

Arrels del trumpisme: Homofília i rebuda social en el suport a Donald Trump a Reddit

Joan Massachs
Universitat Politècnica de
Catalunya, Espanya
joan.massachs@est.fib.upc.edu

Corrado Monti
ISI Foundation, Itàlia
corrado.monti@isi.it

Gianmarco
De Francisci Morales
ISI Foundation, Itàlia
gdfm@acm.org

Francesco Bonchi
ISI Foundation, Itàlia
Eurecat, Espanya
francesco.bonchi@isi.it

RESUM

Estudiem l'emergència del suport a Donald Trump a la discussió política de Reddit. Amb gairebé 800k subscriptors, "r/The_Donald" és una de les comunitats més grans de Reddit i un dels nuclis principals de partidaris de Trump. Es va crear el 2015, poc després que Donald Trump comencés la campanya presidencial. Utilitzant només dades del 2012, prediem la versemblança de ser un partidari de Donald Trump el 2016, l'any de les darreres eleccions presidencials dels EUA. Per caracteritzar el comportament dels simpatitzants de Trump, partim de tres hipòtesis sociològiques diferents: l'homofília, la influència social i la rebuda social. Operacionalitzem cada hipòtesi amb un conjunt de característiques per cada usuari i entrenem classificadors per predir-ne la participació en r/The_Donald.

Trobem que les característiques basades en l'homofília i la rebuda social són els senyals més predictius. Per contra, no observem un fort impacte dels mecanismes d'influència social. També realitzem una introspecció del model amb més bons resultats per construir una "persona" del típic partidari de Donald Trump a Reddit. Trobem evidències que els trets més prominents inclouen una predominança d'interessos masculins, una inclinació política conservadora i llibertariana i vincles amb contingut políticament incorrecte i conspiratori.

CCS CONCEPTS

• **Applied computing** → *Sociology*; • **Information systems** → **Web mining**; • **Computing methodologies** → *Machine learning*.

ACM Reference Format:

Joan Massachs, Corrado Monti, Gianmarco De Francisci Morales, and Francesco Bonchi. 2020. Arrels del trumpisme: Homofília i rebuda social en el suport a Donald Trump a Reddit. In *12th ACM Conference on Web Science (WebSci '20)*, July 6–10, 2020, Southampton, United Kingdom. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3394231.3397894>

1 INTRODUCCIÓ

L'emergència i succés de Donald Trump durant les eleccions presidencials dels EUA de 2016 va agafar molts tertulians per sorpresa.¹

¹<https://www.forbes.com/sites/stevedenning/2016/11/13/the-five-whys-of-the-trump-surprise>

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WebSci '20, July 6–10, 2020, Southampton, United Kingdom

© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 978-1-4503-7989-2/20/07...\$15.00
<https://doi.org/10.1145/3394231.3397894>

Les raons darrere de tal malestar han estat subjecte d'un intens debat: es remunten a un ressorgiment del populisme autoritari, al context socioeconòmic dels EUA a un món globalitzat i, fins i tot, a la seva atracció crua com a candidat divisiu i antiestablishment, per dir-ne uns quants [1, 6, 9, 29, 33].

Encara que entendre les causes precises del succés de Trump pot ser impossible, les dades disponibles sense precedents via la web i les xarxes socials ens donen una oportunitat per, almenys, entendre'n els partidaris. De fet, l'objectiu d'aquest treball és estudiar l'emergència del suport a Donald Trump a la discussió política de Reddit. La campanya de Donald Trump es va basar en gran manera en les xarxes socials i Reddit va ser una plataforma fonamental pel seu succés [14]. A més, Reddit permet d'estudiar aquesta emergència amb una perspectiva més àmplia, identificant quins factors van anticipar el suport a Trump anys abans.

Reddit és un lloc web social d'agregació de notícies; el 2012, va atraure 46 milions de visitants únics; el 2016, va ser el setè web més visitat als Estats Units, amb més de 200 milions de visitants.² Els usuaris utilitzen pseudònims i les publicacions i els comentaris són públicament disponibles. És comú d'utilitzar Reddit per discutir notícies i temes polítics. Aquestes característiques el fan un lloc prometedor per la recerca social. A més a més, una de les comunitats en línia més grans de partidaris de Donald Trump és la comunitat de Reddit r/The_Donald.

Malgrat que aquesta comunitat nasqué el 2015, gràcies a la disponibilitat de les dades de Reddit històriques al llarg dels anys, podem emmarcar la nostra investigació com una tasca de predicció. Per tant, la nostra metodologia a aquest treball és la següent. Primerament, construïm un *grup de focus computacional* [18] de 44 924 usuaris políticament actius a Reddit, que han participat en discussions polítiques tant el 2012 com el 2016. Llavors, dividim el nostre grup de focus en dues classes: els que participen en r/The_Donald el 2016 i els que no ho fan. La participació en r/The_Donald és una representació vàlida per estudiar el suport a Donald Trump, ja que les normes del subreddit diuen explícitament que la comunitat és "únicament per partidaris de Trump" i que els usuaris dissidents seran eliminats. Basant-nos en això, identifiquem 7083 usuaris (el 15,8%) amb una presència significativa a aquesta comunitat.

Per tant, modelem la nostra qüestió com una tasca de predicció binària: *donades les característiques d'un usuari el 2012, podem predir si el 2016 participarà en r/The_Donald?*

Per aquest propòsit, definim un conjunt de característiques partint de teories sociològiques existents sobre la formació d'opinió. En particular, les nostres característiques capturen tres mecanismes

²<http://web.archive.org/web/20121231152526/http://www.reddit.com/about/>
<http://web.archive.org/web/20161213123205/https://www.alexa.com/topsites/countries/US>

socials: la influència, la conformitat i l'homofilia. Cada mecanisme és el producte d'un tipus diferent d'interacció entre un usuari i el seu ambient. Primer, considerem les comunicacions directes –un usuari parant atenció a un comentari. Aquesta interacció pot conduir a un canvi d'actitud mitjançant persuasió o reacció; en general, parlem d'influència *directa*. Determinar si les interaccions a xarxes socials poden fer reconsiderar les opinions d'un mateix ha atret una atenció considerable [8] i diverses preocupacions [19]. El segon tipus d'interacció que considerem és la *rebuda social*. Pot conduir a un canvi d'actitud via *conformitat* [5], ja que els usuaris poden voler coincidir amb la norma percebuda de les seves comunitats. També pot passar el contrari: l'*anticonformitat* [36] pot conduir els usuaris a desafiar les normes percebudes que experimenten. Operacionalitzem la rebuda social com la puntuació rebuda per un usuari en una comunitat particular. Finalment, considerem interaccions indirectes: interessos comuns, proximitat, grups socials. Poden explicar actituds comunes via *homofilia* [24]. Observem les interaccions indirectes com la participació en comunitats de Reddit. No són necessàriament polítiques i inclouen també aficions, interessos, religions, situacions geogràfiques i, fins i tot, addiccions. Distingir la influència de l'homofilia és un problema de llarga durada a l'anàlisi de xarxes socials [16].

Agregant aquests tres conjunts de característiques, construïm un ric conjunt de dades sobre el nostre grup de focus d'usuaris políticament actius. Compartim aquest conjunt de dades, anomenat `reddit-politics-12-16`, per investigacions futures sobre aquest tema. A aquest treball, l'utilitzem per respondre les següents qüestions de recerca:

- *Podem predir qui donarà suport a Donald Trump quatre anys abans?*
- *Quina mena d'interacció és més predictiva de la participació en `r/The_Donald`?*
- *Quins són els principals trets d'un futur partidari de Trump a Reddit?*

El nostre millor model aconsegueix una puntuació F1 del 35,3%, més del doble que la del classificador bàsic aleatori, un 15,2%, i una àrea sota la corba ROC del 0,70. Trobem evidències que l'homofilia és el millor predictor d'entre els considerats, mentre que la conformitat també juga un rol notable. No observem evidència significativa d'influència directa. Emergeixen diversos trets interessants entre els que prediuen el suport a Donald Trump, que descrivim en detall a la secció 5. La “persona” del partidari de Trump té punts de vista conservadors i llibertarians i participa en comunitats políticament incorrectes i conspiratòries. D'entre els seus interessos, els més importants són l'emprenedoria, les armes i els videojocs. D'entre els trets més negativament correlacionats amb el trumpisme, trobem l'ateisme i l'ecologisme i interessos com la cuina o l'electrònica DIY.

