Supplementary Information

Bots are Less Central than Verified Accounts during Contentious Political Events

Sandra González-Bailón[1, *] and Manlio De Domenico [2, *]

[1] Annenberg School for Communication, University of Pennsylvania, 19104 Philadelphia, US

[2] Center for Information and Communication Technology, Fondazione Bruno Kessler, 38123 Trento, Italy

* Corresponding authors: Sandra González-Bailón and Manlio de Domenico

Emails: sgonzalezbailon@asc.upenn.edu; mdedomenico@fbk.eu

This PDF file includes:

Sections

1. Data
2. Retweet and Mention Networks
3. Bot Identification
4. Sentiment Analysis
5. Message-Level Regression Models
6. References

Figures

SI1. Longitudinal Changes in the Volume of Twitter Messages
SI2. Centrality Distributions for Media, Bot, and Human Accounts
SI3. Correlation of Rank Position in Percentage Reach and Centrality in the Mention Networks
SI4. Centrality in the Mention Network
SI5. Composition of the Ten Largest Communities in the Retweet and Mention Networks
SI6. Factors Explaining Strength Centrality in the Retweet Networks
SI7. Factors Explaining Strength Centrality in the Retweet Networks (Upper Decile)
SI8. Factors Explaining Strength Centrality in the Mention Networks
SI9. Factors Explaining Strength Centrality in the Mention Networks (Upper Decile)
SI10. Number of Bots Mentioning News Outlets
SI11. Number of Mutual Ties in the Retweet and Mention Networks
SI12. Comparison of Model Performance in the Classification of Bot Accounts
SI13. Cross-validation of the DL Model Performance on a New Independent Data Set
SI13. Distribution of Sentiment Scores
SI14. Predictors of Number of RTs Received by Messages (with and without Neutral Messages)

Tables

SI1. List of News Sites in Web Tracking Data (France, Nov-Dec 2018)
SI2. List of News Sites in Web Tracking Data (Spain, Sep-Oct 2017)
SI3. Network Statistics for the Retweet and Mention Networks (LCCs)

1. Data

We collected social media data through Twitter's publicly available API by retrieving all messages that contained at least one relevant hashtag. For the Gilets Jaune (GJ, Yellow Vest) mobilizations, the list of keywords included *#GiletsJaunes*; *#GiletsJaune*; *#YellowVests*; *#giletJaunes*; *#giletjaune*; *#giletJaunes*; *#giletjaune*; *#GiletsJaunes*; *#GiletsJaune*; and *#giletgialli*. For the Catalan referendum (1-O), the list included *#Catalunya*, *#Catalonia*, *#Catalogna*, *#1Oct*, *#votarem*, *#referendum*, and *#1O*. Data collection involved monitoring the Twitter stream and collecting messages using the Search API, which allowed us to collect all tweets containing any of the search terms. Based on Twitter's rate limits, we estimate that our data collection missed less than 1% of all messages with those hashtags during the period we consider (mid-November to late December 2018 for the GJ dataset, and mid-September to early October for the 1-O dataset). Using a state-of-the-art bot detection technique (more in section 3), we classified users in our datasets as being bot or human. We then distinguished bot accounts that are also verified by Twitter from unverified bot accounts. We labelled the first category "media" because verified accounts that have bot-like behavior tend to belong to media organizations, journalists and public figures. Only 0.48% of all the accounts in our data fell in the 'media' category. Accounts classified as 'bots' amounted to 38% of all users; the rest were classified as 'human'. Figure SI1 summarizes changes in the number of tweets published by each group over the period we consider.

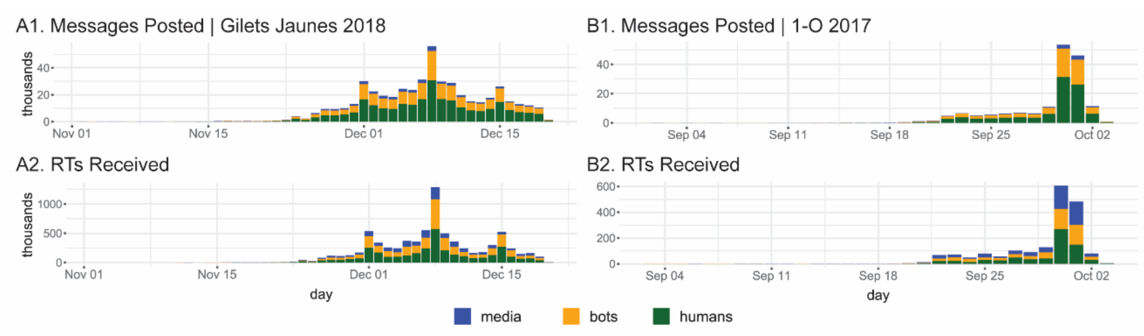

Figure SI1. Longitudinal Changes in the Volume of Twitter Messages. Panels A1 and A2 show totals in messages posted and RTs received by the three types of accounts during the Gilets Jaunes protests. Panels B1 and B2 show the same information during the Catalan Referendum protests.

The web-tracking data was provided by Comscore, a media measurement company that maintains representative panels of the online population in different countries, including France and Spain. We used their MMX Multi-Platform panel, which offers estimates for multi-platform media usage, including desktop, tablet, and mobile access. We used the "news and information" category to identify relevant web domains, and we then checked that list manually to eliminate social media sites and other irrelevant domains. Comscore counts the number of unique visitors accessing a given site and offers monthly aggregates of market share or audience reach at the domain level, so we averaged the estimates for November-December 2018 (GJ data) and for

September-October 2017 (1-O data). Tables SI1 and SI2 give the full list of news sites included in our web-tracking data, sorted in decreasing order by audience reach.

