some topic. Traditionally, attitude distributions are built from opinion polls, but the adoption of online social media as a means of public debate has stimulated the emergence of alternative opinion inference techniques capable of estimating attitudes for massive populations. We are talking about Sentiment Analysis [MHK14], a discipline that consists in analyzing text samples (in this case online messages) and assigning them a label according to the sentiment they convey. Some of the most researched and successful techniques for Sentiment Analysis are those based in Supervised Machine Learning classification models, but been supervised means that a large quantity of manually labeled text samples are needed to train the models. In this thesis, we develop a methodology to mitigate this inconvenience.

However, besides the drawbacks discussed above, Supervised Sentiment Analysis suffers from additional burdens which are difficult to avoid for its application to the study of polarization. For example, the output of classification models is discrete, as they assign a label out of a set of labels of limited size. In fact, the models often use only two labels (positive / negative) or five at most. But when working with opinion distributions, the greater the number of possible attitudes the more detailed the information that can be extracted. Ideally, it is preferable that the opinions be continuous.

For these reasons, we also use an alternative technique for opinion inference based in the network of interactions between users (who do you communicate with) instead of the content of the interaction (what message do you send). This technique, based on the DeGroot learning model [DeG74], has the advantage of providing a continuum of opinions instead of a crisp adherence to a given one. However, like the sentiment analysis approach, it is only useful for bipolar situations, as opinions take values along a one-dimensional line between two extreme and opposed positions. Another contribution of this thesis is the generalization of this framework for contexts with an arbitrary number of poles by considering an ideological spectrum of higher dimensions.

Opinion however does not appear isolated in a social context. It is related to other social dimensions. In order to get insight into the main features of a polarized context, it is often useful to analyze the way people interact and debate. In many significant scenarios, the differences in the use of language or even the use of different languages are key to understand the mechanisms underlying polarization. After all, it is well known that the use of different languages result in different cognitive mechanisms in the people who use them [GM+08; Sus99]. Following this idea, we also study how language is related to opinion in the context of a territorial dispute (the Catalan independence movement), as this kind of conflict has been consistently linked to the use of different languages [IW67; SG82; Cra11].

To further study polarization on social networks [MF13; Bor+12b; Mor+15] we develop two new network metrics: the Network Variance and the Network Covariance. Nevertheless, just like the usual variance and covariance, these are general-purpose measures with applications in many other fields.

To summarize, the main goal of this thesis is to improve our understanding of collective human behavior both at a fundamental level, by developing models of social interaction, and at the level of the emergent properties that arise from these interactions, placing a special emphasis on the study of social polarization. To achieve this global goal, we have set the following specific objectives:

1. Develop models based on the interactions between individuals that explain the relationship between the individual actions and the corresponding stimulated collective responses.
2. Characterize online user behavior in polarized contexts and search for regular behavioral patterns.
3. Develop a methodology to build training sets for sentiment analysis automatically, thus avoiding hand labeling large amounts of text samples.
4. Develop an opinion inference technique and polarization metrics suitable for multipolar contexts.
5. Apply the multipolar methodology to real-world social systems and reveal their polarization patterns.
6. Derive expressions to calculate the variance and the covariance of distributions defined over the nodes of a network.

The rest of the chapters of this thesis are organized as follows:

- In Chapter 2 we propose three statistical models of social interaction developed in a parsimonious way to explain the relationship between the actions of individuals embedded in a social system and the collective response of the system. In particular, we focus on reproducing the empirical distribution of efficiency (η), a metric defined as the ratio between the number of reactions of the system (R) and the number of actions performed by the individual (A); that is, $\eta = R/A$.
- In Chapter 3 we provide a broad perspective of user behavior in polarized contexts by analyzing a Twitter conversation around the electoral campaign of the Spanish General Elections of 2015 and 2016. We show that there are recurrent patterns of user behavior in the users' activity and interaction networks.
- In Chapter 4 we approach the problem of opinion inference from the perspective of sentiment analysis. Here, user opinion is extracted from the content of the communication (the text of the message). We develop a methodology to semi-automatically build a training set for supervised machine learning models in a polarized context.
- In Chapter 5 we tackle the problem of multipolar opinion inference using user interactions instead of content. We extend and generalize a previous technique only suitable for bipolar contexts so that it can be used for an arbitrary number of opinion poles.
- In Chapter 6 the framework developed in 5 is applied to empirical data from several multipolar scenarios.
- In Chapter 7 we analyze a Twitter conversation around the Catalan independence issue taking into account the content of the messages to study the use of the different languages (Catalan or Spanish) and explore their relationship with ideology.
- In Chapter 8 we present the Network Variance and Covariance metrics and show how they can be applied to measure the strength of cohesion of different groups of nodes in a network.
- In Chapter 9 we lay out the conclusions of this work and potential lines of research for future projects.

The final part of this thesis includes nine appendices and the bibliographic references. Appendix A provides a brief introduction to random variables and probability distributions; Appendices B and C include details about the datasets used in this work; in Appendices D and E we present complementary results for Chapter 2; in Appendix F we describe the sociopolitical contexts of systems studied in Chapters 3, 6 and 7; in Appendix G we briefly explain the PageRank centrality metric; in Appendix H we discuss the use of the ROC curve for the evaluation of classification models and in Appendix I complementary results for Chapter 6 are presented.

MODELS OF SOCIAL INTERACTION: THE RELATIONSHIP BETWEEN INDIVIDUAL ACTIONS AND COLLECTIVE RESPONSE

*It can scarcely be denied
that the supreme goal of all theory
is to make the irreducible basic elements
as simple and as few as possible
without having to surrender
the adequate representation
of a single datum of experience.*

— Albert Einstein - *On the Method of Theoretical Physics*

While we as individuals have the ability to influence those that are close to us (friends, family, colleagues...), the structure of our social networks can propagate that influence far beyond our personal social circle. Social interactions are usually studied from the perspective of social influence by focusing on how ideas are spread and *adopted* [Fla+17]. Although we will discuss this approach to the topic of social influence in later chapters (after all, opinion polarization is one of the core topics of this thesis), in the present one we take a broader point of view. We are not interested in how ideas are adopted, but in the extent of the response they trigger. Specifically, our goal in this chapter is to find the laws that relate the actions of an individual that is part of a social system to the stimulated collective reactions.

This topic has received considerable attention and has been studied in different but closely related disciplines [JP19; DDA20; Muc+13; Ryb+09]. From diffusion on networked systems, a field which studies the spread of diseases or information and the emergence of cascading phenomena [G+13; GG+15; Zha+16b; PS+15; MPL17] to *virality*, a property of certain pieces of information that generate a wide response in social systems [Goe+16; IM11; WMA13]. Other works focus on the Influence Maximization problem, taking advantage of the diffusion mechanisms to find a set of individuals whose actions would maximize the response [KKT03; Zha+16a; Li+18; Azc+18]. Taking an alternative approach, the field of control theory aims to steer the collective behavior of a system by controlling the activity of a few individuals [LB16; CC17].

In this chapter we develop a theoretical framework that relates the number of actions performed by an actor (an agent or individual) embedded in a social system; that is, her activity (A), and the number of reactions that these actions trigger in her peers, or response (R) [MGLB20b].

2.1 THE EFFICIENCY METRIC AND ITS UNIVERSAL DISTRIBUTION

To relate activity and response we generalize the efficiency metric, introduced by Morales et al. in the context of Twitter [Mor+14], to other social systems. The

efficiency metric is defined as the ratio between collective response R and individual activity A :

$$\eta = \frac{R}{A} \quad (2.1)$$

It can be considered as a proxy for how efficient an individual is at triggering reactions in her peers or as a measure of the system's inertia to react to the stimuli of the individual. The higher the individual's efficiency, the lower the system's inertia.

We follow a well established modeling approach in social physics: explain the macroscopic properties of the system assuming the simplest microscopic interactions between the actors to extract the most fundamental laws [Say15; New10; Iac+19; WLL19; Zen+19; Can+19; All+17; Tam+18]. The macroscopic property in which we focus is the distribution of efficiency (see Appendix A for a brief introduction to random variables and probability distributions). We have used this metric to analyze three kinds of social systems of different nature: social networks, collaborative networks and citations networks. In particular, we have worked with 14 Twitter conversations around different issues in Spain, Turkey, Palestine, Argentina and Colombia, the editions of the English Wikipedia and the scientific citations data of authors from 14 different countries extracted from the Web of Science. We provide a detailed description of the datasets in Appendices B and C.

In Twitter, the activity is the number of original messages posted by a user and the response of the system is the number of retweets received by that user. Another magnitude used in our analysis is the partial response to single actions (r), which is the number of retweets obtained by a single tweet. In the scientific citations network, A is the number of publications of an author and R the number of citations obtained. The variable r in this case is the number of citations obtained by one paper. In the context of the Wikipedia collaboration network, we consider A as the aggregated number of editions performed by a particular user in any Wikipedia page. The corresponding R is the number of editions made by other users in her personal user page. These editions can be considered as messages directed to that particular user. In this case there is no data for the response to a single edition. Therefore, we have defined r as the number of editions made on the pages of users whose activity is $A = 1$.

In a previous study [Mor+14] it was shown that the shape of the distribution of efficiency in Twitter conversations of very different nature and size is universal. We have found further evidence of this universality for Twitter and also for the other two systems considered (see Appendix D). The efficiency distribution, an example of which is shown in Figure 2.1, always presents a concave shape with two distinct tendencies for $\eta < 1$ (an individual gets less than one reaction per action) and $\eta > 1$ (an individual triggers more than one reaction per action).