2 ESTAT DE LA QÜESTIÓ

Reddit, com a font de dades interessants i disponibles públicament, ha atret molta atenció en treballs recents. Medvedev et al. [25] en van fer una compilació exhaustiva el 2017. Més recentment, alguns treballs han utilitzat dades de Reddit per estudiar l'evolució de creences i tendències específiques, així com la relació entre la política

i diferents comunitats de Reddit. Kane i Luo [13] utilitzen LDA per caracteritzar les tendències polítiques de subreddits no polítics; malgrat això, la presència d'arguments fa difícil d'interpretar els resultats. Klein et al. [15] caracteritzen els usuaris de Reddit que s'han unit al subreddit `r/conspiracy`, per estudiar els punts de vista conspiratoris. Troben que el llenguatge difereix clarament entre els usuaris conspiratoris i el grup de comparació; en particular, observen diferències en l'ús de paraules relacionades amb el crim, el govern i el poder, mentre que no observen diferències significatives en emocions positives o negatives. També analitzen quins subreddits actuen de “camí de pas” cap a `r/conspiracy` construint una xarxa entre comunitats utilitzant una similitud basada en usuaris. Els subscriptors de `r/conspiracy` estan sobrerrepresentats en comunitats relacionades amb la pornografia, la cultura tecnològica i la música. Com mostrem a la secció 5, trobem una correlació significativa entre `r/conspiracy` i `r/The_Donald`. Grover i Mark [12] analitzen patrons de comportament a `r/altright`, trobant-ne alguns que alerten d'una possible radicalització, com la fixació o la identificació de grup.

Un petit nombre de treballs es focalitzen a `r/The_Donald` explícitament. Zannettou et al. [37] estudien la propagació de mems a través de diverses comunitats `alt-right` a xarxes socials, incloent-hi Reddit i `r/The_Donald`. Flores-Saviaga et al. [10] investiguen el comportament dels usuaris de `r/The_Donald`, trobant que adopten sovint un argot trol, sobretot a l'hora de discutir teories de la conspiració. També troben que els missatges que atrauen més participació són els que expliquen en detall algunes circumstàncies polítiques i criden als usuaris a actuar. A les conclusions, observen la necessitat d'una mirada més profunda cap a aquesta comunitat, investigant-ne les arrels.

La nostra tasca de predicció es pot considerar relacionada amb la detecció de posicionament, ja que identifiquem l'opinió d'un conjunt predeterminat d'individus sobre un tema específic. Normalment, però, la detecció de posicionament involucra determinar el posicionament d'un text curt, on sovint l'autor menciona explícitament l'objecte sobre el qual es posiciona. Per exemple, Mohammad et al. [28] a la tasca 6 del SemEval-2016 classifiquen els posicionaments d'un conjunt d'usuaris de Twitter sobre diferents temes. Un dels temes de la tasca és la candidatura presidencial de Donald Trump, on el millor classificador aconsegueix una puntuació F1 del 0,56 (comparada amb puntuació F1 de 0,29 del classificador bàsic constant). Les mètriques de rendiment de classificació del nostre millor model van en la línia d'aquests resultats. Altres exemples de detecció de posicionament polític inclouen el treball de Lai et al. [17] sobre classificar posicionaments sobre el referèndum d'Itàlia de 2016 i el de Taulé et al. [34] sobre posicionaments sobre la independència de Catalunya.

Normalment, els mètodes de detecció de posicionament confien molt en característiques lingüístiques [17, 28, 34] per predir punts de vista explícits. Tanmateix, també es pot utilitzar l'homofilia per identificar correlacions significatives entre creences polítiques i altres trets. Per citar DellaPosta et al. [7], “*les dinàmiques d'autoreforç de l'homofilia i la influència amplifiquen dramàticament fins i tot afinitats electives molt petites entre l'estil de vida i la ideologia*”. Gari-mella i Weber [11] han estudiat aquest fenomen a Twitter, analitzant trets significatius d'usuaris de Twitter demòcrates i republicans.

Magdy et al. [22] empren una barreja d'aquestes característiques per predir opinions islamòfobes a Twitter abans que s'expressin. Les característiques de xarxa poden aconseguir per si soles una precisió del 79% a aquesta tasca, confirmant així la importància de l'homofilia en la predicció de punts de vista no expressats.

3 DADES

Agafem el nostre conjunt de dades de Reddit [2]. Reddit s'organitza en comunitats sobre diferents temes, anomenades *subreddits*. Els usuaris poden *publicar* a aquests subreddits i *comentar* a publicacions i comentaris, creant així una discussió amb estructura d'arbre. Anomenem *missatge* a una peça genèrica de contingut generat per un usuari, quan la distinció entre publicació i comentari no és rellevant. A més, els usuaris també poden *votar positivament* un missatge per mostrar-ne aprovació o acord (i els seus oposats amb un *vot negatiu*). La *puntuació* d'un missatge és el nombre de vots positius menys el nombre de vots negatius que ha rebut.³

Per definir el nostre grup de focus, primer hem de definir el conjunt de subreddits que volem considerar. Com que ens interessa la discussió política, triem com a llavor *r/politics*, el subreddit polític més gran. Llavors agafem els 50 subreddits més similars a *r/politics* segons una similitud cosinus sobre una representació vectorial dels subreddits basada en anàlisi semàntica latent, que captura subreddits que tenen una base d'usuaris similar a la de la llavor.⁴

Considerant aquests *subreddits polítics*, anomenem *actius* els usuaris que han escrit almenys 10 comentaris el 2012 i 10 comentaris el 2016 a algun d'aquests subreddits. El conjunt d'usuaris actius conté 44 924 usuaris i constitueix el nostre grup de focus computacional [18]. A més, anomenem *populars* els 1000 subreddits amb més comentaris.

Ara centrem-nos en la tasca a fer. Volem predir quins usuaris donaran suport a Trump el 2016, l'any que va ser elegit president dels Estats Units, mirant únicament dades del 2012, l'any de les eleccions presidencials anteriors.

Assignació de classe. Utilitzem la participació en *r/The_Donald* el 2016 per inferir la classe de cada usuari actiu a Reddit. Convé mencionar que el 2012 encara no existia el subreddit *r/The_Donald*, de manera que no tenim cap idea dels partidaris de Trump el 2012. Tanmateix, agafar tots els usuaris que van comentar a *r/The_Donald* és massa sorollós per una definició operacional. Com a primera aproximació, definim un usuari com a partidari de Trump si té almenys 4 comentaris a *r/The_Donald* i la suma de les seves puntuacions és almenys 4. Això correspon a 7427 usuaris; però ens adonem que 1200 d'aquests usuaris també havien publicat al subreddit dedicat a l'altra candidata, Hillary Clinton (*r/hillaryclinton*). Per tant, per tenir en compte l'activitat política general d'un usuari, considerem que un usuari és partidari de Trump el 2016 si ha fet almenys 4 comentaris més a *r/The_Donald* que a *r/hillaryclinton* i la suma de les puntuacions (tant positives com negatives) a *r/The_Donald* és almenys 4 punts superior a la de les puntuacions a *r/hillaryclinton*. Aquesta definició ens permet tenir un conjunt de dades amb un

desequilibri de classes limitat i alhora maximitzem la confiança de la classe assignada a cada usuari. Amb aquest mètode, descartem 344 usuaris (4,6% del nostre conjunt inicial) que, segons aquesta definició, no eren partidaris clars de Trump el 2016. Finalment, del nostre grup de focus de 44 924 usuaris, n'etiquetem 7083 (15,8%) com a partidaris de Trump i 37 841 (84,2%) com a no partidaris de Trump. Aquestes dues classes són les que adoptem per tota la nostra anàlisi.

Influència directa. Diem que un usuari actiu *u* interacciona amb el subreddit polític *r* quan *u* respon un missatge, a qualsevol subreddit popular, fet per un altre usuari *v* que ha publicat al subreddit *r* el 2012. Aquesta noció d'influència directa captura la idea que *u* interacciona amb *v*, que és un usuari que pertany a la comunitat *r* i, per tant, és possiblement exposat a les actituds d'aquesta comunitat, independentment d'on s'efectuï la interacció. Optem per aquesta noció d'influència per evitar que les interaccions siguin extremadament esparses, tal com passaria si les consideréssim d'usuari a usuari.

A més a més, considerem que una interacció és *conflictiva* quan un dels dos missatges té una puntuació de com a mínim 10 i l'altre en té una de com a màxim -10. Aquesta definició captura la noció que les dues actituds expressades als missatges difereixen i que la interacció possiblement representa un conflicte. Per cada usuari actiu i subreddit polític, calculem quants cops l'usuari ha interaccionat amb el subreddit i quantes d'aquestes interaccions han estat conflictives.