Table SI1. List of News Sites in Web Tracking Data (France, Nov-Dec 2018)

France			
linternaute	latribune	batiactu	amnestyinternational
lefigaro	buzzfeed	nationalgeographic	lelibrepenseur
huffingtonpost	rfi	freenews	lesmoutonsrebelles
groupelemonde	nbcnewsdigital	lamanchelibre	dw-deutschewelle
aufeminin	businessinsider	lagazettedescommunes	meretmarine
francetvinfo	net-iris	lenouveaudetective	lechorepublicain
bfmtv	lanouvellerepublique	reseauinternational	lesmoutonsenrages
leparisien	valeursactuelles	leberry	lcp
20minutes	lunion	destinationsante	streetpress
ouest-france	marianne	nordlittoral	nouvelleseclair
tf1	lalsace	vosgesmatin	lavenirdelartois
lci	lesinrocks	fdesouche	scoopnest
msnnews-france	euronews	lemainelibre	herault
nouvelobs	dna	contrepoints	nextwarez
rtl	leparticulier	theconversation	questionnezvoselus
lexpress	republicain-lorrain	lardennais	aisnenouvelle
europe1	jsl	wikistrike	arretsurimages
lesechos	courrierinternational	nouvelordremondial	madeinmarseille
liberation	france24	breizh-info	lobservateur
lepoint	courrier-picard	epresse	leveil
orangenews-france	lindependant	m6	philomag
ladepeche	lamontagne	alternatives-economique	montceau-news
sputnik	forbes	scienceshumaines	lavie
ledauphine	bienpublic	afp	politis
sudouest	humanite	niooz	non-stop-zapping
russiatoday	espritsciencemetaphysique	worldnewsnetwork	pourlascience
capital	vie-publique	lepopulaire	kaizen-magazine
lavoixdunord	medium	actualitte	lepharedunkerquois
parismatch	jeuneafrique	lafranceagricole	lejournaldesentreprises
franceinter	atlantico	medias-presse	herault-tribune
telerama	nextinpact	frenchweb	laviedesidees
challenges	ripostelaique	1jour1actu	en-marche
leprogres	paris-normandie	lyonmag	titrespresse
mediapart	kelnews	lepetitjournal	ledevoir
letelegramme	prismamedia	zinfos974	coe
cnnnetwork	larep	lejdc	nordeclair
slate	faitsdivers	mrmondialisation	lagauchematuer
midilibre	lemoniteur	lesobservateurs	centrepresseaveyron
estrepublicain	alterinfo	lyonne	voltairenet
francesoir	agoravox	linfo	
lejdd	egaliteetreconciliation	neonmag	
notretemps	lest-eclair	debout-la-france	
laprovence	courrierdelouest	aljazeeramedianetwork	
cnews	revolutionpermanente	lamarseillaise	
la-croix	lavenir	stopmensonges	
sondagenational	lopinion	dreuz.info	

Note: Comscore MMX Multi-Platform, Total Audience, Age 18+, November-December 2018, France. Custom Defined List.

Table SI2. List of News Sites in Web Tracking Data (Spain, Sep-Oct 2017)

Spain			
elpais	hoy	estrelladigital	elmunicipio
lavanguardia	elnortedecastilla	valenciaplaza	eladelantado
elmundo	cope	efe	horajaen
abc	diariovasco	diaridetarragona	huelvaya
elconfidencial	codigonuevo	segre	rioja2
20minutos	laopiniondemurcia	eldia	espormadrid
elperiodico	gaceta	malagahoy	tercerainformacion
okdiario	republica	diariodevalladolid	aragondigital
eleconomista	infolibre	ccma	melillahoy
eldiario	lavozdeasturias	elimparcial	aragonradio
huffingtonpost	laprovincia	elcorreogallego	canariasenhora
elespanol	eldiariomontanes	diariodejerez	
cadenaser	laopiniondemalaga	invertia	
antena3	diariodesevilla	huelvainformacion	
telecinco	vilaweb	praza	
expansion	diariodemallorca	elboletin	
rtve	lavozdigital	gasteizhoy	
publico	diariodeleon	lamarea	
europapress	laopinioncoruna	pontevedraviva	
lavozdegalicia	elperiodicoextremadura	diariodesoria	
lasexta	delcamp	directa	
libertaddigital	diariodenavarra	8tv	
periodistadigital	canarias7	informe21	
ideal	elconfidencialdigital	diariodelaltoaragon	
larazon	diaridegirona	e-noticies	
lainformacion	diariocordoba	teinteresa	
cuatro	larioja	tiempodehoy	
lasprovincias	galiciae	diagonalperiodico	
levante-emv	laopinion	cuartopoder	
atresplayer	elmundotoday	elcritic	
diariosur	jotdown	lanuevacronica	
elcomercio	eltiempo	actuall	
ara	naiz	lahaine	
economiadigital	finanzas	granadahoy	
lne	elcorreoweb	eltemps	
laverdad	lagacetadesalamanca	miciudadreal	
elplural	elpuntavui	sermosgaliza	
heraldo	lavozdealmeria	mirajerez	
diarioinformacion	que	muhimu	
elcorreo	diariodecadiz	el9nou	
vozpopuli	diariodeibiza	lavozdelsur	
naciodigital	laopiniondezamora	eldiadecordoba	
farodevigo	ctxt	diarideterrassa	
buzzfeed	alertadigital	teldeactualidad	
rac1	regio7	extraconfidencial	
ondacero	diariojaen	espiaenelcongreso	

Note: Comscore MMX Multi-Platform, Total Audience, Age 18+, September-October 2017, Spain. Custom Defined List.

2. Retweet and Mention Networks

Using the data obtained from the Twitter API, we built the RT and mentions networks that emerged during the two mobilizations. Table SI3 shows descriptive statistics for the largest connected component (LCC) used in the analyses. In addition to the RT network (on which most analyses reported in the main text are based) the table also shows data for the mention networks, which includes all the @s interactions that exclude RTs. As the table shows, the mention networks are substantially smaller. In all networks, reciprocity is very low.

Table SI3. Network Statistics for the Retweet and Mention Networks (LCCs)

	Gilets Jaunes				1-O			
	RTs		mentions		RTs		mentions	
	all ties	mutual ties	all ties	mutual ties	all ties	mutual ties	all ties	mutual ties
size	869710	6063	273146	2445	616424	1496	255034	414
number of edges	4367776	10740	1302085	3294	2104927	1946	901507	776
mean degree	10	4	10	3	7	3	7	4
mean indegree	5	4	5	3	3	3	4	4
maximum indegree	36164	154	56719	109	59310	31	32253	43
mean outdegree	5	4	5	3	3	3	4	4
maximum outdegree	15459	154	6887	109	1448	31	931	43
mean strength	15	21	14	14	8	12	9	17
mean in-strength	7	21	7	14	4	12	4	17
maximum in-strength	55648	1263	153017	938	137016	255	48481	385
mean out-strength	7	21	7	14	4	12	4	17
maximum out-strength	32838	1263	19094	938	2561	255	2704	385
reciprocity	0.006	1	0.009	1	0.004	1	0.005	1
clustering	0.12	0.169	0.178	0.244	0.133	0.235	0.166	0.49
degree correlation	-0.075	-0.07	-0.065	-0.066	-0.091	-0.197	-0.131	0.223

Figure SI2 summarizes the distribution of centrality measures in the RT and mention networks (for the directed, asymmetrical version) as they compare with centrality in the larger topology of the Twitter networks (i.e., number of followers and number of friends).