2.2 MODELS OF SOCIAL INTERACTION

The emergence of a similar structure in very different systems may indicate that the mechanism governing the $A - R$ relationship is general. To reveal that mechanism we have developed three domain-independent minimal statistical models. Taking a parsimonious approach, we start from the most naive model and progressively

Figure 2.1: Example of efficiency distribution where the two distinct behaviors that are manifested to each side of the point $\eta = 1$ can be appreciated. The data corresponds to the Twitter conversation around the 2015 Spanish General Elections.

consider more sophisticated theories with increasingly complex levels of dependence between R and A . The models are the Independent Variables model (InV), the Identical Actors model (IdA) and the Distinguishable Actors model (DiA). In the InV model the response of the system is independent with respect to the activity of the individual. In the IdA model, the response of the system depends on the activity of the individual, but the system is agnostic with respect to the individual that stimulates it. Finally, in the DiA model the response is determined not only by the activity of the individual, but also by her features. The models are general because no assumption is made about the particular characteristics of the system or its components. Although the first model shows acceptable agreement with the data and explains the universality of the efficiency distribution, we develop the other two models to refine the fit and improve the concordance between model and data. In Figure 2.2 we present a diagram that summarizes the main ideas behind each model and can be used as a guide to follow the explanations of this section.

Although we have applied the models to the three social systems described above, in this section we describe each model and present the results only for the Twitter datasets. This way we can focus on discussing the models instead of the peculiarities of each system. The results for the other two systems are reported and thoroughly analyzed in Section 2.3.

2.2.1 *Methods to obtain the efficiency distribution*

The theoretical distributions of efficiency have been calculated with three different methodologies: Monte-Carlo (MC) simulation, direct computation with discrete probability distributions and derivation of an analytical expression.

Once the basic mechanism of the model is laid out, MC simulation allows a direct implementation of the model's assumptions. Thus, we use it to compare model and empirical data as well as to verify the results of the other methodologies.

Figure 2.2: Diagram that summarizes the main characteristics of the three models: Independent Variables (InV), Identical Actors (IdA) and Distinguishable Actors (DiA).

To directly compute the efficiency distribution with the discrete joint probability distribution $p(R, A)$ we start from the cumulative probability distribution of efficiency:

$$P(\eta \leq \eta_*) = \sum_{A=1}^{\infty} \sum_{R=0}^{\lfloor A\eta_* \rfloor} p(R, A) \quad (2.2)$$

Where $\lfloor \cdot \rfloor$ is the floor operator and η_* a particular value of the random variable η . Then, average probability densities for efficiency intervals $[\eta_a, \eta_b]$ are computed as:

$$\bar{f}(\eta \in [\eta_a, \eta_b]) \approx \frac{P(\eta \leq \eta_b) - P(\eta \leq \eta_a)}{\eta_b - \eta_a} \quad (2.3)$$

Equation 2.3 is used to compare the theoretical distributions obtained with this method with the empirical Probability Density Functions (PDFs) obtained from histograms and the analytical expression of the PDF that will be derived below. The computation is carried out by aligning the extremes of each bin of the empirical histogram with the extremes of the interval $[\eta_a, \eta_b]$.

The resulting efficiency distribution is asymptotically exact in the sense that, since the support for the distributions of A and R is $\mathbb{N}$, an infinite number of terms would be required to actually obtain exact results, but larger values of A and R have increasingly smaller probabilities, carrying progressively lower weight on the computation and enabling the results to converge for a finite number of terms.

Although A and R are discrete in nature (we have an integer number of tweets and retweets, publications and citations, etc.) it is possible to approximate them as continuous random variables and, if the model is not too convoluted, obtain an analytical expression for the efficiency distribution. We have carried out this

computation for the InV and IdA models. In the InV model the computation is straight forward and the analytical expression perfectly matches the simulations, but in the IdA model additional approximations are needed and the agreement is not as good (this will be discussed in Sections 2.2.2 and 2.2.3). The general procedure to obtain the theoretical PDF of efficiency $f(\eta)$ analytically from the continuous joint probability distribution $\varphi(R, A)$ is as follows.

Since $\eta = R/A$, the PDF of efficiency $f(\eta)$ should be easily computed using random variable algebra, as the computation of a ratio distribution is a well known result in statistics [Cur41; Spr79]. However, the fact that A and R have a lower bound $A_m, R_m > 0$ introduce some subtleties that, if not taken into account, would lead us to a wrong result. For the sake of clarity, only for this section we will adopt the usual convention in statistics of writing the name of a random variable with an uppercase letter (X) and a specific value of that variable as the corresponding lowercase letter (x). Therefore, when we write r in this section, we are referring to a specific value of the response random variable R , not to the partial response to a single action.

Let A and R be two random variables with support $a \in [a_m, \infty)$ and $r \in [r_m, \infty)$ for some $a_m, r_m > 0$. Efficiency (H - uppercase η) is another random variable defined as $H = \frac{R}{A}$. Let η be a particular value of H . The cumulative probability that $H \leq \eta$ can be written as:

$$P(H \leq \eta) = P\left(\frac{R}{A} \leq \eta\right) = P(R \leq \eta A) = P\left(A \geq \frac{R}{\eta}\right) \quad (2.4)$$

Where we have replaced H by R/A . Let us now make the approximation that A and R are continuous variables with joint probability distribution $\varphi_{R,A}(r, a)$ (which will be called $\varphi(r, a)$ to simplify notation). We can compute the cumulative distribution function (CDF) of H ; that is, $F_H(\eta) = P(H \leq \eta)$, by integrating $\varphi(r, a)$ in the region where equation (2.4) is fulfilled:

$$F_H(\eta) = \begin{cases} \int_{r_m}^{\infty} \int_{r/\eta}^{\infty} \varphi(r, a) da dr & \text{if } \eta \leq \frac{r_m}{a_m} \\ \int_{a_m}^{\infty} \int_{r_m}^{\eta a} \varphi(r, a) dr da & \text{if } \eta > \frac{r_m}{a_m} \end{cases} \quad (2.5)$$

In Figure 2.3 we show a diagram explaining the regions of integration. Notice that in the two-dimensional (r, a) space, the subspaces of constant efficiency are straight lines that pass through $(0, 0)$ because, for a specific efficiency $\eta_* \neq 0$, we can write $a = \frac{1}{\eta_*} r$. Therefore, lines with lower η_* have higher slope and vice versa. The probability that the efficiency takes a value smaller than a given η_* ; that is, $p(H \leq \eta_*)$, can be found by integrating the probability density $\varphi(r, a)$ to the *left* of the $a = \frac{1}{\eta_*} r$ line (integrate for all the efficiencies $\eta \leq \eta_*$, so lines with higher slope than $1/\eta_*$).

Now, because the minimum values a_m, r_m are larger than 0, the regions of integration must be defined differently to each side of the line with $\eta_o = \frac{r_m}{a_m}$ (thick dashed line of the diagram). The green area is an example of the region of integration associated to a particular efficiency value $\eta_1 < \eta_o$. There, the integral in variable a runs from the line defined by $a = \frac{1}{\eta_1} r$ to infinity and in variable r from r_m to infinity. When we integrate for an efficiency value $\eta_2 > \eta_o$, the region

Figure 2.3: Diagram showing the regions of integration defined to each side of the line $\eta_o = \frac{r_m}{a_m}$ to obtain the cumulative distribution function of the efficiency.

of integration, an example of which is shown in orange, is defined by taking variable r running from r_m to the line $r = \eta_2 a$ and variable a from a_m to infinity. If we used integration regions of the *green kind* in the area where $\eta > \eta_o$, we would incorrectly include the red area in the integral, while if we used integration regions of the *orange kind* in the area where $\eta < \eta_o$, we would incorrectly include the yellow area in the integral. Therefore, we must compute $F_H(\eta)$ differently to each side of $\eta_o = \frac{r_m}{a_m}$ to obtain the right result.

Once we have the cumulative distribution function $F_H(\eta)$, the Leibniz integral rule can be used to get $f_H(\eta) = \frac{dF_H(\eta)}{d\eta}$. For $\eta \leq \frac{r_m}{a_m}$ we have:

$$f_H(\eta) = \frac{d}{d\eta} \left[\int_{r_m}^{\infty} \int_{r/\eta}^{\infty} \varphi(r, a) da dr \right] = \int_{r_m}^{\infty} \varphi\left(r, \frac{r}{\eta}\right) \left(\frac{r}{\eta^2} \right) dr \quad (2.6)$$

$$\text{Change of variables } r = \eta a \Rightarrow f_H(\eta) = \int_{r_m/\eta}^{\infty} \varphi(\eta a, a) a da$$

And for $\eta > \frac{r_m}{a_m}$:

$$f_H(\eta) = \frac{d}{d\eta} \left[\int_{a_m}^{\infty} \int_{r_m}^{\eta a} \varphi(r, a) dr da \right] = \int_{a_m}^{\infty} \varphi(\eta a, a) a da \quad (2.7)$$

Now, if we rewrite 2.6 and 2.7 dropping the uppercase/lowercase convention adopted only for this section (so use R instead of r and A instead of a) the general expression to obtain $f(\eta)$ from $\varphi(R, A)$ is:

$$f(\eta) = \begin{cases} \int_{R_m/\eta}^{\infty} \varphi(\eta A, A) A dA & \text{if } \eta \leq \frac{R_m}{A_m} \\ \int_{A_m}^{\infty} \varphi(\eta A, A) A dA & \text{if } \eta > \frac{R_m}{A_m} \end{cases} \quad (2.8)$$

In our case, $A_m = R_m = 1$ for every dataset. It is worth noting that the two branches of $f(\eta)$ in equation (2.8) correspond to the two characteristic tails of the efficiency distribution.