Rebuda social. Considerem les puntuacions rebudes per un usuari *u* a un subreddit polític *r* el 2012 com una representació de la rebuda social d'*r* a *u*. Les puntuacions positives i negatives les considerem separades, com a formes de reforçament positiu i negatiu, respectivament. Utilitzem les puntuacions mitjanes per normalitzar-les a través dels diferents nivells d'activitat dels usuaris. Com més alta sigui la mitjana de puntuacions positives d'un usuari, més bé rep la comunitat la seva actitud. Per contra, la mitjana de puntuacions negatives mostra quant desaprova una comunitat l'actitud de l'usuari.

Homofilia. Els usuaris poden tenir comportaments semblants –donar suport a Trump– perquè ja tenen característiques i interessos semblants. Capturem aquesta noció mirant la participació d'un usuari actiu *u* en un subreddit popular *r*. Els usuaris amb interessos semblants és probable que pertanyin a les mateixes comunitats, que és una forma d'homofilia. Hem experimentat amb versions d'aquestes característiques tant numèriques (nombre de comentaris) com binàries i hem trobat que els resultats són semblants. Com que la segona versió és més senzilla d'interpretar, reportem els resultats de la característica binària.

Així, doncs, el nostre conjunt de dades final té les següents característiques per cada usuari:

³<https://www.reddithelp.com/en/categories/reddit-101/reddit-basics/how-post-s-or-comments-score-determined>

⁴<https://www.shorttails.io/interactive-map-of-reddit-and-subreddit-similarity-calculator>

Participació:

- La característica *r part.* és certa quan l'usuari participa en el subreddit *r*, és a dir, ha escrit un comentari a *r*.

Puntuació:

- La característica *r p. pos.* és la mitjana de les puntuacions positives dels comentaris de l'usuari al subreddit *r*.
- La característica *r p. neg.* és la mitjana de les puntuacions negatives dels comentaris de l'usuari al subreddit *r*.

Interacció:

- La característica *i. nom.* és el nombre total d'interaccions directes que l'usuari ha tingut.
- La característica *r i. dist.* és la fracció de les interaccions directes que l'usuari ha tingut amb usuaris que participen en el subreddit *r*.
- La característica *r i. pos.* és la fracció d'interaccions directes *no conflictives* amb usuaris que participen en el subreddit *r* d'entre les interaccions directes amb usuaris que participen en *r*.

Aquest conjunt de dades és l'artefacte principal resultant de la nostra recerca. Creiem que és de valor independent per la recerca en ciència social computacional i, per tant, la posem a disposició de la comunitat.⁵

Tant per les puntuacions com per les interaccions positives, si l'usuari no té comentaris al subreddit *r* i, per tant, les característiques no estan definides, el valor de la característica l'agafem com la mitjana de la població. Així, l'algorisme de classificació no pot distingir una puntuació mitjana de la d'un usuari que no participa. És a dir, aquest mètode d'imputació elimina de les característiques la informació sobre la participació, amb la intenció de separar l'homofília de la rebuda social i la influència directa.

A més a més, hem extret dos altres conjunts de característiques bàsiques interpretables fonamentats en la mineria de textos:

Sentiment: La característica *r polaritat* és la polaritat mitjana dels títols de les publicacions de l'usuari al subreddit polític *r*. Calculem la polaritat utilitzant TextBlob.⁶

Bossa de paraules: La característica *x bossa* és el pes tf-idf de la paraula *x* als títols de les publicacions a subreddits polítics que l'usuari ha fet.

A més d'això, creem dos conjunts de característiques derivats: puntuacions bisecades i interaccions bisecades. Aquestes característiques es basen en les característiques de puntuació i interacció, dividint els subreddits en dos conjunts. L'agrupament es defineix en funció de si la fracció d'usuaris partidaris de Trump el 2016 al subreddit donat és superior o inferior a la mitjana. Anomenem aquests dos conjunts de subreddits *T* i *N*, respectivament. Com que aquest agrupament de característiques utilitza la classe, no les utilitzem per investigar-ne el poder predictiu. Més aviat, les aprofitem per obtenir informació sobre quines correlacions hi ha entre les característiques i el suport a Trump. Per les *puntuacions bisecades*, en comptes de tenir un valor positiu i un de negatiu per cada subreddit, només en tenim quatre: mitjanes de puntuacions positives i

negatives per cada un dels dos grups de subreddits. De manera semblant, per les *interaccions bisecades*, les interaccions d'un usuari es resumeixen en tres valors: (i) la fracció d'interaccions directes que un usuari ha tingut amb usuaris que participen en un subreddit de *T*, (ii) la fracció d'interaccions directes *no conflictives* amb usuaris que participen en un subreddit de *T* i (iii) la fracció d'interaccions directes *no conflictives* amb usuaris que participen en un subreddit a *N*.

4 METODOLOGIA

Per cada conjunt de característiques descrit a la secció anterior, entrenem diferents algorismes de classificació per predir quins usuaris esdevindran partidaris de Trump el 2016. A més, també provem les combinacions possibles entre les característiques de participació, puntuació i interacció.

Abans d'entrenar cada algorisme de classificació, preprocessem les dades i realitzem una selecció de característiques per evitar sobreajustament i obtenir models més parsimoniosos i interpretables. En particular, realitzem els següents passos de preprocessament: (i) eliminem les característiques esparses, (ii) estandarditzem els valors numèrics, (iii) seleccionem només les característiques significativament correlacionades i (iv) eliminem la multicolinealitat.

Al primer pas, eliminem les característiques que estan definides per menys de 500 usuaris (dels 44 924 totals); per les característiques de participació, utilitzem una norma més estricta i eliminem els subreddits amb menys de 250 usuaris del nostre grup; per les característiques de la bossa de paraules, eliminem les paraules utilitzades per menys de 45 usuaris (el 0,1% del nostre grup de focus). Al pas d'estandardització, fem una translació i reescalem cada característica numèrica per fer que tingui mitjana zero i variància unitària. Per la selecció de característiques, eliminem totes les que no estan significativament correlacionades ($p < 0,05$) amb la variable objectiu, d'acord amb la correlació de Pearson. Finalment, per eliminar la multicolinealitat, eliminem iterativament les característiques més significativament col·lineals mitjançant un enfocament voraç per eliminació de característiques cap enrere; mesurem la col·linealitat amb el factor de la inflació de la variància (VIF).

Després de la selecció de característiques, entrenem els següents algorismes d'aprenentatge automàtic: regressió logística, arbre de decisió i bosc aleatori. Per cada un, totes les mesures que reportem les obtenim a partir d'una validació encreuada de 5 parts. Optimitzem els hiperparàmetres de cada algorisme de classificació utilitzant la validació encreuada imbricada, per no sobreajustar l'etapa de selecció del model. Reportem la mesura F1 mitjana i la desviació estàndard a través de les 5 parts del millor model (d'acord amb la validació encreuada imbricada).

5 RESULTATS

A aquesta secció, presentem els nostres resultats experimentals i proporcionem respostes a les nostres qüestions de recerca originals. Primerament, mesurem i discutim l'exactitud de la predicció de cada conjunt de característiques, per determinar com de bé podem predir el suport a Trump i quina mena d'interacció és més predictiva. En segon lloc, analitzem les característiques més predictives, per esquematitzar els principals trets que distingeixen futurs partidaris de Trump a Reddit.

⁵<https://github.com/JoanMG/reddit-data>

⁶<https://textblob.readthedocs.io>

Taula 1: Per cada algorisme i per cada conjunt de característiques, reportem la puntuació F1 (%) i la seva desviació estàndard sobre la validació encruada de 5 parts. Els tres algorismes utilitzats són la regressió logística, l'arbre de decisió i el bosc aleatori. Per una descripció detallada de cada conjunt de característiques, vegeu la secció 5.1.

	RL		AD		BA	
	F1 (%)	σ	F1 (%)	σ	F1 (%)	σ
Participació	34,8	0,7	31,8	0,5	33,7	0,7
Puntuació	29,5	1,2	31,0	1,7	33,7	1,0
Interacció	26,7	0,7	26,3	1,0	25,5	0,6
Sentiment	7,3	0,8	16,4	13,4	10,7	13,1
Bossa de paraules	25,9	1,0	13,1	10,7	23,1	0,6
Punt. (biscades)	29,0	0,8	29,5	0,8	29,8	0,9
Int. (biscades)	25,4	0,6	26,9	0,8	24,6	1,5
Int. + Part.	34,7	1,3	31,5	0,7	33,8	0,8
Int. + Punt.	30,4	1,0	30,6	1,8	33,6	0,9
Part. + Punt.	35,3	0,9	32,3	0,7	35,0	0,6
Int. + Part. + Punt.	35,5	1,2	32,0	0,7	35,2	0,8

F1 de la classificació bàsica aleatòria: 15,2%

5.1 Exactitud de les prediccions

Els resultats per cada conjunt de característiques i classificador estan resumits a la taula 1. Primer, observem que la regressió logística supera els resultats dels altres dos algorismes en la majoria de casos, encara que hi ha algunes excepcions –característiques basades en la puntuació, el sentiment i les interaccions biscades– que discutim als següents paràgrafs. Ara comparem el poder predictiu de cada conjunt de característiques mirant-ne la puntuació F1 aconseguida.