Figure SI3 shows the correlation between visibility on the web and visibility in the mentions network. The correlation is moderate and stronger than the association identified with the network of RTs, which means that news organizations with a larger audience base on the web are also more likely to be mentioned in the coverage of the two protest events. Figure SI4 shows that, on the aggregate, media accounts are the only ones receiving a significantly higher number of mentions than expected by chance – a pattern that is similar to that observed with RTs. This higher centrality holds even after controlling for account-level attributes (figures SI6 and SI8) and when using only the top percentile data (figures SI7 and SI9). Even though both the RT and the mention networks have a clear modularity structure (figure SI5), there is no clear clustering of accounts in specific communities. Verified accounts are more visible in the larger components of the mention networks but the fraction of bot and human accounts in each of these communities is roughly proportional to their overall representation in the data. Figure SI10 shows that bots are, again, not creating the discrepancy in visibility when looking at mentions and figure SI11 shows that most reciprocal ties exist between human accounts.

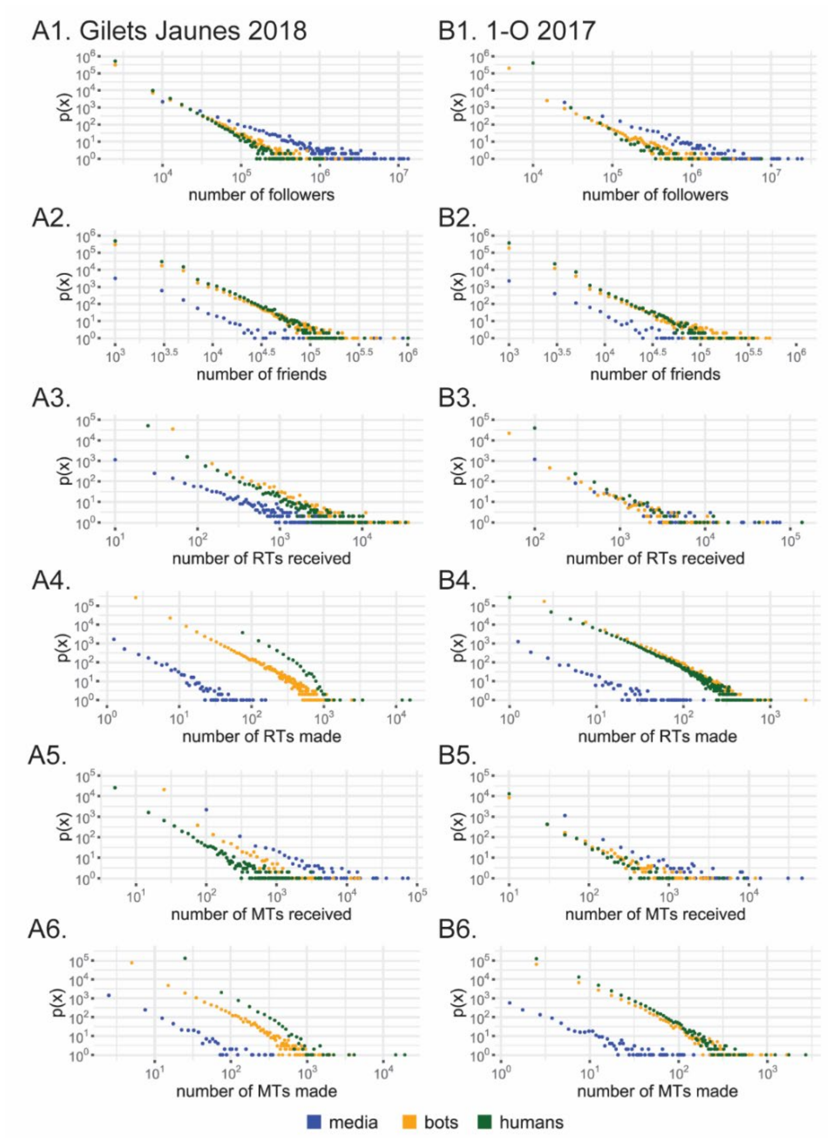

Figure SI2. Centrality Distributions for Media, Bot, and Human Accounts. The left column shows the distributions for the Gilets Jaunes data, the right column shows the distributions for the Catalan Referendum data. Centrality is assessed as number of followers (first row), friends (second row), retweets received (third row), retweets made (fourth row), mentions received (fifth row), and mentions made (sixth row). The findings reported in the main paper focus mostly on the RT network because retweets are the main mechanism for information diffusion and, therefore, overall visibility in the spread of protest-related information.

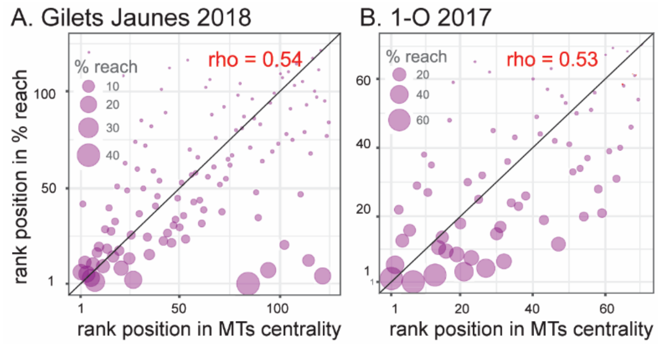

Figure SI3. Correlation of Rank Position in Percentage Reach and Centrality in the Mention Networks. The association between audience reach on the web and number of mentions received is higher than the association with RTs (reported in figure 3 of the main text). This suggests that Twitter users try to gain the attention of the larger news outlets, in terms of audience base, by targeting them more often with mentions.