2.2.2 *Independent Variables: explaining the universality of the efficiency distribution*

2.2.2.1 *InV Model*

The fundamental assumption of the Independent Variables (InV) model is that A and R are independent random variables with probability distributions $p(A)$ and $p(R)$. The InV model is inspired in a previous work [Mor+14] where it was shown that the response distribution for Twitter conversations was independent with respect to the $p(A)$ distribution, so it was hypothesized that R was not affected by A .

The Monte-Carlo simulation of the InV model is computed as follows: In a system with N individuals indexed by $i = 1, 2, \dots, N$, store the empirical data of activity and response in two vectors $\vec{A}$ and $\vec{R}$ such that component i of vector $\vec{A}$ corresponds to the same individual as component i of vector $\vec{R}$. Next, shuffle each of them independently, such that the correlations that may have been present when each couple (A_i, R_i) corresponded to the same individual vanish. The randomized versions of the vectors, $\vec{A}_{rnd}$ and $\vec{R}_{rnd}$, hold the same values as the originals but with the order of the elements randomly altered. Finally, the efficiency vector $\vec{\eta}_{rnd} = \vec{R}_{rnd} / \vec{A}_{rnd}$ is used to compute the efficiency distribution according to the InV model.

To compute the efficiency distribution directly with the discrete joint probability distribution, we take into account that A and R are considered independent, so their joint probability distribution is $p(R, A) = p(R)p(A)$. The PDF of efficiency can be obtained by plugging this expression in equations (2.2) and (2.3). However, for this model we have left out the results of the discrete methodology because the analytical expression that will be derived below is exact¹.

For the analytical computation of the efficiency distribution we consider A and R as continuous variables with PDFs $f_A(A)$ and $f_R(R)$. Again, due to their independence, the continuous joint probability distribution can be written as $\varphi(R, A) = f_A(A)f_R(R)$. Plugging this in equation (2.8) we obtain:

$$f^{InV}(\eta) = \begin{cases} \int_{R_m/\eta}^{\infty} f_R(\eta A) f_A(A) A dA & \text{if } \eta \leq \frac{R_m}{A_m} \\ \int_{A_m}^{\infty} f_R(\eta A) f_A(A) A dA & \text{if } \eta > \frac{R_m}{A_m} \end{cases} \quad (2.9)$$

¹ Leaving aside the approximation of considering A and R as continuous.

This expression provides an explanation for the previous empirical result [Mor+14] that the right tail of the efficiency distribution remains unaltered when the activity distribution is modified. To reach that result, let us assume that $f_R(R) \propto R^{-\gamma_R}$. This power law distribution was used in [Mor+14] as well as in other works to model the distribution of retweets [Bor+12b], scientific citations [CSN09] and incoming editions in Wikipedia [Muc+13]. Then, the right tail ($\eta > \frac{R_m}{A_m}$) of the PDF shown in equation (2.9) can be written as:

$$\begin{aligned} f^{\text{InV}}(\eta) &\propto \eta^{-\gamma_R} \int_{A_m}^{\infty} A^{1-\gamma_R} f_A(A) dA \\ &= \mathbb{E}_A[A^{1-\gamma_R}] \eta^{-\gamma_R} \\ &\propto f_R(\eta) \end{aligned} \quad (2.10)$$

Where $\mathbb{E}_A[\cdot]$ is the expected value with respect to the activity distribution, i.e., a constant. Therefore, when $f_R(R) \propto R^{-\gamma_R}$, the right tail of the efficiency distribution is proportional to $\eta^{-\gamma_R}$. That is, in addition to being independent of the activity distribution, its shape is completely determined by the exponent of the response distribution.

To apply the analytical computation of the efficiency distribution for the InV model to empirical data we have fit the empirical distributions of A and R to a power law with exponential cutoff (or truncated power law) using the powerlaw python module [ABP14]. The functional form of this distribution is the following:

$$f(x) = \frac{\lambda^{1-\alpha}}{\Gamma(1-\alpha, \lambda x_{\min})} x^{-\alpha} e^{-\lambda x} \quad (2.11)$$

Where $\Gamma(s, x)$ is the upper incomplete gamma function. When the PDFs of activity and response are power laws with exponential cutoff, the PDF of efficiency adopts the following form:

$$f^{\text{InV}}(\eta) = \begin{cases} g(\eta) \Gamma(2 - \alpha_R - \alpha_A, (\lambda_R \eta + \lambda_A) \frac{R_m}{\eta}) & \text{if } \eta \leq \frac{R_m}{A_m} \\ g(\eta) \Gamma(2 - \alpha_R - \alpha_A, (\lambda_R \eta + \lambda_A) A_m) & \text{if } \eta > \frac{R_m}{A_m} \end{cases} \quad (2.12)$$

With

$$g(\eta) = C (\lambda_R \eta + \lambda_A)^{(\alpha_R + \alpha_A - 2)} \eta^{-\alpha_R} \quad (2.13)$$

and

$$C = \frac{\lambda_R^{1-\alpha_R}}{\Gamma(1-\alpha_R, \lambda_R R_m)} \frac{\lambda_A^{1-\alpha_A}}{\Gamma(1-\alpha_A, \lambda_A A_m)} \quad (2.14)$$

2.2.2.2 Results of the InV model applied to Twitter datasets

In Figure 2.4 we present the results of the MC simulation (orange squares) and the analytical calculation (green line) of the efficiency distribution according to the InV model applied to Twitter datasets together with the empirical distribution (blue dots). The data used consists of Twitter messages retrieved with the Twitter Streaming API (see Appendices B.1 and C for more details). It can be seen that the InV model captures the general shape of the Twitter efficiency distributions. This result, together with the derivation in equation (2.10) of the independence of the right tail of the efficiency distribution with respect to $p(A)$, shows that a very plausible explanation for the efficiency distribution's universality is that the efficiency is the ratio of two independent random variables (R and A).

Figure 2.4: Efficiency distributions corresponding to the InV model applied to the Twitter datasets. The plots show the empirical efficiency distribution (blue dots), the Monte-Carlo simulation (orange squares) and the analytical expression of equation (2.12) (green line). See Appendix D for the results of the other 10 datasets.

However, on closer examination we see that in some cases the model does not fully agree with the empirical data, especially when considering the right tails (see Spanish elections 2015 and Argentinian retirement plans), although for others the concordance is quite good (see Madrid-Barcelona and PSOE primary). On the other hand, the analytical expression of equation (2.12) shows a very good agreement with the Monte-Carlo computation in every case.

The small discrepancies found between the model and the empirical data may be due to the fact that the independence assumption is not fully adequate. To verify

this hypothesis, we check if the $A - R$ correlations neglected by the independence assumption affect the quality of the fit given by the InV model. To do this we first compute the empirical Spearman's rank correlation ρ_e [CF14] between R and A . Secondly, we measure the discrepancy Δ_{InV} between the InV model and the data. Finally, we test if ρ_e and Δ_{InV} are positively correlated.

As can be seen in Figure 2.4, the disagreement between the InV model and the data mostly lie on the extreme values of the right tail of the efficiency distribution, which have very low probabilities. Since the right tail can be approximated by a power law, we define Δ_{InV} as the difference between the theoretical exponent ($\gamma_{\text{InV}} = \gamma_R$ according to equation (2.10)) and the empirical exponent (γ_{emp}) of that power law. Thus,

$$\Delta_{\text{InV}} = \gamma_{\text{emp}} - \gamma_{\text{InV}} = \gamma_{\text{emp}} - \gamma_R \quad (2.15)$$

We would like to remark that, since the power law exponent reflects the global trend including the effect of the tail, the contributions to the error of low probability values are not underestimated.

We have computed Δ_{InV} and ρ_e for each of the 14 Twitter datasets and measured the relationship between the two variables by means of a linear regression. In Figure 2.5 each blue dot corresponds to one Twitter dataset and the linear regression is shown as a black line. The positive Pearson correlation of the regression ($r = 0.80$) indicates that Δ_{InV} increases monotonously with ρ_e and corroborates the hypothesis presented above.

Figure 2.5: The discrepancy of the InV model with respect to the Twitter empirical data (Δ_{InV} , computed with equation (2.15)) is related to the Spearman rank correlation (ρ_e) between A and R . The figure shows a monotonous increase of Δ_{InV} with ρ_e , which has been characterized by means of a linear regression.

2.2.3 *Identical Actors: introducing a parsimonious dependency between R and A*

2.2.3.1 *IdA Model*

Since the $A - R$ correlations seem to be the cause of the discrepancy between the data and the InV model, the next logical step is to develop a new model where the response of the system depends on the activity of the individual. To do this in a parsimonious way, we realize that the stimuli to which the system reacts occur in a discrete fashion, so we can assume that it reacts to each action (a tweet, a scientific publication, an edition on Wikipedia, etc.) individually, as if they were isolated events. Then, while in the InV model the *aggregate* response of the system was independent of the aggregate activity of the actor, in the Identical Actors (IdA) model the *partial* response of the system to each single action is independent of the actor. But, as the aggregate response of the system to the activity of an individual is the sum of the partial responses to each of her A actions, a dependence between R and A is induced.

To formalize this idea we introduce the new variable r as the response of the system to a single action by any individual. In the IdA model, this random variable has the same distribution $p(r)$ for all the actors. The aggregate response R associated to an actor that performed A actions and triggered partial responses $\{r_1, r_2, \dots, r_A\}$ is $R = \sum_{j=1}^A r_j$. The dependence of R on A resides precisely on the number of terms of this sum.