Homofília. La participació és la característica amb més bons resultats d'entre els conjunts bàsics; aconseguix una puntuació F1 mitjana del $34,8\% \pm 0,7$. Aquest resultat suggereix que l'homofília és el predictor més potent del suport a Donald Trump entre els que hem considerat: el rol dels grups socials compartits supera en poder predictiu les característiques basades en interaccions directes, la rebuda social, la bossa de paraules i el sentiment. Aquest resultat confirma la importància de l'homofília com a determinant del comportament social [7]. Mostrem quins grups temàtics específics són més predictius del suport a Trump a la secció 5.2.

Rebuda social. Les puntuacions de Reddit obtenen una puntuació F1 del $33,7\% \pm 1,0$, gairebé tan alt com la participació. Remarquem que, per separar tant com fos possible la participació i les puntuacions, hem agafat la puntuació mitjana de la població als subreddits on un usuari concret no havia participat. Per tant, una puntuació tan alta ens suggereix un rol rellevant de la rebuda social i la conformitat: els individus que van ser benvinguts positivament o negativament en certes comunitats aterren a r/The_Donald quatre anys més tard. Mirem a quines comunitats hi ha aquest efecte a la secció 5.2. La independència de les puntuacions i la participació es confirma amb l'increment de la puntuació F1 quan s'utilitzen

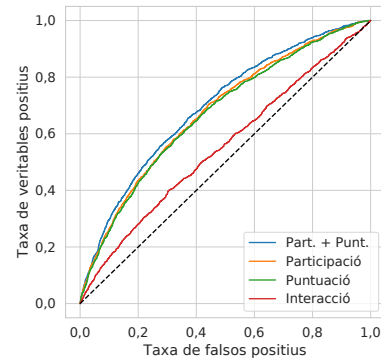


Figura 1: Corbes ROC dels conjunts de característiques més predictius: participació, puntuacions, interaccions directes, i la combinació de la participació i les puntuacions. Només reportem el resultat obtingut pel millor algorisme d'entre la regressió logística i el bosc aleatori. Tots els classificadors utilitzen informació del 2012 per predir el suport a Trump el 2016.

tots dos conjunts de característiques junts, tal com mostrem al final d'aquesta secció.

Mentre que la regressió logística és el millor classificador pels altres conjunts de característiques, el bosc aleatori té un resultat més bo amb les característiques relacionades amb la puntuació. Tenint en compte que el bosc aleatori és un classificador no lineal, el seu avantatge ens suggereix una relació no lineal entre les puntuacions a Reddit i la versemblança de donar suport a Donald Trump.

Influència directa. L'efecte de les interaccions, amb una puntuació F1 del $26,7\% \pm 0,7$, sembla ser molt més baix que el de les puntuacions i la participació. Utilitzant un classificador bàsic aleatori proporcional amb les classes, obtenim una puntuació F1 del 15,2% (pròxim al 15,8%, la proporció de partidaris de Trump). Les interaccions directes són, per tant, encara un predictor més bo que l'aleatori. Investiguem en profunditat les correlacions descobertes a les interaccions directes utilitzant el conjunt de característiques d'interaccions biscades al final d'aquesta secció i analitzant quines característiques són les més importants a la secció 5.2.

Llengua. Finalment, les característiques lingüístiques tenen un resultat bastant pobre. El sentiment, amb una puntuació F1 del $16,4\% \pm 13,4$ és tan predictiu com el classificador aleatori i qualsevol classificador més complex que un arbre de decisió acaba sobrestimant. En altres paraules, no observem cap correlació entre el to en escriure i la versemblança de tornar-se partidari de Trump. Les característiques de la bossa de paraules tenen un resultat més bo, però amb un $25,9\% \pm 1,0$ de puntuació F1 són molt pitjors que la participació i pitjors que la interacció. Aquest resultat ens suggereix que els models lingüístics simples són més mal predictors del trumpisme que els grups socials comuns.

Característiques combinades. Ara, mesurem el poder predictiu de parells de conjunts de característiques utilitzats conjuntament: participació i puntuacions, participació i interaccions, i interaccions i puntuacions. Els resultats mostren que, primerament, afegir el conjunt de característiques d'interacció a qualsevol altre no en millora el poder predictiu. Els resultats per participació i interaccions són

Taula 2: Per cada un dels conjunts de característiques més predictius, en reportem la precisió, el reclam, la puntuació F1 i l'àrea sota la corba ROC. Només reportem el resultat obtingut pel millor algorisme entre la regressió logística i el bosc aleatori. Tots els classificadors utilitzen una validació encreuada de 5 parts i informació del 2012 per predir el suport a Trump el 2016.

	Precisió	Reclam	F1	AUC
Participació	0,25	0,56	0,34	0,68
Puntuació	0,24	0,60	0,33	0,67
Interacció	0,18	0,52	0,26	0,55
Part. + Punt.	0,27	0,56	0,35	0,70

els mateixos que els de participació. I per interaccions i puntuacions també són els mateixos que els d'únicament puntuacions. Aquests resultats enforteixen la nostra conclusió que les interaccions directes a Reddit no són un factor decisiu en determinar qui es torna un partidari de Trump quatre anys després. En canvi, quan combinem participació i puntuacions, els resultats milloren lleugerament comparant-ho amb els millors dels dos. Aquest fet ens suggereix que aquestes dues menes d'interaccions ens proporcionen uns senyals parcialment ortogonals. Els senyals més importants que trobem són, per tant, l'homofília i la rebuda social, mentre que trobem només un efecte limitat de la influència social. Combinar participació i puntuació constitueix, doncs, el nostre millor classificador basat en característiques socials.

Analitzem en detall els resultats d'aquest model anterior a l'hora de predir el suport a Trump amb quatre anys d'antelació. Aquest model obté una precisió del 27% i un reclam del 56%. Recordem que la fracció de partidaris de Trump al nostre grup de focus és el 15,8%. Agafant la probabilitat assignada pel millor classificador a cada usuari obtenim una puntuació indicant la propensió d'un usuari de Reddit de tornar-se partidari de Trump. Avaluem el poder predictiu d'aquesta puntuació de propensió amb la corba ROC a la figura 1. L'àrea sota la corba ROC per aquest model és 0,70. Reportem aquests resultats, juntament amb els dels models de participació, puntuacions i interaccions agafats individualment a la taula 2.

Característiques bisecades. Ara centrem l'atenció a les característiques bisecades. Recordem que amb *bisecades* ens referim a dividir els subreddits dins d'un cert conjunt de característiques (puntuacions o interaccions) en dos grups, en funció de si un subreddit té una fracció de futurs partidaris de Trump superior (T) o inferior (N) que la mitjana. Per això, aquestes característiques contenen informació futura, no disponible originalment el 2012, però tenen una granularitat més grollera. Ens permeten d'investigar l'efecte de la influència dels (futurs) usuaris partidaris de Trump en contrast amb la resta, tant per la influència directa com per la rebuda social. Primer, en mesurem els resultats en termes de precisió de la predicció, mirant a la taula 1. Les interaccions bisecades obtenen uns resultats semblants als de les interaccions dividides per subreddits. Aquesta troballa ens suggereix que l'efecte de la influència social és bastant semblant entre els subreddits amb predomini de partidaris de Trump. Sorprenentment, en canvi, les puntuacions perden poder predictiu. Aparentment, la granularitat més grollera fa menys precís

Taula 3: Coeficients de la regressió logística per predir el suport a Trump, per totes les característiques d'interaccions bisecades. Indiquem amb T el conjunt de subreddits amb més partidaris de Trump que la mitjana i amb N el conjunt dels que en tenen menys que la mitjana.

Descripció de la característica	β
Interaccions amb usuaris que participen en T	0,076163
Interaccions no conflictives amb usuaris que participen en T	-0,005322
Interaccions no conflictives amb usuaris que participen en N	-0,029029

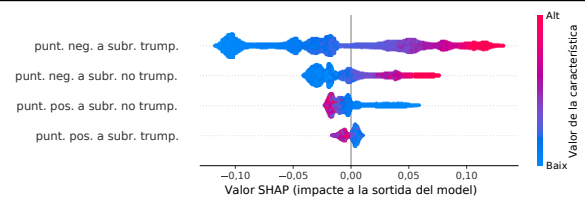


Figura 2: Valors SHAP per totes les característiques de puntuacions bisecades. Indiquem amb *punt. pos.* les característiques obtingudes de les puntuacions positives i *punt. neg.* les de les puntuacions negatives. Per cada característica, s'indiquen en vermell els valors més alts i en blau els més baixos. A la dreta, tenim els valors de característiques més associats amb el suport a Trump.

el classificador. Aquest resultat ens mostra que l'efecte de la rebuda social d'una certa comunitat no és senzillament un reflex de si la comunitat tindrà més o menys predomini de partidaris de Trump, sinó que hi ha una estructura més fina.