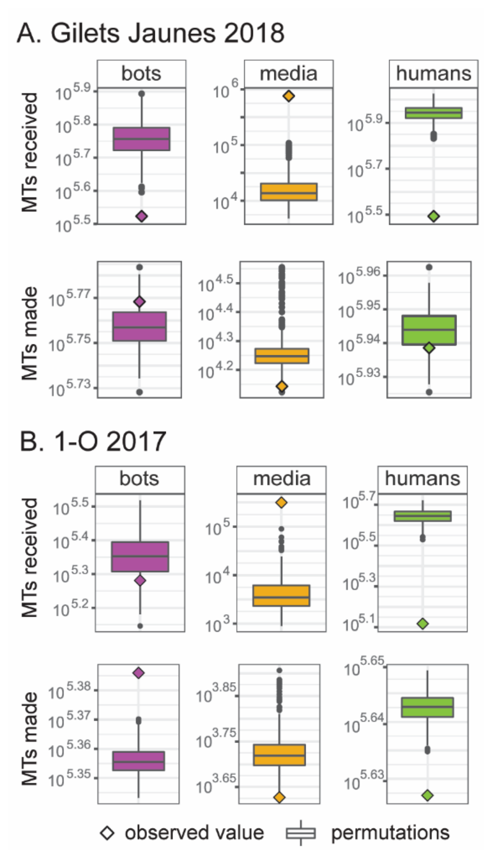

Figure SI4. Centrality in the Mention Network. These boxplots summarize the values obtained from permutations of the data where the category labels were randomly reshuffled across accounts. The observed centrality of media accounts in the mentions network is, again, significantly higher than expected by chance in both mobilizations. Human accounts receive and make significantly less mentions. The axes preserve different scales to allow visual identification of distance between permutations and observed values.

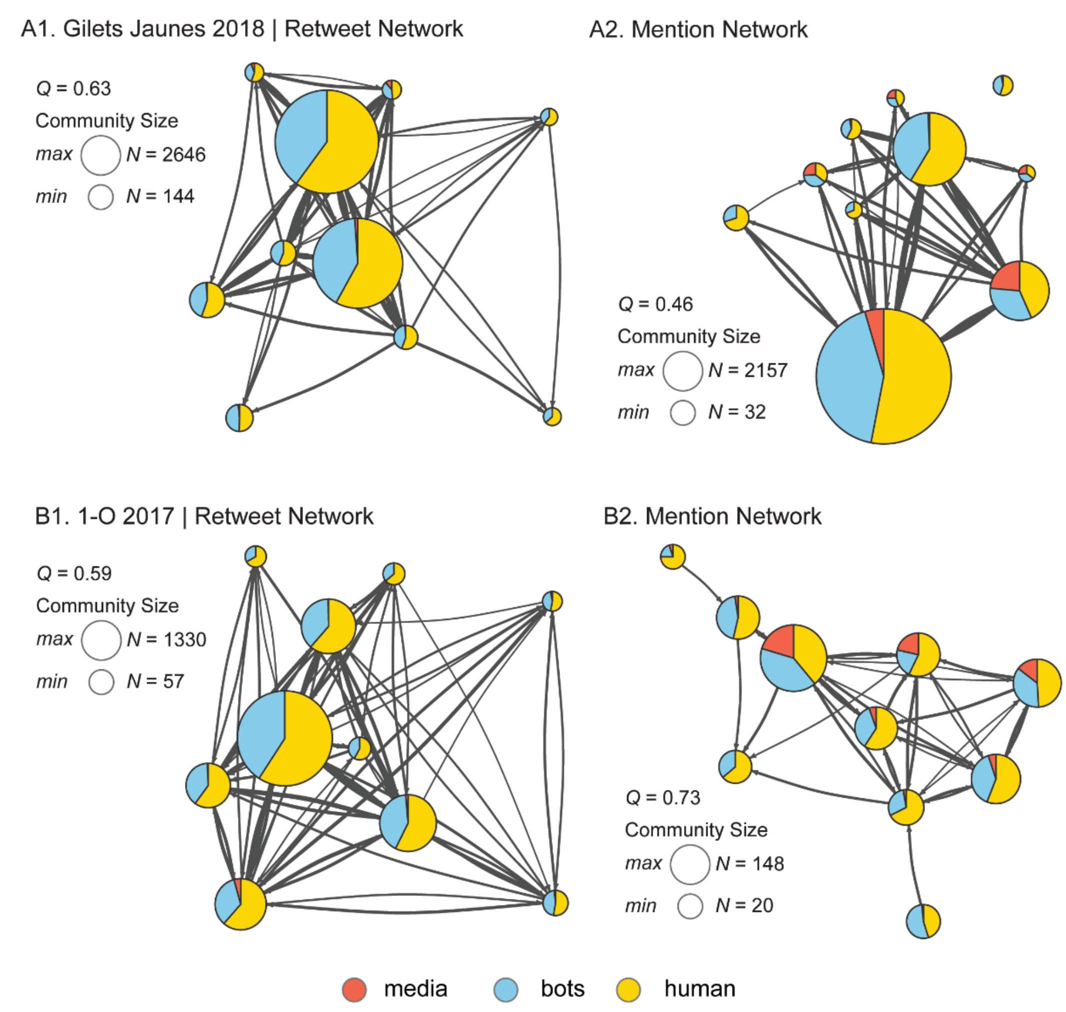

Figure SI5. Composition of the Largest Communities in the Retweet and Mention Networks. The modularity scores *Q* are derived from a random-walk community detection algorithm [1]. Each community is represented by a pie-chart summarizing the composition of the ten largest communities in the RT (left column) and mentions (right column) networks. The analysis reveals no clear evidence of clustering by account type.