For the Monte-Carlo simulation of the IdA model we first fit the $p(r)$ with the hybrid methodology detailed in Appendix E.1 and $p(A)$ to a discrete truncated power law (see Appendix E for the results). Then, we generate a set of individuals whose activity is assigned according to $p(A)$. The responses for each of the A actions of an individual is randomly generated with $p(r)$ and then aggregated to obtain her R . The efficiency according to this model is directly computed from the (R, A) tuple associated to each actor.

To get the efficiency distribution for the IdA model from the discrete $p(R, A)$ distribution, we start with the conditional discrete probability distribution of R given an activity A , which is computed as the $A - \text{fold}$ discrete convolution of $p(r)$ with itself:

$$p(R|A) = p(r_1) * p(r_2) * \dots * p(r_A) = p(r) * p(r) * \dots * p(r) = p^{*A}(r) \quad (2.16)$$

Then, the joint probability distribution can be obtained as:

$$p(R, A) = p(R|A)p(A) = p^{*A}(r)p(A) \quad (2.17)$$

The efficiency PDF is finally computed by plugging expression (2.17) in equations (2.2) and (2.3). The $p(r)$ and $p(A)$ distributions used in this methodology are the same as those used in the Monte-Carlo simulations.

To carry out the previous computations with infinite precision we would need an infinite number of values for the $p(r)$, $p(A)$ and $p(R, A)$ distributions. To be able to perform the numerical computations, we have used distributions that are bounded at a certain value and we have verified that further increasing the number

of values employed do not affect the results. The cut-off values used for the three systems considered are shown in table 2.1.

SYSTEM	$r_{\max}$	$A_{\max}$	$R_{\max}$
Twitter	10^6	3×10^4	5×10^4
Wikipedia	3×10^5	3×10^4	5×10^4
Citations	3×10^5	2×10^4	3×10^4

Table 2.1: Cut-off values used to perform the numerical computations for the IdA model.

To obtain the analytical expression of the efficiency distribution for the IdA model we first need to find an expression for the sum of A independent random variables $\{r_1, r_2, \dots, r_A\}$ following the same distribution $p(r)$; that is, an expression for the A -fold convolution of $p(r)$ with itself.

Although finding an analytical expression for the sum of random variables is not always feasible, there exist approximations that work in some cases. We have modeled $p(r)$ as a power law and adopted an approximation proposed by Zaliapin et al. [ZKS05] to obtain the distribution of the sum of power-law distributed random variables. This approximation consists in replacing the sum by its larger term $R = \sum_{i=1}^A r_i \approx \max(\{r_i\}_{i=1}^A)$. The reason to do that resides in the heavy-tailed nature of the $p(r)$ distribution: when we consider a set of partial responses r , we can expect to have many small values and a few large ones, and the difference between the small and the large values can be so great that even if the small ones are summed, the largest value will still, with high probability, be orders of magnitude higher. Thus, the sum can be replaced by the maximum term with small error. In that case, if the PDF of r is:

$$f_r(r) = (\gamma_r - 1)r_m^{\gamma_r-1}r^{-\gamma_r} \quad (2.18)$$

The conditional cumulative distribution function of R given A can be approximated by:

$$F(R|A) \propto e^{-(R^{1-\gamma_r}A)} \quad (2.19)$$

Then, if the distribution of A is modeled as a power law [Muc+13; Bor+12b] with the form $p(A) \propto A^{\gamma_A}$ (see Appendix E for the fits) and take into account that the minimum values of activity and response are $A_m = 1$ and $R_m = 1$ (which is what is observed for every dataset), the approximated joint probability density is:

$$\varphi(R, A) = \frac{(\gamma_r - 1)(\gamma_A - 1)}{1 + (1 - \gamma_A)\Gamma(1 - \gamma_A, 1)} A^{1-\gamma_A} R^{-\gamma_r} e^{-AR^{1-\gamma_r}} \quad (2.20)$$

Plugging this in equation (2.8), we obtain the following expression:

$$f^{\text{IdA}}(\eta) = \begin{cases} \begin{cases} C \frac{1}{2-\gamma_r} \eta^{\gamma_A-3} E_{\frac{\gamma_A-1}{2-\gamma_r}}(\eta^{-1}) & \text{if } \eta \leq 1 \\ C \frac{1}{2-\gamma_r} \eta^{-\gamma_r} E_{\frac{\gamma_A-1}{2-\gamma_r}}(\eta^{1-\gamma_r}) & \text{if } \eta > 1 \end{cases} & \text{when } \gamma_r < 2 \\ \begin{cases} C \eta^{\left(-1 - \frac{(-1-\gamma_r)(1-\gamma_A)}{2-\gamma_r}\right)} \frac{1}{\gamma_r-2} \Upsilon\left(\frac{3-\gamma_r-\gamma_A}{2-\gamma_r}, \eta^{-1}\right) & \text{if } \eta \leq 1 \\ C \eta^{\left(-1 - \frac{(-1-\gamma_r)(1-\gamma_A)}{2-\gamma_r}\right)} \frac{1}{\gamma_r-2} \Upsilon\left(\frac{3-\gamma_r-\gamma_A}{2-\gamma_r}, \eta^{1-\gamma_r}\right) & \text{if } \eta > 1 \end{cases} & \text{when } \gamma_r > 2 \end{cases} \quad (2.21)$$

Where $E_n(\cdot)$ is the generalized exponential integral, $\Upsilon(s, x)$ the lower incomplete gamma function and C the following normalization constant:

$$C = \frac{(\gamma_r - 1)(\gamma_A - 1)}{1 + (1 - \gamma_A)\Gamma(1 - \gamma_A, 1)} \quad (2.22)$$

In principle, the expression of equation (2.19) proposed in [ZKS05] is valid when $\gamma_r < 2$. However, we have found that the expression of equation (2.21) also provides a good fit for the data when $\gamma_r > 2$, although the agreement quickly deteriorates when γ_r exceeds a value of around 2.5 (see Appendix E.3.4 for a thorough discussion).

2.2.3.2 Results of the IdA model applied to Twitter datasets

In Figure 2.6 we present the empirical distribution of efficiency for the Twitter datasets (blue dots) and the theoretical distribution according to the IdA model computed by MC simulation (orange squares), with the discrete joint probability distribution $p(R, A)$ (dashed red line) and with the analytical approximation of equation (2.21) (green line). A clear improvement in the agreement between theory and data can be appreciated on the right tails of all the datasets. However, there is a higher discrepancy in the left tail of the distribution, which will be discussed below. If we compare the results of the model according to the analytical expression given by equation (2.21) with the other computational methods, we find a good correspondence for some datasets (like PSOE primary), while others present discrepancies that can be explained by the approximated power-law fit of $p(r)$ (see Appendix E.3.4 for more details).

In the IdA model, contrasting with the InV model, R and A are correlated, so we would like to verify that the correlations generated by the model match the empirical ones. To this end we have generated 300 MC simulations of the IdA model for each Twitter dataset and for every simulation we have computed the corresponding $A - R$ correlation (ρ_t). Then, we have computed the mean and the standard deviation of the 300 MC simulations to get a confidence interval for the theoretical correlations.

In Figure 2.7 we have plotted (as blue dots) the theoretical Spearman's correlation according to the IdA model (ρ_t) against the empirical one (ρ_e) for the Twitter datasets and carried out a linear regression between both magnitudes. As it can

Figure 2.6: Efficiency distributions corresponding to the IdA model applied to the Twitter datasets. The plots show the empirical efficiency distribution (blue dots), the Monte-Carlo simulation (orange squares), the discrete computation (dashed red line) and the analytical expression of equation (2.21) (green line). See Appendix D for the results of the other 10 datasets.

be appreciated, there is a significant correlation ($r^2 = 0.88$) between ρ_t and ρ_e . Moreover, the slope is not far from 1 and the value of the intercept is close to 0, corroborating the ability of the IdA model to reproduce the correlations of the real data to a reasonable extent. For comparison the values corresponding to the InV model, where R and A are not correlated, have also been plotted (as orange squares).

Figure 2.7: Linear regression of the correlation induced by the IdA model (blue dots) with respect to the empirical correlation computed for the Twitter datasets. Each point corresponds to a different conversation. For comparison the values corresponding to the InV model, where R and A are not correlated, have also been plotted (orange squares).

Summarizing, there is an excellent agreement between the IdA model and the data in the right tail of the efficiency distribution. Besides, the correlations induced by the model present a high correspondence with the empirical correlations between A and R . However, the left tail of the efficiency distribution falls faster in the model than in the data. A low efficiency implies that an individual performs many actions obtaining a very low aggregate response (high A and low R), but under the IdA model, which considers that *every individual and every action have identical ability* to trigger a response on the system, low efficiencies are very unlikely, as performing many individual actions guarantee at least a moderate aggregate response.

Rather, the empirical evidence suggests that individuals from the left tail of the efficiency distribution ($\eta < 1$) have lower capabilities to trigger a response than those of the right tail ($\eta > 1$). This is backed by the fact that the InV model, which considers response independent from activity, reproduces the left tail of the efficiency distribution with better accuracy than the IdA model. Therefore, we hypothesize that the response associated to actors with lower efficiencies do not depend on their activity (so they should show low $A - R$ correlation), contrasting with the users with higher efficiencies, whose behavior can be characterized with the IdA model and should present high $A - R$ correlation.