Analitzem en profunditat les característiques pels dos models bisecats (puntuacions i interaccions directes) per caracteritzar més quins tipus d'interaccions anticipen el suport a Donald Trump.

Mirem primer els coeficients de la regressió logística per diferents característiques d'interaccions bisecades. Aquí tenim tres característiques, en funció de si la interacció és conflictiva o no, i en funció de si involucra un subreddit amb un nombre alt o baix de futurs partidaris de Trump. Utilitzar aquesta mena d'informació futura ens permet cercar evidències d'efecte backfire. La taula 3 mostra que tenir *qualsevol* interacció directa amb subreddits que tindran un predomini de partidaris de Trump és predictiu del suport a Trump. A més, les interaccions conflictives (independentment de l'objectiu) es correlacionen amb el suport a Trump, tal com mostra el coeficient negatiu de les interaccions no conflictives. Aquesta troballa és una manifestació del comportament conflictiu que els partidaris de Trump tenen a Internet, més que no pas un efecte backfire. Aquesta interpretació és consistent amb anàlisis prèvies [26] i d'acord amb els resultats que mostrem al següent paràgraf.

En segon lloc, analitzem els resultats del conjunt de característiques de puntuacions bisecades. A aquest conjunt de característiques, dividim els subreddits en dos grups, en funció del nombre de futurs partidaris de Trump. Per tant, considerant puntuacions positives i negatives, tenim quatre característiques. Com que el millor classificador per aquest conjunt de característiques és el bosc aleatori, utilitzem SHAP, un algorisme de l'estat de la qüestió per explicar les

característiques a models de bosc aleatori [20, 21]. Aquests valors poden ser interpretats de forma semblant als coeficients β de la regressió logística. Cada punt representa un usuari i , per tant, per cada característica, la figura mostra la distribució dels valors SHAP a través del conjunt de dades. Horitzontalment, distribucions més amples indiquen un impacte absolut més gran de la característica en la classificació total, mentre que el color de cada punt (de blau a vermell) codifica el valor de la característica (baix o alt). Una característica amb valors alts corresponents a valors SHAP positius (a la dreta) es correlaciona positivament amb el trumpisme. Per contra, una característica amb valors alts corresponents a valors SHAP negatius (a l'esquerra) es correlaciona negativament amb el suport a Trump. Reportem els valors SHAP a la figura 2.

Els resultats són interessants: les puntuacions negatives a subredits amb una presència de futurs partidaris de Trump superior a la mitjana s'associen a un futur suport a Trump. Semblaria que, per tant, el desafiament de normes de grups socials que anticipa el suport a Trump també és present a les comunitats més alineades amb el trumpisme. És consistent amb altres troballes de comportaments "tols" dels partidaris de Trump [26].

5.2 Trets predictius

A aquesta secció, investiguem la importància de cada un dels nostres models, per respondre la nostra última qüestió de recerca: quins trets van anticipar el desenvolupament del suport a Trump? Per fer-ho, realitzem una anàlisi de característiques en profunditat pels models més reeixits: la bossa de paraules, la participació, les puntuacions, les interaccions i el model combinat.

Tal com hem vist a la secció anterior, el millor algorisme de classificació és, en general, la regressió logística. Pels conjunts de característiques de puntuació, el bosc aleatori aconsegueix uns resultats semblants o millors, possiblement per la seva no linealitat. Per tant, a la nostra investigació de la importància de les característiques, analitzem les característiques del bosc aleatori quan hi ha involucrades les puntuacions i les de la regressió logística en cas contrari. Gràcies a la normalització descrita a la secció 4, per la regressió logística podem mirar senzillament als coeficients obtinguts per cada característica. En canvi, pels boscos aleatoris, emprem altre cop SHAP, un algorisme per explicar la sortida de models d'arbre ensemble [20, 21].

Característiques lingüístiques. El primer model que investiguem és el de la bossa de paraules. El model prova de capturar les diferències estadístiques en l'ús de paraules dels partidaris de Trump. La taula 4 reporta les paraules més discriminants. En general, aquestes característiques no són fàcilment interpretables, però podem destriar alguns patrons notables.

Els partidaris de Trump el 2012 eren més propensos a utilitzar la paraula *liberal* 'liberal' i la paraula *libertarian* 'llibertarià'. Podem conjecturar que el primer és un insult i el segon una autodescripció, però no hi ha cap manera directa de confirmar-ho mirant només el model. Malgrat això, podem veure alguna evidència confirmatòria en l'anàlisi de les característiques de participació. A més a més, utilitzen termes com *cop* 'policia', possiblement lligat a les visions de llei i ordre promogudes per Trump; i *home* 'casa', potser relacionada amb l'atenció pronunciada als conceptes com els valors familiars o la pàtria (*homeland* en anglès).

Taula 4: Coeficients de la regressió logística de les característiques més importants de la bossa de paraules. A l'esquerra, tenim les 10 característiques amb un coeficient β més gran i, a la dreta, les 10 característiques amb un coeficient β més petit.

Partidaris de Trump		No partidaris de Trump	
Paraula	β	Paraula	β
liberal	0,000784	abuse	-0,000399
guy	0,000691	reporter	-0,000345
debate	0,000650	similar	-0,000338
politic	0,000635	contribution	-0,000326
libertarian	0,000604	century	-0,000322
come	0,000604	honor	-0,000321
think	0,000593	palestinian	-0,000318
cop	0,000591	writer	-0,000314
tell	0,000587	context	-0,000313
home	0,000570	voting	-0,000306

Al costat oposat –les paraules menys utilitzades per partidaris de Trump el 2012– trobem termes vagament relacionats amb drets civils tals com *abuse* 'abús', *reporter* 'reporter', i la paraula *palestinian* 'palestí', possiblement reconeixent reclamacions dels palestins. Malgrat això, en general les característiques d'aquest costat són difícils d'interpretar. Ara veurem com, utilitzant el classificador basat en la participació, que és més predictiu, podem dibuixar un retrat més clar.

Característiques de participació. Hem vist que aquesta és la millor característica individual en termes d'exactitud de la predicció. La taula 5 mostra les 30 característiques més importants per cada una de les dues classes. Aquí, cada característica representa la participació (escriure un comentari) en aquest subreddit el 2012. Els coeficients del model són més grans que els de les característiques de la bossa de paraules.

Les característiques més discriminadores estan relacionades amb visions polítiques. Els grups conservador i llibertarià són els més correlacionats amb el suport a Donald Trump. Aquesta troballa és consistent amb la idea que la coalició de Trump és part de l'anomenat "autoritarisme llibertarià", on conflueixen necessitats de tots dos camps ideològics [4].

També reconeixem temes i comunitats que s'associen amb el suport a Trump. *r/conspiracy* és una comunitat dedicada a teories de la conspiració [15]. Per exemple, va cobrir extensivament l'engany del "pizzagate" sobre cercles d'abusos sexuals infantils operats per oficials del Partit Demòcrata. Aquesta observació recolza la teoria que alguns grups marginals s'han fusionat en el discurs polític majoritari [30].

El web 4chan, un fòrum de discussió "políticament incorrecte", s'ha vinculat al "moviment de dreta alternativa" en anàlisis prèvies [26]. Trobem que la participació en el subreddit *r/4chan* el 2012 és la quarta característica més predictiva d'aquest conjunt. Altres grups políticament incorrectes també es correlacionen amb el suport a Trump. Per exemple, *r/ImGoingToHellForThis* és una comunitat dedicada a l'humor xocant i àcid.

Taula 5: Coeficients de la regressió logística de les característiques més importants de participació. A l'esquerra, tenim les 30 característiques amb un coeficient β més gran i, a la dreta, les 30 característiques amb un coeficient β més petit.

