

Hacia una exploración de las representaciones sociales en torno al big data

Gastón Becerra ^{1,2}, Juan Pablo López-Alurralde ¹

¹ Universidad Abierta Interamericana – Centro de Altos Estudios en Tecnología Informática, Ciudad Autónoma de Buenos Aires, Argentina

² Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, Argentina

gaston.becerra@sociales.uba.ar

Resumen. Compartimos los primeros resultados de una investigación en curso cuyo objetivo es explorar las representaciones sociales del big data. Para ello se usó un cuestionario dirigido compuesto por la técnica de evocación libre de palabras a una muestra de 247 estudiantes y graduados universitarios argentinos de distintas carreras. Entre las preguntas que exploramos se encuentran: ¿qué ideas se evocan en la representación social del big data y cómo se estructuran? ¿qué temas se pueden distinguir, por la coocurrencia de evocación en una respuesta, y cuál es la orientación o valoración, en tanto indicador de actitud hacia el big data? ¿cuáles son las fuentes de consulta acerca del big data?

1 Introducción

En este trabajo compartimos los primeros resultados de una investigación en curso, cuyo objetivo es explorar las representaciones sociales en torno del big data, junto a otros fenómenos complejos en los que confluyen cuestiones tecnológicas, sociales y epistémicas, tales como la inteligencia artificial, la irrupción de las ciencias de datos, y los fenómenos 2.0-5.0.

Este interés surge como una línea complementaria a una investigación más amplia que busca problematizar el fenómeno del big data desde la perspectiva de los sistemas sociales y el constructivismo de Niklas Luhmann [1, 2]. En dicho enclave teórico habíamos asumido el objetivo de buscar el significado social del big data en las diversas formas en que se lo tematiza desde distintos sistemas, tales como la política, la economía, la ciencia, los medios de comunicación, entre otros.

Aquí consideramos un espacio complementario: el sentido común. Nos orientamos por desarrollos de la psicología social, particularmente, por la teoría de las representaciones sociales [3, 4]. Por “representación social” entendemos a una forma de conocimiento de carácter práctico, orientado hacia la comunicación, la comprensión y el dominio del entorno social, material e ideal [5]. Las representaciones se construyen activamente ante la necesidad de dar sentido a fenómenos novedosos que nos interpelan diferencialmente en vistas de nuestra posición en el entramado de relacio-

nes sociales [6–8]. En este sentido, la representación social se vuelve particularmente relevante para el estudio de desarrollos científicos que pudieran tener un impacto social, siendo el estudio de Moscovici sobre el psicoanálisis y su recepción en Francia un ejemplar paradigmático [9]. Particularmente, aquí hemos adoptado el modelo estructural de las representaciones sociales propuesto por J. C. Abric [10], que postula que toda representación social se compone, por un lado, de un núcleo central que aúna los sentidos más comunes y estables de la representación, y por el otro, de elementos periféricos que concretizan la representación y la vinculan a los distintos contextos específicos de su uso.

Aquí presentamos al big data como un fenómeno interesante para ser tratado en términos de representación social. Su relevancia se desprende de que es uno de esos fenómenos tecnológicos, sociales y epistémicos en torno a los cuales se ha desarrollado una “retórica” que augura una profunda transformación de todos los aspectos de la vida social. Por mencionar sólo uno de los espacios alcanzados, podemos referirnos a la economía, donde actores como *The Economist* afirman que los datos constituyen el “nuevo petróleo”¹, o como cuando *IBM* señala que “hoy todo está hecho de datos” (IBM, 2014)². Como señalamos en otro trabajo centrado en el tratamiento del big data por parte de la prensa digital argentina [11], en el centro de este discurso se encuentra una “promesa” de carácter epistémico: que los grandes datos habilitarían formas superiores de inteligencia y conocimiento, que su exploración permitiría arribar a verdades más objetivas o con mayor grado de precisión y previsión, y que la cuantificación de todos los aspectos del comportamiento humano es clave para desarrollar soluciones novedosas [12, 13]. Se debe notar que si bien esta promesa se encuentra claramente expresada en los discursos “promocionales” del big data, también hay algo de ella en los discursos más pesimistas que ven en el big data un renovado mecanismo de control social [14, 15]. En cualquier caso, la promesa no tendría asidero si no estuviera sustentada en una “premisa” que pudiera presumirse evidente: que vivimos en un mundo donde hay vastas cantidades de datos disponibles, y que los mismos pueden ser técnicamente manipulados para darles sentido. Diferentes estudios sociales críticos del big data han puesto de manifiesto los variados mecanismos retóricos en la comunicación usual de esta premisa y esta promesa, tales como esconder el carácter de producto humano de los datos al equipararlos con recursos naturales que pueden ser “minados” o “colectados”, o esconder las diferencias en las posibilidades reales en torno al acceso y la explotación de los datos, o incluso negar su carácter propietario [16, 17].

Hasta donde conocemos, no hay trabajos empíricos que hayan explorado la representación social del big data. En este trabajo adelantamos algunos resultados preliminares de una investigación en curso con dicho objetivo. Nuestras preguntas son:

RQ1. ¿Qué ideas se evocan en la representación social del big data? ¿Cómo se organizan y estructuran?

¹ <https://www.economist.com/briefing/2017/05/06/data-is-giving-rise-to-a-new-economy> (consultado Junio/2020)

² https://www.youtube.com/watch?v=QCgZrOUd_Dc (consultado Junio/2020)

RQ2. ¿Qué temas se pueden distinguir en la representación del big data? ¿Cuál es su orientación o valencia?