The previous hypothesis can be verified by comparing the Spearman's rank correlation between A and R for individuals with $\eta < 1$ and $\eta > 1$. However, we should be careful and factor out the artificial correlations induced by performing

this filtering. This is achieved by subtracting the correlation associated to the InV model (averaged from 300 realizations). We have performed this computation for the Twitter datasets and, as can be appreciated in Figure 2.8, there is a clear difference between both sets of individuals.

Figure 2.8: Histogram of the Spearman's rank correlation between A and R for the Twitter datasets computed for individuals with low efficiency ($\eta < 1$) and high efficiency ($\eta \geq 1$). The small vertical lines in the bottom of the horizontal axis are a rug plot and represent the value of $\rho_e - \rho_{InV}$ for each individual dataset.

2.2.4 Distinguishable Actors: letting the system discriminate between individuals

2.2.4.1 DiA Model

As discussed above, on the one hand, actors with low efficiency have a small impact on the system. Therefore, the InV model, where the interaction between system and individual is *weak* (because the system's response is not influenced by the individual's activity) explains their distribution better. On the other hand, actors with high efficiency have a greater impact on the system, so it is the IdA model, where the interaction between individual and system is *strong*, the one that explains their distribution better. Since two models with different assumptions explain different intervals of the efficiency distribution and users with different efficiencies also present different levels of correlation between A and R , we realize that the response of the system should depend on the features of each actor. Thus, we have developed the Distinguishable Actors (DiA) model, where the system is sensitive to the individual who makes the action and reacts in a different manner depending on her characteristics.

This idea can be formalized by considering that the probability distribution of response to single actions depends on the features of the individual that performs the action, summarized in a vector $\vec{s}$. The distribution of aggregate response R of

the system is computed as the A — fold convolution of the $p(r|\vec{s})$ distribution with itself:

$$p(R|A, \vec{s}) = p^{*A}(r|\vec{s}) \quad (2.23)$$

If $\{s_1, s_2, \dots, s_N\}$ are the components of the feature vector (assume the features are independent discrete variables), the discrete joint probability distribution $p(R, A)$ is obtained as follows:

$$p(R, A) = \sum_{s_1} \dots \sum_{s_N} p^{*A}(r|\vec{s})p(A)p(\vec{s}) \quad (2.24)$$

Finally, $p(R, A)$ can be used to compute the efficiency distribution with equations (2.2) and (2.3).

A key point of the implementation of the DiA model is to find the conditional probability distribution $p(r|\vec{s})$ that characterizes the relationship between the features $\vec{s}$ of the individual and the response r of the system to her actions. Although this task is not trivial in most cases, in Twitter the number of followers F of a user can be considered as a proxy of her ability to trigger a response, since the follower layer is the substrate through which the retweets are spread [Bak+11; Bor+15]. Therefore, we have adopted F as the characterizing feature of the actors.

To implement these ideas and derive an expression for $p(r|\vec{s})$, we have formalized the Independent Cascade (IC) model in terms of probability distributions. Due to the reasons stated above, agent-based simulations of the IC model have been used in previous works [Mor+14] to reproduce the Twitter dynamics using the follower network as the diffusion substrate.

The IC model consists in a diffusion mechanism [KKT03] that can be outlined as follows. Assume that we have a set of nodes that are connected with a set of links forming a directed network. Choose one node randomly to be activated. The active node tries to activate each of her neighbors one by one. There is only one attempt per neighbor and all the attempts are independent. The success probability for each activation attempt depends on the link that joins the source node and the target node. The successfully *infected* nodes become active and try to activate their neighbors. This process is iterated until the diffusion cascade dies out.

Now, drawing on empirical results [Mor+14], let us assume that the probability of a given node to activate her neighbors only depends on her distance to the source of the cascade. If she is the source, the *infection* probability will be p_{inf}^1 . If she is l layers away from the source, the infection probability will be p_{inf}^l . The number of *neighbors* of a node is her out-degree k_{out}^l (considering the direction of the links as the direction of propagation of influence, i.e., from followee to follower in the case of the follower network), which we will call k^l in order to simplify the notation.

Taking the above into account, let us try to find the response distribution conditioned on the activity. Notice that if we look at a given active node trying to *infect* her neighbors we can model this process with a binomial variable: In the first layer, the node has k^1 Bernoulli trials to activate a neighbor and each trial has a success probability p_{inf}^1 . The number of successfully activated neighbors in that

layer corresponds to the response r^1 , which can take values between 0 and k^1 . The response distribution for the first layer of the diffusion cascade would be:

$$p(r^1|k^1) = B(r^1; k^1, p_{\text{inf}}^1) \quad (2.25)$$

With $B(x; n, p)$ a distribution of a binomial variable x with n Bernoulli trials and p success probability.

An individual that performs A actions and has k^1 neighbors will trigger an aggregated response $R^1 = \sum_{i=1}^A r_i^1$ in her immediate neighbors (the first layer). The distribution of the aggregated response can be computed straight forward, as the sum of two binomial variables ($y = x_1 + x_2$) with n_1 and n_2 trials and same success probability p also follows a binomial distribution of the form $B(y; n_1 + n_2, p)$. The probability distribution for R^1 given a specific activity A , an out-degree k^1 and an infection probability p_{inf}^1 is then:

$$p(R^1|A, k^1) = B(R^1; Ak^1, p_{\text{inf}}^1) \quad (2.26)$$

Our next approximation is that a response of R^1 implies that R^1 nodes were activated (the reality is that there were R^1 activations distributed among k^1 nodes, some of them possibly activated more than once). The R^1 active nodes have $k_1^2, k_2^2, \dots, k_{R^1}^2$ neighbors (the superscript 2 indicates that they are in the second diffusion layer). Each of the active nodes tries to *infect* her neighbors with the same infection probability p_{inf}^2 . Hence, there are a total of $k_T^2(R^1) = \sum_{i=1}^{R^1} k_i^2$ Bernoulli trials with success probability p_{inf}^2 , resulting in the following distribution of aggregated response for the second diffusion layer:

$$p(R^2|R^1, k_T^2(R^1)) = B(R^2; k_T^2(R^1), p_{\text{inf}}^2) \quad (2.27)$$

We can marginalize this with respect to R^1 and $k_T^2(R^1)$ in order to get the distribution of R^2 , but first we need to obtain the distribution of $k_T^2(R^1)$, which can be computed as the R^1 – fold convolution of the empirical $p(k)$ distribution with itself:

$$p(k_T^2|R^1) = p^{*R^1}(k) \quad (2.28)$$

In order to be able to do this, we have to assume that $p(k^{l+1}|k^l) = p(k)$; that is, that there exists no degree correlations. This is a reasonable approximation for Twitter because it has been observed that the degree correlations are small ($|\rho| < 0.2$) [Bil+15; MGLB18; Bor+12b], a property that is also present in other online social networks [HW09].

The probability distribution of R^2 is finally:

$$\begin{aligned}
 p(R^2|A, k^1) &= \sum_{k_T^2=0}^{\infty} \sum_{R^1=0}^{\infty} p(R^2|R^1, k_T^2(R^1))p(k_T^2|R^1)p(R^1|A, k^1) \\
 &= \sum_{k_T^2=0}^{\infty} \sum_{R^1=0}^{\infty} B(R^2; k_T^2(R^1), p_{inf}^2) p^{*R^1}(k) B(R^1; A k^1, p_{inf}^1)
 \end{aligned} \tag{2.29}$$

This process can be iterated to get the response distribution for layer l as a function of the one obtained in layer $l-1$:

$$\begin{aligned}
 p(R^l|A, k^1) &= \sum_{k_T^l=0}^{\infty} \sum_{R^{l-1}=0}^{\infty} p(R^l|R^{l-1}, k_T^l(R^{l-1}))p(k_T^l|R^{l-1})p(R^{l-1}|A, k^1) \\
 &= \sum_{k_T^l=0}^{\infty} \sum_{R^{l-1}=0}^{\infty} B(R^l; k_T^l(R^{l-1})) p^{*R^{l-1}}(k) p(R^{l-1}|A, k^1)
 \end{aligned} \tag{2.30}$$

The total aggregated response triggered in all the layers would be $R = \sum_{l=1}^{\infty} R^l$. Hence, the distribution of the response given an activity A and a number of neighbors k^1 is the convolution of the distributions of response for each layer:

$$p(R|A, k^1) = p(R^1|A, k^1) * p(R^2|A, k^1) * \dots \tag{2.31}$$

Finally, the joint probability distribution for R and A can be obtained as:

$$p(R, A) = p(A) \sum_{k^1=0}^{\infty} p(R|A, k^1) p(k^1) \tag{2.32}$$

Notice that $p(k^1) = p(k)$. In this model, the infection probabilities $\{p_{inf}^l\}_{l=1}^{\infty}$ are free parameters that must be determined in some way, be it by direct computation if the data is available, by Maximum Likelihood Estimation or any other fitting technique.

The computation of the complete IC model formalized in this way comprehend an infinite number of diffusion layers, making it numerically intractable. Taking into account that most information cascades in Twitter are shallow [Kwa+10; Bak+11], we have performed a first-neighbors approximation and focused on the response generated only in the first layer of diffusion. Moreover, the computations involved for the second and farther layers are very demanding in terms of memory and computation time, and the results that we obtain for the first-layer approximation are already in good agreement with the data.