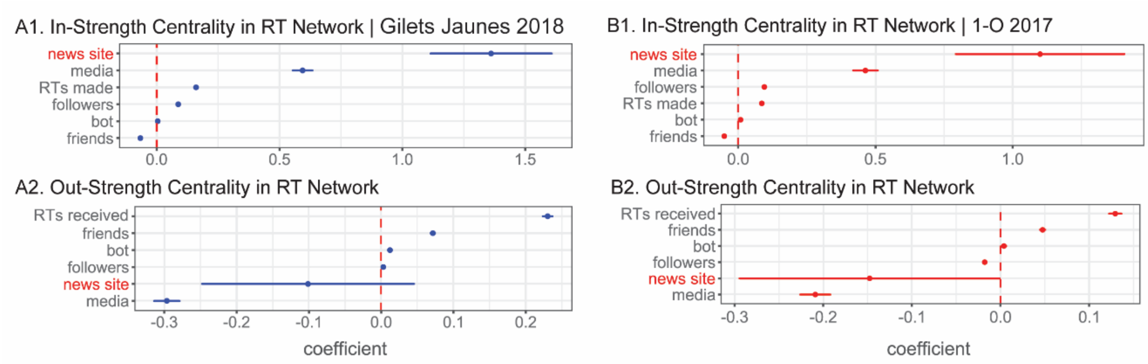

Figure SI6. Factors Explaining Strength Centrality in the Retweet Networks. Results of linear regression with robust confidence intervals predicting in- and out- centrality in the weighted RT network. Media accounts are verified by Twitter; the 'news site' category includes the accounts for which we also have web tracking data. Human accounts are the base category.

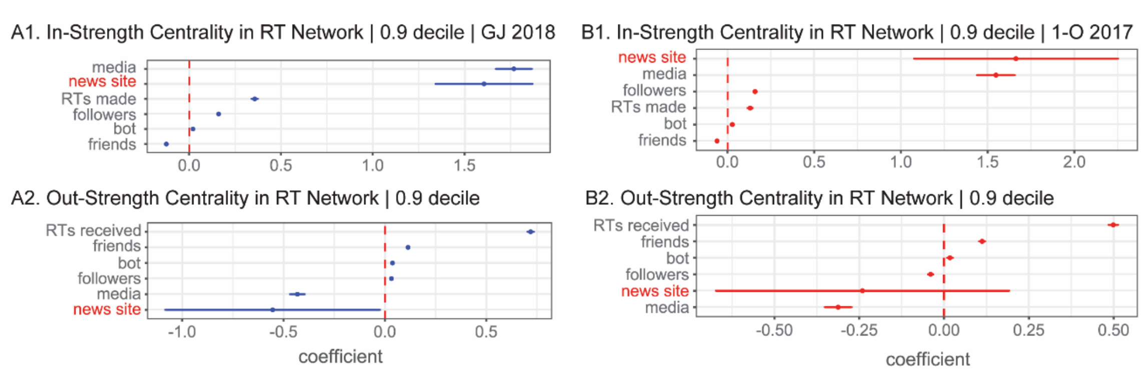

Figure SI7. Factors Explaining Strength Centrality in the Retweet Networks (Upper Decile). Results of quantile regression with bootstrapped confidence intervals predicting in- and out- centrality in the weighted RT network. Media accounts are verified by Twitter; the 'news site' category includes the accounts for which we also have web tracking data. Human accounts are the base category.

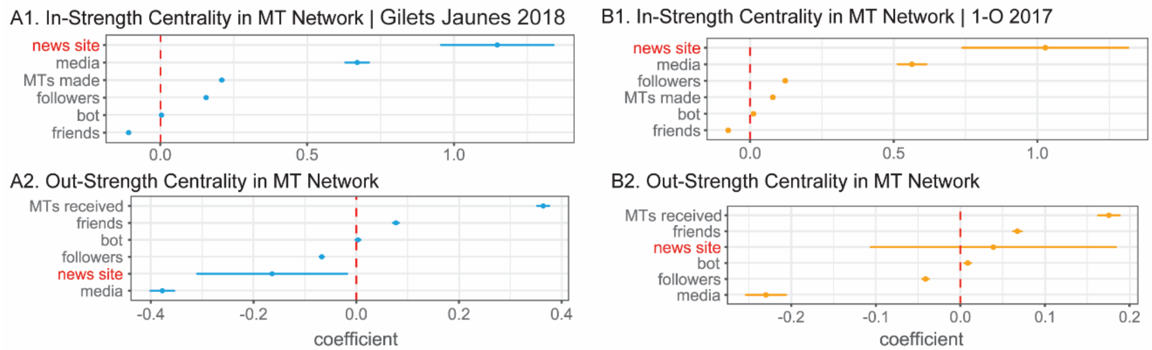

Figure SI8. Factors Explaining Strength Centrality in the Mention Networks. Results of linear regression with robust confidence intervals predicting in- and out- centrality in the weighted mention network. Media accounts are verified by Twitter; the 'news site' category includes the accounts for which we also have web tracking data. Human accounts are the base category.

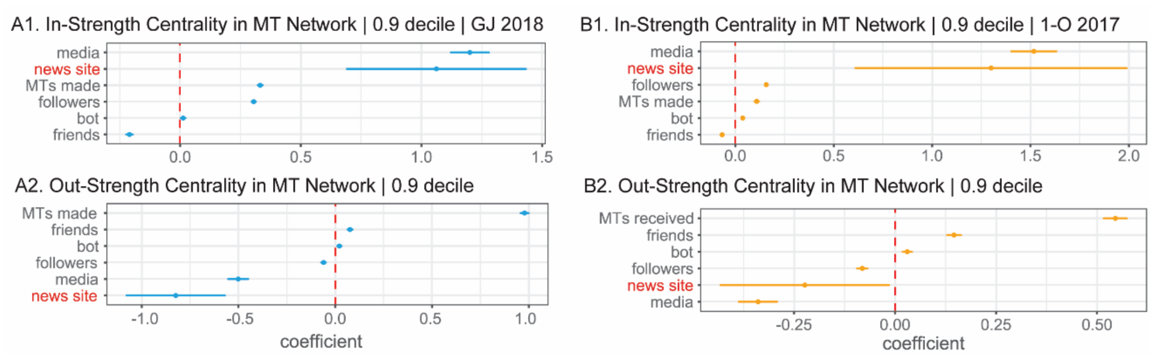

Figure SI9. Factors Explaining Strength Centrality in the Mention Networks (Upper Decile). Results of quantile regression with bootstrapped confidence intervals predicting in- and out-centrality in the weighted mention network. Media accounts are verified by Twitter; the 'news site' category includes the accounts for which we also have web tracking data. Human accounts are the base category.