Partidaris de Trump		No partidaris de Trump	
Subreddit	β	Subreddit	β
r/Conservative	0,3815	r/raspberry_pi	-0,2847
r/Libertarian	0,3740	r/TrueAtheism	-0,2577
r/conspiracy	0,3733	r/AskCulinary	-0,2355
r/4chan	0,3341	r/comics	-0,2249
r/circlejerk	0,3107	r/rpg	-0,2186
r/NoFap	0,2918	r/ireland	-0,2034
r/Entrepreneur	0,2539	r/Fantasy	-0,1983
r/ImGoingToHellForThis	0,2510	r/explainlikeimfive	-0,1944
r/trees	0,2482	r/environment	-0,1892
r/MensRights	0,2482	r/doctorwho	-0,1878
r/guns	0,2293	r/polyamory	-0,1806
r/blackops2	0,2110	r/scifi	-0,1777
r/runescape	0,2031	r/books	-0,1772
r/Anarcho_Capitalism	0,1937	r/askscience	-0,1738
r/Catholicism	0,1931	r/london	-0,1691
r/leagueoflegends	0,1920	r/britishproblems	-0,1687
r/nfl	0,1843	r/Homebrewing	-0,1632
r/starcraft	0,1714	r/programming	-0,1521
r/CCW	0,1638	r/gadgets	-0,1501
r/breakingbad	0,1631	r/AndroidQuestions	-0,1463
r/investing	0,1624	r/listentothis	-0,1462
r/AdviceAnimals	0,1589	r/hiphopheads	-0,1397
r/DeadBedrooms	0,1577	r/boardgames	-0,1336
r/Firearms	0,1551	r/asoiaf	-0,1292
r/Advice	0,1537	r/whatisthisthing	-0,1244
r/seduction	0,1518	r/lgbt	-0,1187
r/Christianity	0,1455	r/criangepics	-0,1175
r/golf	0,1453	r/ukpolitics	-0,1136
r/mylittlepony	0,1437	r/Python	-0,1089
r/POLITIC	0,1423	r/baseball	-0,1080

Alguns interessos i aficions emergeixen clarament entre els subreddits més predictius del suport de Trump, mentre d'altres semblen que s'anticorrelacionen amb el suport a Trump. Un interès en les armes de foc es correlaciona fortament amb el trumpisme (r/guns, r/Firearms, r/CCW). Això mateix és cert per diverses comunitats de videojocs (r/blackops2, r/runescape, r/leagueoflegends, r/starcraft). En canvi, altres aficions són anticorrelacionades, per exemple, jocs de taula (r/boardgames, r/rpg). Cuinar i aficions *do it yourself* són entre les més importants: r/raspberry_pi, r/AskCulinary, r/Homebrewing estan fortament anticorrelacionades amb el suport a Trump. Interessos en la literatura i l'art és un predictor igualment important (r/books, r/comics, r/ListenToThis, r/Fantasy, r/scifi).

La religió també és central a la separació: entre les correlacionades amb el suport a Trump trobem r/Catholicism i r/Christianity.

Entre les anticorrelacionades, en canvi, una de les més predictives és r/TrueAtheism. Aquesta troballa és consistent amb la idea que, per molts americans, Trump era “una defensa simbòlica de l'herència cristiana percebuda dels Estats Units” [35].

Algunes de les comunitats correlacionades amb el suport a Trump estan relacionades a interessos com l'emprenedoria i la inversió. Això ens podria suggerir tant el suport de gent rica com de qui té una actitud *self-made*. L'amenaça de l'estatus (oposada a les penúries econòmiques) s'ha identificat com un tret comú en el suport a Trump [29].

Diversos subreddits amb demografies predominantment masculines apareixen entre els correlacionats amb el suport a Trump, consistent amb troballes anteriors [3]. Entre ells, r/MensRights, centrar en la defensa dels interessos masculins contra el feminisme. D'un punt de vista de l'orientació sexual, observem una clara divisió entre els subreddits associats a Trump i els anticorrelacionats. Aquest darrer grup inclou minories de gènere, sexuals i romàntiques, com r/polyamory i r/lgbt. Els subreddits més positivament correlacionats amb Trump són majoritàriament masculins: per exemple, r/seduction, un subreddit part del moviment Pick-Up Artists;⁷ r/NoFap, un grup que proporciona autoajuda per l'addicció a la pornografia; i el ja citat r/MensRights. Val la pena adonar-se que r/DeadBedrooms, que descriu a si mateix com “un grup de suport pels reditaires que s'enfronten a una relació que no té intimitat sexual”, és també entre els més associats al suport a Trump.

D'entre els subreddits restants al grup, molts s'associen a la cultura popular (en tots dos costats), tal com esports o programes de televisió. Altres subreddits apareixen anticorrelacionats amb el suport a Trump senzillament perquè normalment s'associen a usuaris de Reddit no americans: aquest és el cas de r/ukpolitics, r/london, r/ireland, r/britishproblems. Una troballa curiosa és que un dels millors predictors del suport a Trump és r/trees, un subreddit pels entusiastes del cànnabis. Sospitem d'un possible factor de confusió: per exemple, Miech et al. [27] mostren que, als Estats Units, el consum diari de cànnabis en joves de 19 - 24 anys és tres vegades més alt entre aquells que no van a la universitat (13% respecte al 4%). Això és consistent amb troballes anteriors que Trump ha atret més suport del segment de la població amb menys educació [32].

Rebuda social. Ara centrem l'atenció a les característiques de la rebuda social. Tal com hem dit, com que el millor model per aquest conjunt de característiques és el bosc aleatori, emprem SHAP [20, 21] per explicar les relacions apreses pel model. La figura 3 reporta els valors SHAP resultants.

Alguns dels subreddits que havíem trobat que la seva participació era un fort predictor del suport a Trump també apareixen aquí, encara que amb un format diferent: les puntuacions negatives a r/Conservative, r/trees i r/conspiracy es correlacionen amb la falta de suport a Trump, mentre que puntuacions negatives a r/atheism es correlacionen amb el suport a Trump.

Al subreddit r/NFL, observem una anticorrelació entre puntuacions positives i suport a Trump. Com que aquest és un subreddit que també apareix entre les característiques més importants de participació, aquest resultat ens suggereix que la participació en el

⁷<https://www.dailydot.com/irl/ken-hoinsky-pua-reddit-seduction-book-the-ga-me>

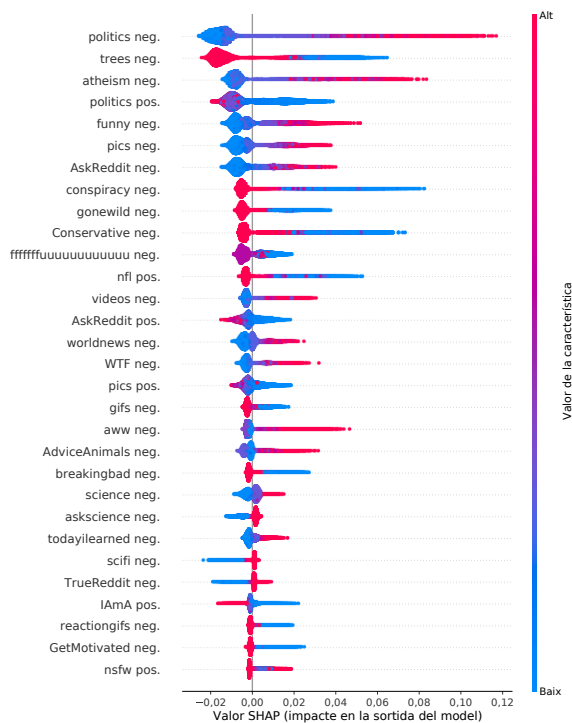


Figura 3: Valors SHAP per les 30 característiques més importants de puntuació. Indiquem amb pos. les característiques obtingudes de les puntuacions positives i neg. les de les negatives. Per cada característica, el vermell indica els valors més alts i el blau els més baixos. A la dreta, tenim els valors de les característiques més associades al suport de Trump. Per exemple, la primera fila indica que una puntuació negativa elevada a r/politics és indicativa de suport a Trump.

subreddit sense ser apreciat per la comunitat és un predictor del suport a Trump.

Alguns subreddits generalistes, com r/funny, r/pics o r/AskReddit també hi apareixen. En tots aquests casos, les puntuacions negatives s'associen amb el suport a Trump, igual que a r/politics. Recordem que la participació en aquests subreddits no és entre les característiques més importants. Aquestes observacions ens suggereixen que una rebuda negativa per comunitats de Reddit majoritàries i àmplies el 2012 es vincula amb el suport a Trump el 2016.

Això podria ser també el cas de r/gonewild, un subreddit que s'autodescriu com “un lloc per redditaires adults de ment oberta que ensenyen els seus cossos nus per diversió”: els usuaris que obtenen una rebuda negativa a aquesta comunitat són més propensos a donar suport a Trump quatre anys més tard.