Since we work with the follower network of Twitter, the out-degree distribution $p(k)$ corresponds to the follower distribution $p(F)$. Hence, for our first-neighbors

approximation to the independent cascade model, the distribution of response R given the activity A and a certain number of followers F is:

$$p(R|A, F) = B(R; AF, p_{\text{inf}}) \quad (2.33)$$

Notice that p_{inf} can be considered as an effective infection probability that includes the effect of the higher layers. We have determined p_{inf} by fitting the empirical distribution of response to single actions $p(r)$ to the corresponding theoretical distribution for this model with MLE (see Appendix E.4):

$$p(r) = \sum_{F=0}^{\infty} B(r; F, p_{\text{inf}})p(F) \quad (2.34)$$

Where $p(F)$ is the empirical follower distribution. Once the p_{inf} has been determined it is straight forward to compute the joint probability distribution $p(R, A)$:

$$p(R, A) = p(A) \sum_{F=0}^{\infty} B(R; AF, p_{\text{inf}})p(F) \quad (2.35)$$

Finally, the efficiency distribution can be determined with equations (2.2) and (2.3). The numerical computations for the DiA model have been carried out with Twitter data considering the following cut-off values for the followers, the activity and the response: $F_{\text{max}} = 500000$; $A_{\text{max}} = 1000$; $R_{\text{max}} = 1000$.

Notice that F is the only component of the feature vector $\vec{s}$ of the individual. The discrete computation of the DiA model requires a fit for the $p(F)$ distribution, which has been performed with the hybrid methodology detailed in Appendix E.1. The $p(A)$ was fit to a discrete truncated power law.

A Monte-Carlo simulation of the DiA model can be performed as follows: Generate a set of individuals with a random number of followers $F \sim p(F)$ and a random activity $A \sim p(A)$. Then, for each action $j = 1, 2, \dots, A$ performed by an individual, the partial response of the system r_j is computed as:

$$p(r|\vec{s}) = p(r|F) = B(r; F, p_{\text{inf}}) \quad (2.36)$$

And the aggregate response with $R = \sum_{j=1}^A r_j$. For this model, we have found that an analytical derivation of the PDF of efficiency is too cumbersome to be tackled.

2.2.4.2 Results of the DiA model applied to Twitter datasets

In Figure 2.9 we present the results of applying the DiA model to the Twitter datasets by considering the number of followers as the feature that characterizes each actor. It can be appreciated that the DiA model behaves as expected and is able to reproduce both branches of the efficiency distribution for every Twitter dataset.

Figure 2.9: Efficiency distributions corresponding to the DiA model applied to the Twitter datasets. The plots show the empirical efficiency distribution (blue dots), the Monte-Carlo simulation (orange squares) and the discrete computation (dashed red line). See Appendix D for the results of the other 10 datasets.

The correlations that the DiA model induces between A and R (ρ_t) have been represented *vs* the empirical correlations (ρ_e) in Figure 2.10. The linear regression that is included in the aforementioned figure shows that, analogously to the IdA model, there exists a significant correlation between empirical and theoretical results. Additionally, in the case of the DiA model the slope is closer to 1 and, more importantly, consistent with that value according to the standard error. Hence, the DiA model improves the results of the IdA model and reproduces the correlations between A and R more faithfully.

Figure 2.10: Linear regression of the correlation induced by the DiA model with respect to the empirical correlation computed for the Twitter datasets. Each point corresponds to a different conversation. The reference values obtained for the InV model are also displayed.

The previous results confirm that the DiA model shows a very good agreement between theory and data both for the efficiency distribution and for the $A - R$ correlations. We conclude that in the case of Twitter it is necessary to consider the actors distinguishable to correctly describe the relationship between A and R .

2.3 APPLICATION TO WIKIPEDIA AND THE SCIENTIFIC CITATIONS NETWORK

In this section we present and discuss the results of applying the InV and IdA models to Wikipedia and the scientific citations network.

2.3.1 Wikipedia

Wikipedia is an online encyclopedia whose articles are collaboratively created and edited by a community of users. For the Wikipedia analysis we have worked with metadata of the editions made on articles of the English Wikipedia (see Appendix B.3 for more details). Let us recall that in this system, the activity A is the number of editions made by a Wikipedia user and the response R , the number of messages received by that user.

In Figures 2.11 and 2.12 we show the theoretical efficiency distributions for the InV and the IdA models, respectively. The MC simulations are shown as orange squares, the discrete computation as a red dashed line and the analytical calculation (exact in the case of the InV model and approximated for the IdA model), as a continuous green line. The empirical data are shown as blue dots.

We see in Figure 2.11 that the general shape of the distribution is correctly captured by the InV model and the MC simulation coincides with the analytical computation. However, while a partial agreement between theory and empirical data can be appreciated in the left tail, there are discrepancies elsewhere, especially in the far ends of both tails.

Figure 2.11: Same as Figure 2.4 for the InV model applied to the Wikipedia dataset.

With respect to the IdA model, in Figure 2.12 we can appreciate a clear improvement of the correspondence between model and data in the right tail. However, analogously to Twitter, the agreement in the left tail is slightly better for the InV model. The discrete computation perfectly matches the MC simulation, as expected, and the analytical expression coincides with them in the right tail, but not in the left, indicating that the rough approximations used for this calculation have a limited applicability (more details can be found in Appendix E.3.4).

2.3.2 Scientific citations network

The scientific citations network emerges from the references included in scientific papers to previous works. To analyze it we have used citations data of authors from 14 different countries retrieved from the Web of Science [Cla; Oli+18]. As a reminder, in this context activity A is defined as the number of publications of an author and response R , as the number of citations she obtains. Notice that, although the authors are associated to a given country, the citations they receive may come from any author in any country (more details in Appendix B.2).

Figures 2.13 and 2.14 show respectively the theoretical efficiency distributions according to the InV and IdA models. In these figures we can see that, while the IdA model improved the results for Twitter and Wikipedia by increasing the agreement in the right tail, in this system the IdA model is clearly outperformed by the InV

Figure 2.12: Same as Figure 2.6 for the IdA model applied to the Wikipedia dataset.

model in both tails. The InV model not only reproduces the general shape of the distribution, but also coincides remarkably well with the empirical data. Therefore, the scientific citations network is better characterized by a model in which activity and response are considered independent.

Figure 2.13: Same as Figure 2.4 for the InV model applied to the scientific citations datasets. See Appendix D for the results of the other 10 datasets.

Figure 2.14: Same as Figure 2.6 for the IdA model applied to the scientific citations datasets. See Appendix D for the results of the other 10 datasets.

This result may be caused by the long reaction times of citations and prolonged lifespan of scientific publications (months or years) [CL91], which contrast with the short characteristic times of the interactions in social media (in the order of minutes or hours) [Zha+15; KL16; Yas+12]. Another important factor to take into account is that while the actions (and the corresponding reactions) in the case of Twitter and Wikipedia are associated to a specific individual because a tweet or an edition have one author (who is the recipient of the retweets and incoming messages), a scientific paper usually has several authors and its associated citations are assigned to all of them. Moreover, there are many and varied circumstances that determine the citation counts of publications. As a result, the independence assumption may work for the scientific citations datasets because the different overlapping factors discussed above mask the dependence between R and A .

It should be noticed that in this section we have applied the InV and IdA models to Wikipedia and the scientific citations network, but not the DiA model. The reason to leave out the Distinguishable Actors model is that the identification of the set of features that characterizes the ability of an actor to trigger reactions on these two systems is not straight forward and requires further research that fall beyond the scope of this work. In the case of the citation network the literature shows that the citation counts of publications [Bor+12a] are conditioned by multiple aspects, like the quality of the manuscript, the field of research, the cited references or the

reputation of the authors and their institutions. With respect to Wikipedia, some factors that could determine the response to a user could be the topics she is more active on, the age of her user account or her main role (some users may be focused on editing articles, others on moderating discussion pages, etc.).

2.4 DISCUSSION

In this chapter we have measured the relationship between individual actions (A) and the collective response of social systems (R) with the probability distribution of the efficiency metric. The efficiency distribution is known to have a universal shape for Twitter conversations, but we have shown that this universality can be extended to two more systems of different nature: the Wikipedia collaboration environment and the scientific citations network.

We have reproduced the general shape of the efficiency distribution with the Independent Variables (InV) model, a parsimonious theory that considers activity and response as independent random variables. Consequently, the universality of the efficiency distribution can be explained by the efficiency being the ratio of two random variables with heterogeneous distributions that can be considered independent.

However, we found small discrepancies between model and data in the right tail of the efficiency distribution and showed that they are caused by the $A - R$ correlations that are neglected by the InV model. To correct those discrepancies, we developed the Identical Actors (IdA) model where, although all individuals are considered identical, the collective response of the system R depends on the number of actions A performed by an actor, inducing correlations between the two variables. The IdA model shows an excellent agreement with data in the right tail of the efficiency distribution and reproduces the empirical correlations between A and R to a reasonable extent. However, the InV model performs better than the IdA model in the left tail, as the latter underestimates the number of low efficiency actors.

Summarizing, the low efficiency region of the distribution is reproduced better by a model that neglects the correlations between A and R whereas the high efficiency region is reproduced better by a model that takes into account those correlations. This result prompted us to develop the Distinguishable Actors (DiA) model, in which the system's response depends both on the characteristics of the individual that performs the action and on the number of actions. Thus, the DiA model can explain the existence of both inefficient actors that generate small responses and influential or efficient actors that trigger large reactions, as well as all the possible intermediate situations. This flexibility enables the DiA model to adequately fit both tails of the empirical efficiency distribution for Twitter datasets and faithfully reproduce the empirical correlations between R and A .