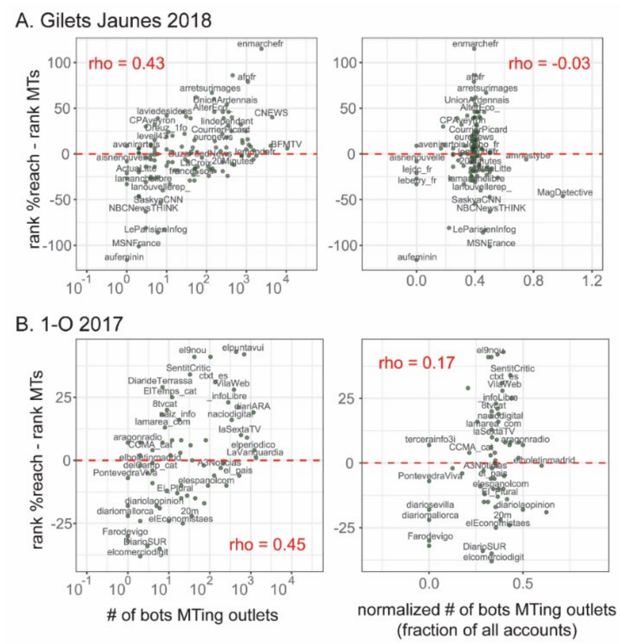

Figure SI10. Number of Bots Mentioning News Outlets. The scatterplots measure the association between the number of unverified bots mentioning news outlets and the differences in visibility rankings. As with the RT network discussed in the main text (figure 5), the moderate association disappears once we normalize the number of bots as a fraction of the neighborhood.

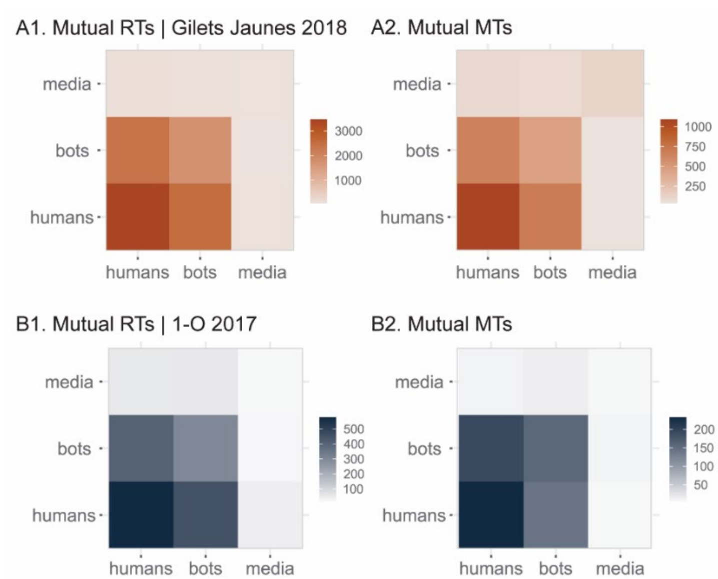

Figure SI11. Number of Mutual Ties in the Retweet and Mention Networks. The networks are very asymmetrical (the reciprocity scores are, as table SI3 shows, very low) but when reciprocated connections exist, most of them connect human accounts.

3. Bot Identification

To identify automated accounts, we use the same classification technique used in previous research [2]. For training and validation, we used publicly available datasets [3, 4] and we added other lists of humans and bots (from [2]). Overall, the training and validation dataset consists of 22,993 users: 14,218 bots and 8,775 humans. We used 80% of the dataset for training and 20% for validation. We carried the division between these two sets respecting the balance between bots and humans present at the level of the original datasets to have different types of bots in the training and validation stages. To fix the parameters, we fitted the models using three-fold cross validation on the training dataset.

The classification of accounts as human or bot relies on ten features, which recent studies have shown to yield the best classification accuracy [5, 6]: (1) statuses count; (2) followers count; (3) friends count; (4) favorites count; (5) listed count; (6) default profile; (7) geo enabled; (8) profile use background image; (9) protected; and (10) verified. As Figure SI4 shows, the deep learning technique (DL) used in this study is comparable in accuracy, specificity, sensitivity and other statistical indicators to previous existing methods including logistic regression (LOGR), ada-boost classfier (AB), random forest (RNF) and stochastic gradient descent (SGD).
For all the models except DL we use the scikit-learn implementation of the algorithms (http://scikit-learn.org). For the DL model, we use the pytorch framework (http://pytorch.org/), which consists of four fully-connected layers of 2 x N_{feats}, 4 x N_{feats}, N_{feats} and 2 hidden nodes respectively. For all layers we use a rectified linear activation unit (or ReLU) function, with the exception of the last layer, for which we use a sigmoid function. A dropout of 0.2 was also applied between the fully-connected layers in order to prevent overfitting.

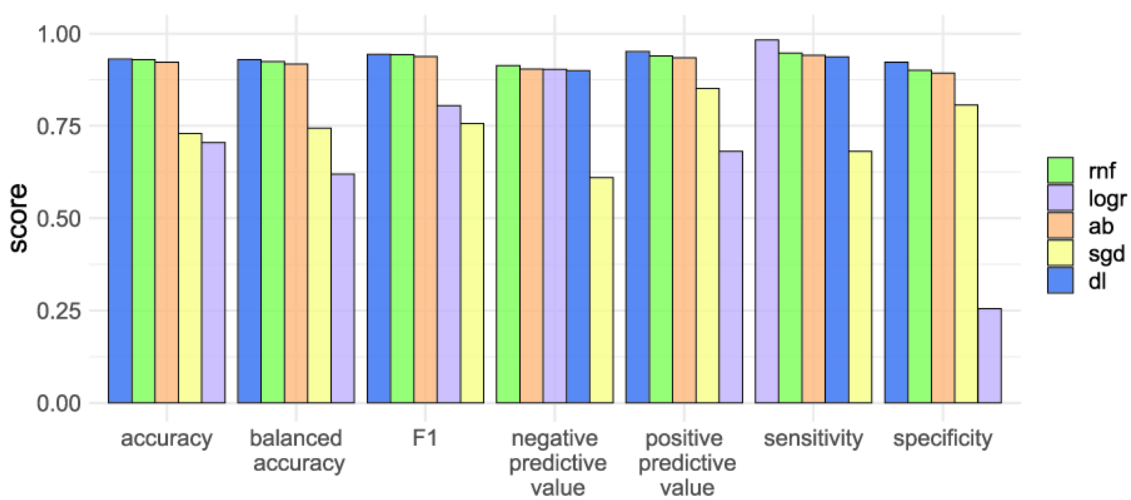

Figure SI12. Comparison of Model Performance in the Classification of Bot Accounts. The deep learning technique (DL) used in this study is comparable in accuracy, specificity, sensitivity and other statistical indicators to previous existing methods including logistic regression (LOGR), ada-boost classfier (AB), random forest (RNF) and stochastic gradient descent (SGD).