Influència directa. El nostre tercer conjunt de característiques bàsiques representa la interacció directa entre un usuari i un altre usuari, en què aquest darrer participa en un cert subreddit. També té en compte quantes d'aquestes interaccions no són conflictives. La taula 6 mostra les característiques més predictives. Malgrat el seu poder predictiu escàs si ho comparem amb la participació, encara

Taula 6: Coeficients de la regressió logística per les característiques més importants d'interacció. A l'esquerra, tenim les 10 característiques amb un coeficient β més gran i, a la dreta, les 10 amb un coeficient β més petit. Indiquem amb “r dist.” la fracció d'interaccions al subreddit r i, amb “r pos.,” la fracció d'interaccions positives entre totes les interaccions amb el subreddit r.

Partidaris de Trump		No partidaris de Trump	
Característica	β	Característica	β
r/ShitPoliticsSays pos.	0.1513	r/todayilearned dist.	-0.0698
r/Republican pos.	0.0868	r/TrueReddit dist.	-0.0584
r/conspiracy dist.	0.0684	r/Futurology dist.	-0.0548
r/moderatepolitics pos.	0.0637	r/dataisbeautiful dist.	-0.0348
r/Conservative dist.	0.0563	r/GaryJohnson dist.	-0.0214
r/Libertarian dist.	0.0543	r/PoliticalDiscussion dist.	-0.0168
r/Libertarian pos.	0.0457	r/Liberal dist.	-0.0149
r/conspiracy pos.	0.0416	r/PoliticalDiscussion pos.	-0.0120
r/POLITIC pos.	0.0274	r/worldnews dist.	-0.0116
r/Economics pos.	0.0219	Nombre total d'interaccions	-0.0105

podem utilitzar aquestes característiques per enriquir el nostre retrat.

El suport a Trump es pot predir per la fracció d'interaccions positives a subreddits políticament actius com r/Republican, r/Libertarian i r/moderatepolitics, així com comunitats que discuteixen temes d'interès pels partidaris de Trump tals com r/conspiracy i r/Economics. Aquests trets concorden amb les nostres anàlisis prèvies i confirmen la idea que el llibertarianisme i el conservadorisme són entre les arrels del trumpisme. Malgrat això, també observem que el nombre d'interaccions amb r/GaryJohnson, candidat contra Trump a les eleccions de 2016, s'anticorrelaciona amb el suport a Trump. La característica més potent d'aquest conjunt és la fracció d'interaccions positives a r/ShitPoliticsSays. Aquest subreddit allotja crítiques i escarni d'altres subreddits i exhibeix visions dretanes.⁸

Finalment, trobem que el nombre total d'interaccions directes s'anticorrelaciona amb el suport a Trump, suggerint que la influència total de Reddit és adversa a Trump.

Característiques combinades. Finalment, la figura 4 mostra les característiques més importants del model combinat que utilitza la participació i les puntuacions. Els dos conjunts de característiques estan ben equilibrats: tots dos conjunts apareixen entre les característiques més predictives (14 a 16). Aquesta observació enforteix la hipòtesi que la rebuda social i l'homofília proporcionen senyals ortogonals diferents a l'hora de predir el suport a Trump.

6 CONCLUSIONS I TREBALLS FUTURS

Hem mirat predictors de convertir-se en un partidari de Donald Trump a Reddit. Hem utilitzat dades del 2012 per predir la participació en r/The_Donald el 2016, que ens ha servit de representació del suport a Trump. Aquesta tasca de predicció és un bon repte, per l'abast temporal de quatre anys (un cicle electoral presidencial dels EUA) entre les dades observades i el comportament objectiu. Tanmateix, el nostre millor model aconsegueix una AUC del 0,70 i

⁸Per exemple, denuncia r/Fuckthealtright i r/AgainstHateSubreddits com a subreddits hostils.

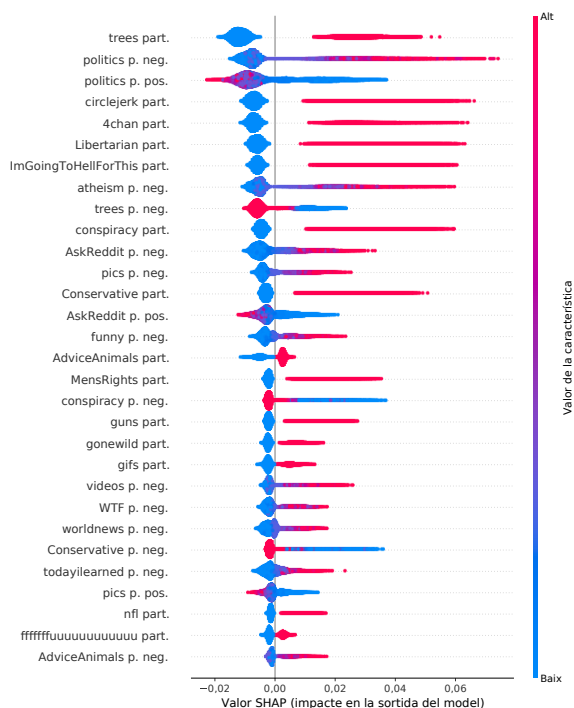


Figura 4: Valors SHAP de les 30 característiques més importants dels conjunts de característiques de participació i puntuació combinats. Indiquem amb *p. pos.* les característiques obtingudes de les puntuacions positives i amb *p. neg.* les de les puntuacions negatives; i amb *part.* les característiques de participació. A la dreta, tenim els valors de les característiques més associats al suport a Trump.

una mesura F1 del 0,36, significativament per sobre dels resultats d'un classificador bàsic aleatori.

Hem explorat diversos conjunts de predictors que representen tres hipòtesis sociològiques pel suport a Trump: l'homofília, la rebuda social i la influència. Hem operacionalitzat cada hipòtesi al context de Reddit mirant la participació d'un usuari en una comunitat (un subreddit), l'aprovació que reben els seus missatges a una comunitat donada i interaccions amb usuaris d'altres comunitats. Comparades amb altres característiques lingüístiques interpretables bàsiques, tals com la bossa de paraules i el sentiment de les publicacions, les característiques socials resulten ser més predictives del comportament objectiu. En particular, les característiques que codifiquen l'homofília i la rebuda social (conformitat i anticonformitat) han demostrat ser els millors predictors del trumpisme, mentre que la influència social ha demostrat una rellevància limitada. A més a més, la combinació de característiques d'homofília i rebuda social (és a dir, la participació i les puntuacions) obté uns resultats lleugerament millors que les característiques soles, demostrant així que els dos senyals són complementaris d'alguna manera.

Finalment, fem una introspecció de les característiques dels millors models per delinear una 'persona' de com semblava a Reddit el 2012 el típic partidari de Trump el 2016. El típic partidari de Trump té punts de vista conservadors i llibertarians, és mal rebut pels corrents polítics majoritaris, és religiós i en conflicte amb l'ateisme,

i té interessos en les armes, les conspiracions, l'emprenedoria i contingut políticament incorrecte. Per contra, el típic usuari que no dona suport a Trump és ateu, LGBT-friendly i té interessos culinaris, literaris i tecnològics.

Limitacions i treballs futurs. L'operacionalització de les teories sociològiques que hem considerat en aquest estudi té, necessàriament, l'oportunitat d'introduir-hi distorsions. Dels tres conjunts de característiques, les d'interacció, que codifiquen la influència social, són les més fràgils per ser naturalment esparses. Hem contrarestat aquesta característica agregant-les per comunitat, però encara han resultat ser les menys predictives dels nostres models. Aquest resultat pot venir de les tries específiques del disseny i cal més treball per quantificar el rol que la influència social juga a l'hora de canviar les actituds polítiques de la gent a les xarxes socials.

El conjunt de característiques de puntuació, que codifiquen la rebuda social, també presenta alguns reptes, ja que la distribució de les puntuacions té una cua pesada. Al nostre treball, hem utilitzat un classificador lineal (el bosc aleatori) per abordar aquest problema, però algorismes més sofisticats poden millorar els resultats.

Més fonamentalment, el disseny de l'estudi actual no ens permet de diferenciar entre diferents interpretacions causals de l'efecte de la rebuda social. Utilitzem tres variables per representar el comportament dels partidaris de Trump: la rebuda social observada, el suport a Trump observat i l'actitud política latent. Per altra banda, un model causal podria veure la rebuda social com una causa del canvi en les actituds polítiques, que al seu torn causi el suport a Trump. En aquest cas, la rebuda social és una *causa* de les arrels del suport a Trump. Per exemple, un usuari pot tenir una experiència negativa amb comunitats polítiques majoritàries, que causi que les seves actituds derivin cap a posicions més extremes, que al seu torn expliquin el suport a Trump. Per altra banda, les actituds polítiques latents podrien ser una causa comuna tant de la rebuda social, perquè de bon principi les opinions expressades no van en la línia de la comunitat, com del suport a Trump. En aquest segon cas, la rebuda social és un *efecte* de les actituds polítiques i el suport a Trump en depèn d'una forma no causal. Per exemple, un usuari pot tenir algunes actituds marginals que són mal rebudes per la comunitat política majoritària i trobar una sortida natural al trumpisme. Una investigació causal d'aquestes hipòtesis a partir de dades observacionals és una extensió interessant del treball actual [31]. Amb aquest marc, podríem formalitzar factors de confusió; entenent, per exemple, si els partidaris de Trump van començar a participar més en alguns subreddits polítics o si ja eren usuaris més actius inicialment. Malgrat això, el nostre treball constitueix un primer pas necessari abans de cap investigació causal.