To conclude, let us discuss the potential domains of application of each model. Taking into account its characteristics, the InV model is more appropriate for systems where R does not depend on A . Nevertheless, its application may be further extended to systems where the relationship between R and A is so complex and depend on so many counteracting factors that the two variables can be considered independent. As discussed above, this may be the case of the scientific citations networks. The InV model may also be useful when the analysis is focused on low

efficiency actors, as their $A - R$ correlations are low enough to be neglected. When the effect of $A - R$ correlations is important and/or the interest is placed on highly efficient actors, the IdA model tends to do a better work at modeling these social dynamics. However, the IdA model is not suitable for the low efficiency region. In Twitter and Wikipedia we have seen examples of this behavior. The DiA model presents the most accurate results and may be useful in cases where the ability of individuals to trigger reactions can be characterized by one or several well chosen metrics. This is the case of Twitter, where the influence of an individual strongly depends on its position in the follower network, as it is the substrate layer for the communication between users [Bor+15].

From the discussion above it is clear that, although the particular characteristics of a social system are relevant to achieve a detailed reproduction of its dynamics, by approximating individual activity and collective response as independent variables we can explain general properties of those systems, like the universality of the efficiency distribution.

USER BEHAVIOR DURING THE SPANISH GENERAL ELECTIONS OF 2015 AND 2016 IN TWITTER

*[...] despite the spontaneity you may exhibit,
you are far more predictable than you are willing to admit [...]
Forget dice rolling or boxes of chocolates as metaphors for life.
Think of yourself as a dreaming robot on autopilot,
and you'll be much closer to the truth.*

— Albert-László Barabási - *Bursts*

The kind of elementary interactions studied in the previous chapter can produce complex emergent phenomena. Using data from social networking sites (SNS), many research works have been devoted to reveal these emergent patterns in our collective behavior and use them to gain insight about relevant social topics like economy [Llo+15], marketing [AH10], politics [Tum+10; Bor+12b], mobility [Haw+14], segregation [Don+20; Mor+19] or polarization [Rob+20; Con+11; Mor+15].

Twitter is one of the most dynamic SNS with respect to the interactions among users and one of the most powerful with respect to the potential information that can be extracted for research purposes. Moreover, in the last quarter of 2015 this social network had 305 million monthly active users, and the last data show that the number of monthly active users has risen to around 330 million [Cle19].

In this thesis we focus mainly on the topic of polarization and, for the reasons stated above, we mainly use Twitter data to study this phenomenon. But before addressing the study of polarization itself, in this chapter we perform an exploratory analysis to get a general picture of the main behavioral patterns of Twitter users in an electoral context (the Spanish general elections of 2015 and 2016), as it is one of the most common scenarios where polarization may emerge. In this kind of contexts, people use social media as a communication medium to exchange opinions. However, there are also some users whose purpose is to influence the conversation in a way that may affect the voting choices of the people. Although these are online actions, they have an impact on the offline world.

3.1 TWITTER AND ELECTORAL CAMPAIGNS

The first time computers were used in an electoral context was in the night of the 1952 USA presidential elections to predict the electoral results [Hen12], so it is no surprise that politicians promptly realized the potential of SNS and began to use them for electoral campaigns. In the Obama campaign of 2008 (Twitter had been born only two years before), the efficacy of Twitter as a communication medium in a political context was established for the first time [AL09; Sol10]. From then on, it has gained more and more relevance in political campaigns, becoming one of the most consequential social media for politicians [Par13; RMCH19]. This trend has also inspired an ever growing number of academic works that use Twitter data to analyze different aspects of the political campaigns.

Some lines of research are centered around predicting the outcome of elections following diverse techniques [Tum+10; Bor+12b; Cal+14]. Other works investigate the interaction among parties to assess the cohesion of the parties [Che+16] or the level of debate [Bor+12b]. There are also studies that analyze the content of the tweets in order to detect their sentiment [Yaq+17; Ant+16]. Two extensive reviews that cover a variety of subjects within the area can be found in [GA13; Jun16].

In a previous paper [Bor+12b] the Spanish general elections of 2011 were studied using Twitter data. That work was focused on the prediction of the outcome of the election using a metric based on user activity. The authors also explore the interactions among users by analyzing the structural and dynamical patterns of the mention and retweet networks. The communication dynamics among politicians are also studied, finding a lack of debate, and a network growth model is proposed to reproduce these interactions.

Although we compare some of our results to those presented in that work, in this study we do not aim to predict the outcome of the elections or simply study the relations among users and politicians. We take advantage of the availability of Twitter data gathered during two electoral campaigns (the Spanish general elections of 2015 and the repetition of the elections in 2016) that are very close in time to compare the collective user behavior manifested in two analogous social systems with a similar context [MGLB18]. The individual activity and the political actors may be different; in fact, in the second election two parties formed a coalition, altering the political landscape (see Appendix F.1 for more details). Our objective is to study and characterize emergent behaviors that are recurrently manifested in electoral contexts.

To this end, we have focused on the temporal patterns of activity and interactions. Thus, we have computed time series and distributions of daily activity for both elections and we have built and analyzed networks of retweets and mentions aggregated at different temporal scales: we have studied the temporal evolution of the daily networks and we have also compared the aggregated networks of the two electoral periods. Additionally, we have explored the influence of politicians as well as regular users using the efficiency metric, defined in Chapter 2. Finally, we have extracted the sub-networks of user interactions comprised of politicians from the different parties to study their cross-mentions and retweets.

3.2 CHARACTERIZATION OF THE DATASETS

The data used consists of Twitter messages retrieved with the Twitter Streaming API (see Appendices B.1 and C for more details). This API allows to download *tweets* matching a set of keywords. In order to avoid biased results, we have chosen a set of neutral keywords (listed in Appendix B.1) to filter the messages. The sociological background of the datasets is thoroughly described in Appendix F.1.

We have downloaded tweets during a time interval that includes a pre-campaign period, the electoral campaign, the election day and a few days after that. However, the core of our analysis has been focused on the 15 days of the official electoral campaign, the *reflection day* (day before the election), the election day and the day after. During that period of 18 days we have retrieved 1793145 *tweets* for the 2015 election and 1755438 for the 2016 election.

Besides the keywords used to retrieve the data, the most used hashtags that appear in our dataset are the following:

- Top hashtags of 2015: #podemos, #psoe, #partidopopular, #ciudadanos, #16elecciones, #7deldebatedecisivo, #votapsoe, #possible, #pp, #mivotocuenta, #españa, #españaenserio, #podemos20dic, #20dicpodemos, #podemosremontada, #votapp, #hevotado, #votapodemos20d.
- Top hashtags of 2016: #afavor, #unidospodemos, #16elecciones, #votapsoe, #Debate13j, #psoe, #cambioamejor, #votapp, #españa, #ciudadanos, #partidopopular, #elcanvipossible, #unsiporelcambio, #avotar, #lasonrisadeunpais, #brexit, #eleccionesgenerales.

We have also compiled lists of Twitter accounts associated to the four main parties (PP, PSOE, Podemos and Cs) and to IU (see Table 3.1 for the full names of the parties). The latter is relevant because in the 2016 election, as we explain in Appendix F.1, this party formed a coalition (UP) with Podemos, fact that is reflected on the data. These lists of accounts allow us to analyze the difference of behaviors of regular users and politicians.

ACRONYM	SPANISH NAME	TRANSLATION
PP	Partido Popular	People's Party
PSOE	Partido Socialista Obrero Español	Spanish Socialist Worker's Party
Podemos	Podemos	We can
Cs	Ciudadanos	Citizens
IU	Izquierda Unida	United Left
UP	Unidos Podemos	United we can

Table 3.1: Names of the four main political parties contending in the 2015 and 2016 Spanish general elections and the minority left-wing party (IU) that formed a coalition with Podemos (named UP) before the 2016 election.

In order to build the set of political accounts, we have looked into the Twitter lists (which are lists of accounts elaborated by the users) defined by relevant official accounts associated to each party. We have downloaded those lists that include politicians, political institutions or supporters of the party. The total number of retrieved political users that participated in the conversation was 5227 in 2015, with an average number of followers of 4044 and 5012 in 2016, with an average number of followers of 4662.

3.3 BUILDING THE NETWORKS OF MENTIONS AND RETWEETS

Here we explain the methodology that we have adopted to build the networks of interactions from Twitter data. The interactions that we have analyzed are the *mention* and the *retweet*.

We have built mention networks by considering each user participating in the conversation as a node. Two nodes j and i are joined with a directed link from j to

i when user j mentions user i ; that is, j posts a *tweet* including *@name_of_i*. The result is a directed and weighted network where the weight of each link corresponds to the total number of tweets where j mentions i gathered during a given interval of time. The in-degree, or k_{in}^i , of a user i in this network corresponds to the total number of users that have mentioned i . The k_{out}^i corresponds to the total number of users that i has mentioned.

The retweet network is built in a similar way as the mention network. The nodes are the users and user j is joined with a directed link to user i if j retweets a message *originally* posted by i . However, if a user j retweets an original tweet by i and k retweets the message broadcast by j , there are links from k to i and from j to i , but not from k to j . Figure 3.1 illustrates this mechanism.

Figure 3.1: Scheme that shows the method used to build the networks of retweets. In panel 1), user i posts a message starting an information flow that reaches user j , who decides to retweet it. In panel 2), the retweeted message posted by j is read by k who in turn retweets it again. In panel 3) the three users are information transmitters and in the last panel the resulting retweet network is shown.

Let us remark this idea: links in the retweet network join retweeters with original posters; there are no links between the middlemen that broadcast the original message. The reason for this choice of methodology is that the Twitter API used to download the data at that moment only provided information about the original poster of a tweet in the retweet metadata. The retweet network is then a directed and weighted network where the weights of the links from j to i are the number of messages of i retweeted by j during a given period of time. The k_{in}^i of a user

i in this network corresponds to the total number of users that have retweeted i 's messages. The k_{out}^i corresponds to the total number of users that have been retweeted by i . The retweet mechanism usually results in networks that display large star subgraphs.