To verify if the bot detection model built by training and validating on publicly available datasets [2, 3, 4] is reliable on new social systems, we applied it on an independent data set, i.e., a list of manually curated bot/human classifications not used to train and validate our model. This data set was built during the 2018 US midterm elections [7] and provides information about 8,092 humans and 42,446 bots. We chose this specific data set because it encodes a political event of national relevance, as the two events analyzed in this study, and it is publicly available and used in the literature.

The statistical indicators of this analysis are reported in Figure SI13. Overall, the results are satisfactory, with accuracy and balanced accuracy close to 60% and an F1-score of 71.5%. The precision (Positive Predictive Value) is 93%, with a recall (Sensitivity) of 58%. In other words, the rate of false positives (i.e., labeling a human as bot) is 7%, while the rate of false negatives (i.e., labeling a bot as human) is 42%. These numbers highlight that, on the one hand, the bot detection model generalizes well enough to be applied on new social systems and, on the other hand, that performance degrades with respect to training/validation sets – as expected [8]. Out-of-domain performance is still an open (and challenging) problem for many online ML systems. Developing a better bot detection classifier is not the primary purpose of this study, but future research will only benefit from advances in this area.

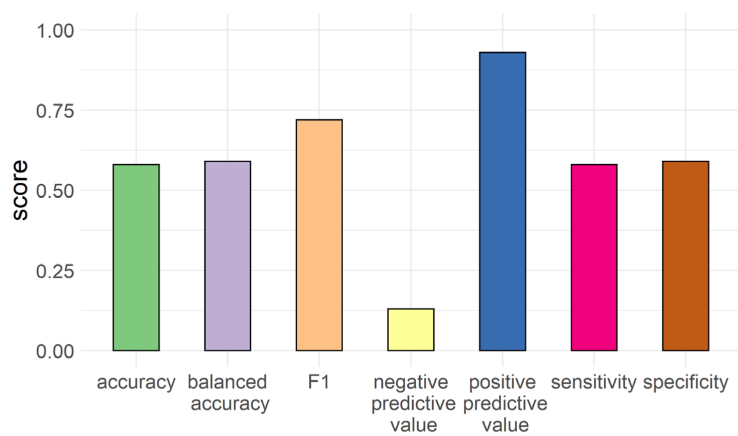

Figure SI13. Cross-validation of the DL Model Performance on a New Independent Data Set. Overall, the results are satisfactory, with accuracy and balanced accuracy close to 60% and an F1-score of 71.5%. The precision (Positive Predictive Value) is 93%, with a recall (Sensitivity) of 58%. In other words, the rate of false positives (i.e., labeling a human as bot) is 7%, while the rate of false negatives (i.e., labeling a bot as human) is 42%.

More generally, what the results of this cross-validation confirm is that, as is common with all automated classification techniques, the best performing models for a given training dataset might still generate inaccurate outputs when applied to novel data. However, when the analyses involve hundreds of thousands of accounts (as is our case), using scalable methods becomes unavoidable, even when those methods will end up misclassifying some accounts. For instance, of all the legitimate news organizations for which we have Twitter and web-browsing data (N = 126 in France and N = 73 in Spain), our bot identification technique gave us the following outputs.

In the case of France, 36 are classified as unverified bots, 75 as verified bots (media); and 15 as human accounts. In the case of Spain, 21 are classified as unverified bots, 44 as media, and 8 as human. Our bot identification approach, in other words, gets most of the classifications right (~ 90% of news accounts are correctly classified as not human), but it still introduces some miss-classifications (it is worth noting here that many of these news organizations did not have the verified badge that Twitter uses to identify accounts of public interest).

4. Sentiment Analysis

To quantify the sentiment of each Tweet, we use a well-established natural language processing technique named VADER (Valence Aware Dictionary and Sentiment Reasoner, [9]). VADER is a lexicon and rule-based sentiment analysis tool that is specifically designed to analyze sentiments expressed in social media. By design, the original version of VADER is built to analyze English text. For the English language, VADER uses a human-rated dictionary of word emotions for 7,513 lexical items annotated through Amazon Mechanical Turk. Lexical features of individual words are summed up and normalized over the text length. VADER accounts for individual words' lexical features as well as word capitalization, punctuation (e.g. “amazing!”), degree modifiers, polarity shifts due to connectors and polarity negation. There is a version of VADER supporting the analysis of text in multiple languages (https://pypi.org/project/vader-multi/). However, this version has the strong drawback of performing the sentiment analysis of the text translated from its original language to English. This is an important limitation, since languages vary substantially in how emotional content is associated to words, and the meaning of humorous or ironic sentences is often lost in translation.

To overcome this limitation, we adopt the same approach used in [5] and modify VADER to natively support sentiment analysis of texts in Spanish and Catalan by enriching the tool with sentiment lexica for those languages as obtained from the ML-SentiCon datasets [10]. The Catalan (Spanish) lexicon included 7,816 (7,377) lexical items with annotated sentiment polarities renormalized between -1 and 1. Accordingly, polarity negation, connectors and degree modifiers were also modified to account for specific lexical rules characterizing both Spanish and Catalan.

Similarly, we have further extended the range of applicability of VADER to include lexical rules specific for the French language (https://github.com/thomas7lieues/vader_FR), which is also relevant for the areas of our interest (Catalonia and France). This expansion and enrichment of VADER allows us to compute the sentiment of English, Catalan, Spanish and French text within one consistent framework for sentiment analysis.

Figure SI14 shows the distribution of these sentiment scores, which range from -1 (extremely negative sentiment) to + 1 (extremely positive sentiment), with 0 values representing neutral messages.