Finalment, descrivim la 'persona' d'un partidari de Trump assumint que n'hi ha només una. Però hi ha evidències que gent provinent de múltiples estrats sociodemogràfics li donen suport [23].⁹ Així doncs, és possible que la persona descrita sigui una amalgama de trets provinents de diferents orígens. En aquest cas, construir múltiples personae podria crear retrats més acurats. També, ajudaria a distingir els partidaris de Trump a Reddit d'altres republicans joves dels EUA. Aquesta anàlisi podria ajudar a entendre quines

⁹<https://fivethirtyeight.com/features/the-mythology-of-trumps-working-class-support>

qüestions van atraure els qui es van polititzar així i, per tant, donar més idees de les arrels del trumpisme.

REFERÈNCIES

- [1] Sara Ahmadian, Sara Azarshahi, and Delroy L Paulhus. 2017. Explaining Donald Trump via communication style: Grandiosity, informality, and dynamism. *Personality and Individual Differences* 107 (2017), 49–53.
- [2] J. Baumgartner, S. Zannettou, B. Keegan, M. Squire, and J. Blackburn. [n.d.]. The Pushshift Reddit Dataset. *arXiv preprint arXiv:2001.08435* [n. d.].
- [3] Mick Brewer. 2019. *From the Ground, to the Ballot, to the System: The (Critical) Interpersonal Reproduction of Masculinity within Homosocial Friendships of Male Donald Trump Supporters*. Southern Illinois University at Carbondale.
- [4] Wendy Brown. 2018. Where the fires are. *Soundings* 68, 68 (2018), 14–25.
- [5] Robert B Cialdini and Noah J Goldstein. 2004. Social influence: Compliance and conformity. *Annu. Rev. Psychol.* 55 (2004), 591–621.
- [6] M. de Wit, A. Roman-Alcalá, A. Liebman, and S. Chrisman. 2019. Agrarian origins of authoritarian populism in the United States: What can we learn from 20th-century struggles in California and the Midwest? *Journal of Rural Studies* (2019).
- [7] Daniel DellaPosta, Yongren Shi, and Michael Macy. 2015. Why do liberals drink lattes? *Amer. J. Sociology* 120, 5 (2015), 1473–1511.
- [8] Trevor Diehl, Brian E Weeks, and Homero Gil de Zuniga. 2016. Political persuasion on social media: Tracing direct and indirect effects of news use and social interaction. *new media & society* 18, 9 (2016), 1875–1895.
- [9] M Fitzduff. 2017. *Why irrational politics appeals: understanding the allure of Trump*. ABC-CLIO.
- [10] Claudia I Flores-Saviaga, Brian C Keegan, and Saiph Savage. 2018. Mobilizing the Trump train: Understanding collective action in a political trolling community. In *Twelfth International AAAI Conference on Web and Social Media*.
- [11] Venkata Rama Kiran Garimella and Ingmar Weber. 2014. Co-following on Twitter. In *Proceedings of the 25th ACM conference on Hypertext and social media*. 249–254.
- [12] Ted Grover and Gloria Mark. 2019. Detecting Potential Warning Behaviors of Ideological Radicalization in an Alt-Right Subreddit. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. 193–204.
- [13] Benjamin Kane and Jiebo Luo. 2018. Do the Communities We Choose Shape our Political Beliefs? A Study of the Politicization of Topics in Online Social Groups. In *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 3665–3671.
- [14] David Karpf. 2017. Digital politics after Trump. *Annals of the International Communication Association* 41, 2 (2017), 198–207.
- [15] Colin Klein, Peter Clutton, and Adam G Dunn. 2019. Pathways to conspiracy: The social and linguistic precursors of involvement in Reddit's conspiracy theory forum. *PloS one* 14, 11 (2019).
- [16] Timothy La Fond and Jennifer Neville. 2010. Randomization tests for distinguishing social influence and homophily effects. In *Proceedings of the 19th international conference on World wide web*. 601–610.
- [17] Mirko Lai, Viviana Patti, Giancarlo Ruffo, and Paolo Rosso. 2018. Stance evolution and twitter interactions in an italian political debate. In *International Conference on Applications of Natural Language to Information Systems*. Springer, 15–27.
- [18] Yu-Ru Lin, Drew Margolin, Brian Keegan, and David Lazer. 2013. Voices of victory: A computational focus group framework for tracking opinion shift in real time. In *Proceedings of the 22nd international conference on World Wide Web*.
- [19] Xiaodan Lou, Alessandro Flammini, and Filippo Menczer. 2019. Information pollution by social bots. *arXiv preprint arXiv:1907.06130* (2019).
- [20] Scott M. Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. 2020. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence* 2, 1 (2020), 2522–5839.
- [21] Scott M. Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., 4765–4774.
- [22] Walid Magdy, Kareem Darwish, Norah Abokhodair, Afshin Rahimi, and Timothy Baldwin. 2016. #ISISisNotIslam or #DeportAllMuslims? Predicting Unspoken Views. In *Proceedings of the 8th ACM Conference on Web Science (WebSci '16)*. ACM, 95–106. <https://doi.org/10.1145/2908131.2908150>
- [23] Jeff Manza and Ned Crowley. 2017. Working class hero? Interrogating the social bases of the rise of Donald Trump. In *The Forum*, Vol. 15. De Gruyter, 3–28.
- [24] Miller McPherson, Lynn Smith-Lovin, and James M Cook. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology* 27, 1 (2001), 415–444.
- [25] Alexey N Medvedev, Renaud Lambiotte, and Jean-Charles Delvenne. 2017. The anatomy of Reddit: An overview of academic research. In *Dynamics on and of Complex Networks*. Springer, 183–204.
- [26] William Merrin. 2019. President Troll: Trump, 4Chan and Memetic Warfare. In *Trump's media war*. Springer, 201–226.
- [27] Richard Miech, Lloyd Johnston, Patrick O'Malley, Jerald Bachman, John Schulenberg, and Megan Patrick. 2019. Monitoring the future national survey results on drug use, 1975-2018: volume I, secondary school students.
- [28] Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. Semeval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. 31–41.
- [29] Diana C Mutz. 2018. Status threat, not economic hardship, explains the 2016 presidential vote. *Proceedings of the National Academy of Sciences* 115, 19 (2018).
- [30] Rishab Nithyanand, Brian Schaffner, and Phillipa Gill. 2017. Online political discourse in the Trump era. *arXiv preprint arXiv:1711.05303* (2017).
- [31] Judea Pearl. 2009. *Causality*. Cambridge university press.
- [32] Jonathan T Rothwell and Pablo Diego-Rosell. 2016. Explaining nationalist political views: The case of Donald Trump. *Available at SSRN 2822059* (2016).
- [33] Ryne A Sherman. 2018. Personal values and support for Donald Trump during the 2016 US presidential primary. *Personality and Individual Differences* 128 (2018).
- [34] Marionna Taulé, M Antonia Martí, Francisco M Rangel, Paolo Rosso, Cristina Bosco, Viviana Patti, et al. 2017. Overview of the task on stance and gender detection in tweets on Catalan independence at IberEval 2017. In *2nd Workshop on Evaluation of Human Language Technologies for Iberian Languages, IberEval 2017*, Vol. 1881. CEUR-WS, 157–177.
- [35] Andrew L Whitehead, Samuel L Perry, and Joseph O Baker. 2018. Make America Christian again: Christian nationalism and voting for Donald Trump in the 2016 presidential election. *Sociology of Religion* 79, 2 (2018), 147–171.
- [36] R Willis. 1963. Two dimensions of conformity-nonconformity. *Sociometry* (1963).
- [37] Savvas Zannettou, Tristan Caulfield, Jeremy Blackburn, Emiliano De Cristofaro, Michael Sirivianos, Gianluca Stringhini, and Guillermo Suarez-Tangil. 2018. On the origins of memes by means of fringe web communities. In *Proceedings of the Internet Measurement Conference 2018*. 188–202.