We have built the mention and retweet networks considering two different temporal scales. On one hand, we have built daily networks, counting the 24 hours of a day from 4AM (UTC)¹ of that day until 4AM of the following day. This way we capture full *human* activity cycles. This approach is based on a work by Morales et al. [Mor+17], where human synchronization is studied at several scales, finding strong periodicities linked to day-night and social cycles. On the other hand, we have built aggregated networks for the 15 days of electoral campaign plus the next three days (pre-election, election and post-election).

3.4 TEMPORAL EVOLUTION OF USER ACTIVITY AND INTERACTIONS

In this section we analyze the users' behavior and their interactions from a dynamical point of view by studying the temporal evolution of several relevant metrics both at the daily scale and comparing the two considered electoral periods.

3.4.1 Temporal evolution of user activity

In order to characterize the global activity of the users (number of tweets posted during a given period of time), we have computed time series of total daily activity for the whole period considered. Additionally, we have analyzed the temporal evolution of the distribution of daily activity per user.

In the left panels of Figure 3.2 we have plotted the time series of daily activity for both electoral campaigns. We have aligned them and labeled the days so that the voting day coincides with day 0. As we can see in that figure, before the electoral campaign the time series of daily activity present low values, but they start to rise around the onset of both campaigns.

In line with the existing literature [Bor+12b; LM12; Jun13], the features of the time series of activity can be explained to a high extent by offline events. There is a huge peak of messages on the election day, which accounts for more than the 30% of the total of messages during the period of study (36% in 2015 election and 32% in 2016). There are also smaller peaks during the period of the electoral campaign that are caused by the main electoral debates.

In Spain, pre-electoral silence is mandatory during the period that spans from 12AM of the day before the election to the closing of voting polls. This period is called *reflection day* and a number of restrictions apply with respect to the behavior of political parties and media. Among others, electoral events can not be held and the spread of electoral propaganda is forbidden [Ley]. The time series experiences a decrement that day, which may be triggered by the fact that politicians cease to campaign in both online and offline media. Because of this lack of stimulus, the conversation among common users decay.

It is worth pointing out the similarity of the temporal evolution of the user activity in both elections. In Figure 3.2, we have plotted (in logarithmic scale) the

¹ In Spain, the local time corresponds to UTC+1 for winter time and UTC+2 for summer time.

Figure 3.2: Left panels: time series of aggregated user activity per day for the 2015 (top) and 2016 (bottom) elections. The shadowed region corresponds to the days of electoral campaign. Right panel: Linear regression between the activity time series of both elections. Each dot corresponds to one day and the lines correspond to the linear fits of the data including the day of the elections (red dashed line) and excluding it (black continuous line).

time series of one period against the other such that the election days coincide. By performing a linear regression of these data, a significant linear correlation has been found. We obtain an $r^2 = 0.99$ when we perform the linear regression including the day of the elections and an $r^2 = 0.84$ if we exclude it. However, notice that the slopes present only a difference of 3.5%. This correlation can be partially explained by the similarity of the contexts and by the recurrent structure of the electoral campaign: It always starts 16 days before the election and lasts 15 days. Moreover, the main debates were both held 13 days before the election. Due to the fact that the number of tweets in the 2016 election was slightly lower than in 2015, the slope of the linear regression is lower than 1.

Another relevant property of the user activity is displayed in Figure 3.3, where we show that the total number of tweets, retweets and mentions per day follow a power law with respect to the number of unique users each day. Accordingly, we have fitted our data to a power law with exponent α . The values of the exponent α and correlation coefficients of the fits are displayed in table 3.2. To understand this behavior we have followed the work by Leskovec et al. [LKF05], where it is shown that when real-world networks evolve through time, the number of links $E(t)$ scale with the number of nodes $N(t)$ as:

$$E(t) \propto N(t)^\alpha \quad (3.1)$$

Notice that, whereas in 2015 the growth for the three quantities was slightly super-linear with respect to the number of unique users, in 2016 we observe an approximately linear behavior. Hence, in 2015 when more users join the conversation, the activity experiences a proportionally higher increment than in 2016.

Figure 3.3: Power law relationships of the total number of tweets, retweets and mentions per day as a function of the number of unique users that participated in the conversation each day for the 2015 campaign (top) and the 2016 campaign (bottom). Note that the data corresponding to the day of the elections (marked with a circle) were not included in the fit.

	α		r^2	
	2015	2016	2015	2016
Tweets	1.20 ± 0.06	1.04 ± 0.11	0.96	0.85
Retweets	1.25 ± 0.08	1.02 ± 0.12	0.94	0.80
Mentions	1.22 ± 0.10	0.9 ± 0.2	0.89	0.66

Table 3.2: Values of the exponents (α) and correlation coefficients (r^2) of the linear regressions of the log-log plots of tweets, retweets and mentions with respect to the number of unique users per day shown in Figure 3.3 for both electoral campaigns.

In order to further explore the characteristics of the user behavior, we have also analyzed the temporal evolution of the distribution of the daily user activity shown in the left panels of Figure 3.4. The activity distributions show a heterogeneous character, result that is in concordance with the literature [Mor+14]. We have fitted them to the following discrete power law:

$$p(x) = \frac{x^{-\gamma}}{\zeta(\gamma, x_{\min})} \quad ; \quad \zeta(\gamma, x_{\min}) = \sum_{x=0}^{\infty} \frac{1}{(x + x_{\min})^{\gamma}} \quad (3.2)$$

Where in this case x is the daily activity and ζ , the Riemann zeta function.

Figure 3.4: Temporal evolution of the distribution of activity per day in both electoral campaigns (2015 in the top panels and 2016 in the bottom panels). Left panels: probability mass functions (PMF) of the distribution of activity for each day in color code. Right panels: daily evolution of the γ exponent of the power law fit of the activity distributions. The error bars correspond to 2σ .

In the right panels of Figure 3.4 we have plotted the temporal evolution of the exponent γ of the power law. It has been computed by numerically solving the following equation to obtain the maximum likelihood estimator [CSN09] (MLE) of γ :

$$\frac{\zeta'(\gamma, x_{\min})}{\zeta(\gamma, x_{\min})} = -\frac{1}{n} \sum_{i=1}^n \ln(x_i) \quad (3.3)$$

Where the prime denotes differentiation with respect to the first argument and n corresponds to the number of users. The uncertainty of γ has been computed as:

$$\sigma = \frac{1}{\sqrt{n \left[\frac{\zeta''(\gamma, x_{\min})}{\zeta(\gamma, x_{\min})} - \left(\frac{\zeta'(\gamma, x_{\min})}{\zeta(\gamma, x_{\min})} \right)^2 \right]}} \quad (3.4)$$

This technique is proven to be more precise than a minimum squares fit to the log-log plot, which usually yields incorrect results when computing the parameters of power law distributions[CSN09]. All the exponents of power laws of discrete data have been computed in the same way. Since the minimum value of activity is 1 (and is also the most abundant in the data), we have fixed $x_{\min} = 1$ when computing the MLE of γ .

We can see that the values of the γ exponent fluctuate within a small interval ($2.0 < \gamma < 2.5$) during the considered period for both elections. These values are perfectly compatible with those presented in [Bor+12b], where the authors obtain a value of $\gamma = 2.275 \pm 0.002$.

The small fluctuations of the γ exponent tell us that the collective behavior does not change much from one day to the other. If we take into account that the value of the exponent of the distribution controls how fast it decays, we see that, when the exponent is smaller, the activity reaches higher values (and vice versa). Consequently, by looking at Figure 3.4, we can see that the day of the elections we obtain a lower value of the exponent due to the increase in activity. Additionally, during the electoral silence, consistently with the results shown in Figure 3.2, we appreciate a decrease in activity. Notice however that, while in the case of Figure 3.4 the points corresponding to the electoral silence are the highest, they are not the lowest in the activity time series of Figure 3.2. This implies that, although individually each user tends to post fewer messages that day, there are still a lot of users taking part in the conversation.

3.4.2 Temporal evolution of mention and retweet networks

We have analyzed the temporal evolution of the aggregated mention and retweet networks at two different temporal scales. On one hand, we have aggregated the networks for the whole campaign period (plus the next three days); on the other hand, we have performed an analysis of the temporal evolution of the networks by aggregating the data in a daily fashion and computing time series for different metrics.

In Figure 3.5 we show the giant strongly connected component² of the aggregated networks of 2015 and 2016. Colors correspond to the communities³ computed with the Louvain algorithm [Blo+08]. The modularity⁴ (Q) [New10] of the community partitions are shown in Table 3.3. Other community detection algorithms were applied obtaining analogous results [Pei14]. We have indicated the most probable affiliations of the nodes of each community. In order to do that we have visually inspected which nodes (or users) are the most central in each community. The centralities of the nodes have been computed with the PageRank metric [BP98], which measures the influence of a node based on its neighborhood (see Appendix G for more details). The size of the nodes in Figure 3.5 is proportional to $\log(\text{PageRank})$.

² The giant strongly connected component of a directed network is the largest sub-network where there is a directed path (a sequence of directed links) in both directions between every pair of nodes.

³ Communities are usually defined by partitioning a network in sets of nodes with many intra-connections and few inter-connections.

⁴ The modularity Q is a measure of the quality of the partition of a network in communities. It takes values strictly smaller than one and gets close to 1 when the communities are very well defined.