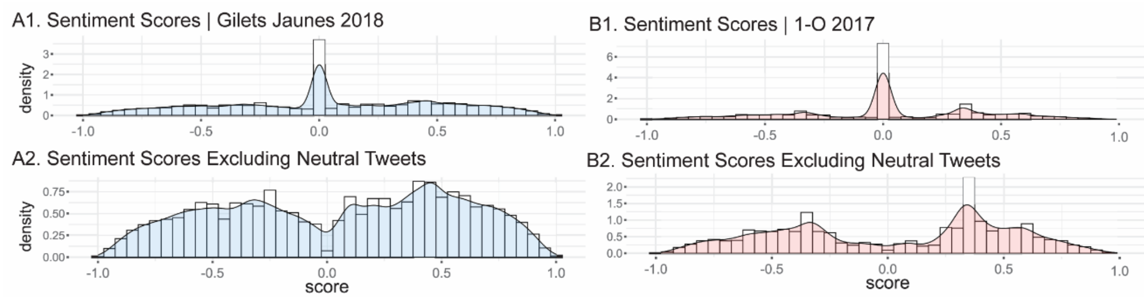

Figure SI14. Distribution of Sentiment Scores. The figure shows the distribution of sentiment scores that range from -1 (extremely negative sentiment) to + 1 (extremely positive sentiment), with 0 values representing neutral messages .

5. Message-Level Regression models

To identify the factors that predict the number of RTs messages receive, we used mixed effects models at the message-level [11, 12]. Messages are nested within unique users, so we use 'account ID' as the random effect, which also allows us to control for unmeasured characteristics at the user-level. Our fixed effects include three control variables (number of followers and friends and the sentiment score of messages), and two explanatory variables (i.e., whether the account posting the messages is classified as media or as human, with 'bots' as the base category). Overall, the model is expressed as:

$$y_i = X_i b + Z_i v + e_i$$

Where y_i is the number of RTs a message *i* receives, X_i are the fixed effects or predictor variables, b are the regression coefficients, Z_i are the random effects, v are the estimated coefficients for the random effects, and e_i are the residuals.

As a robustness test, we fit the models with and without neutral messages in the data (i.e., messages with a sentiment score 0, which are the majority of the messages in our data, see figure SI13). Figure SI15 shows that the results remain qualitatively unchanged.

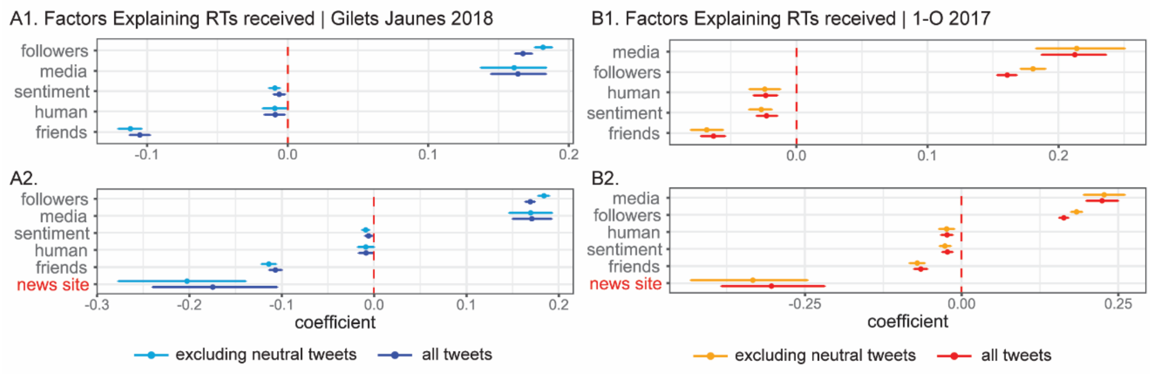

Figure SI15. Predictors of Number of RTs Received by Messages (with and without Neutral Messages). Comparison of estimated effects on number of retweets received when messages classified as neutral are removed. Media accounts are verified by Twitter; the 'news site' category includes the accounts for which we also have web tracking data (i.e., prominent news organizations).

6. References

1. Pons, P. and M. Latapy, *Computing Communities in Large Networks Using Random Walks.* Journal of Graph Algorithms and Applications, 2006. **10**(2): p. 191-218.

2. Stella, M., M. Cristoforetti, and M. De Domenico, *Influence of augmented humans in online interactions during voting events.* PLOS ONE, 2019. **14**(5): p. e0214210.

3. Cresci, S., et al., *Fame for sale: Efficient detection of fake Twitter followers.* Decision Support Systems, 2015. **80**: p. 56-71.

4. Cresci, S., et al., *The Paradigm-Shift of Social Spambots: Evidence, Theories, and Tools for the Arms Race*, in *Proceedings of the 26th International Conference on World Wide Web Companion*. 2017, International World Wide Web Conferences Steering Committee: Perth, Australia. p. 963–972.

5. Stella, M., E. Ferrara, and M. De Domenico, *Bots increase exposure to negative and inflammatory content in online social systems.* Proceedings of the National Academy of Sciences, 2018.

6. Ferrara, E., *Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election.* First Monday, 2017. **22**(8).

7. Yang, K.-C., et al., *Scalable and Generalizable Social Bot Detection through Data Selection.* Proceedings of the AAAI Conference on Artificial Intelligence, 2020. **34**(01): p. 1096-1103.

8. Echeverria, J., et al., *LOBO: Evaluation of Generalization Deficiencies in Twitter Bot Classifiers*, in *Proceedings of the 34th Annual Computer Security Applications Conference*. 2018, Association for Computing Machinery: San Juan, PR, USA. p. 137–146.

9. Hutto, C. and E. Gilbert. *Vader: A parsimonious rule-based model for sentiment analysis of social media text*. in *Eighth International Conference on Weblogs and Social Media (ICWSM-14).* 2014.

10. Cruz, F.L., et al., *Building layered, multilingual sentiment lexicons at synset and lemma levels.* Expert Systems with Applications, 2014. **41**(13): p. 5984-5994.

11. Faraway, J.J., *Extending the Linear Model with R. Generalized Linear, Mixed Effects and Nonparametric Regression Models*. 2005, New York: Chapman and Hall.

12. Galecki, A. and T. Burzykowski, *Linear Mixed-Effects Models Using R*. 2013, New York: Springer.