RQ3. ¿Cuáles son las fuentes de consulta acerca del big data? ¿Cómo lo representan estudiantes de distintas carreras universitarias?

2 Método

En este primer estudio exploratorio indagamos la representación social del big data por parte de estudiantes y graduados universitarios de Argentina a través de la técnica de asociación libre de palabras [18–20]. Decidimos hacer foco en estudiantes universitarios, asumiendo que las nuevas tecnologías que nos interesan pueden ser materia de tratamiento distinto en las diferentes carreras. En cualquier caso, se debe ser consciente de este recorte para la generalización de los datos recabados.

2.1 Instrumentos

La recolección se realizó a través de un formulario online³ basado en la técnica de evocación o asociación libre de palabras (EVOC o TALP) y un control deslizante para indicar una valoración para cada término introducido. El formulario tenía por default 5 campos para ingresar palabras, aunque se podían agregar otros. El estudio se difundió por redes sociales.

La consigna fue la siguiente: *Por favor, indicanos qué palabras o frases te vienen a la mente cuando pensás en "[estímulo]". También te pedimos que, por favor, nos indiques si estas ideas que acabas de introducir se corresponden con algo que valoras positivamente (algo que te agrada) o negativamente (algo que te desagrada), utilizando el deslizador debajo de cada palabra.*

Como uno de los objetivos de la investigación es ubicar al big data en un campo de fenómenos de relevancia tecnológica, social y epistemológica más amplia, este formulario fue programado para rotar los términos estímulos en la siguiente colección (por orden de probabilidad de aparición): *big data, inteligencia artificial, ciencia de datos, conocimiento*. En cada consulta el formulario muestra sólo uno de estos términos para evitar condicionamientos. En este trabajo, nos centramos sólo en los resultados relativos al big data, no haciendo comparaciones entre corpus.

Finalmente, también se incluían preguntas sociodemográficas, de formación (carrera), y referidas a las fuentes consultadas para informarse sobre el tema en cuestión.

³ Disponible en <https://evoc-67321.web.app/>

Investigación nuevas tecnologías

Tema: Inteligencia artificial

Un grupo de estudio e investigación de la Universidad de Buenos Aires, Universidad Abierta Interamericana, y la Universidad Nacional de Quilmes está llevando a cabo una breve encuesta acerca de nuevas tecnologías y los sentidos que estas adquieren en la sociedad.

Esta encuesta está destinada a **estudiantes o graduados recientes de terciarios y universidades, de cualquier carrera y nivel de Argentina.**

Responderla lleva menos de 3 minutos y lo podés hacer tanto desde tu celular como desde una computadora. Las respuestas son anónimas y los resultados serán utilizados exclusivamente para fines académicos.

1 Test de evocación

Por favor, indicanos qué palabras o frases te vienen a la mente cuando pensás en **"Inteligencia artificial"**

También te pedimos que, por favor, nos indiques si estas ideas que acabas de introducir se corresponden con algo que valoras positivamente (algo que te agrada) o negativamente (algo que te desagrada), utilizando el deslizador debajo de cada palabra.

Palabra 1: Palabra 1 _____

Negativo —●— Positivo

Palabra 2: Palabra 2 _____

Fig. 1. Captura del instrumento (test de evocación).

2.2 Muestra

La muestra fue de tipo intencional, compuesta por estudiantes y graduados universitarios de Argentina (N=246), con edades comprendidas entre 17 y 70 años (M=28,6; SD=7,5); 74,7% de mujeres, 24% de hombres y 1,3% de otros. Más del 55,5% son estudiantes o graduados de ciencias sociales y empresariales (incluyendo psicología, la carrera que más estudiantes aporta en la muestra), el 17,5% de humanidades, el 15% de naturales y exactas, el 6% de ingenierías y tecnologías, y finalmente el 6% restante de médicas y de la salud.

2.3 Pre-procesamiento

El pre-procesamiento del corpus de respuestas (todos los estímulos) consistió en varios pasos: 1) se anotaron los términos con UPOS, lo que permitió trabajar con lemmas, remover símbolos, puntos y números, así como también adverbios; 2) se reemplazaron caracteres especiales y acentos; 3) se convirtió el corpus a minúscula. En el caso de que la respuesta fuese una frase o varios términos, se unieron (los que no fueron filtrados en pasos anteriores) por el símbolo “_”.

Al comenzar el pre-procesamiento se contaba con un vocabulario de 1.655 términos únicos (para todos los estímulos), con cerca de 1.300 términos que aparecían una sola vez, y cuyo término más frecuente tenía 70 repeticiones. Al terminar, el vocabulario se redujo a poco más de 1.200 términos únicos, con cerca de 800 términos que aparecían una sola vez, y cuyo término más frecuente tenía 155 repeticiones. La base de términos específicos del big data incluye 590 términos únicos, donde el término

más frecuente tiene 110 repeticiones, y hay cerca de 400 términos mencionados una sola vez, delineando una distribución que sigue la ley de Zipf. Para todos los análisis siguientes se fijó un umbral mínimo de 5 repeticiones.

Todas las tareas de pre-procesamiento, análisis, y visualización fueron programas en R [21]. Los datos se estructuraron, siguiendo principios “tidy”, y fueron manipulados con herramientas del Tidyverse [22]. Para los análisis factoriales se reutilizó código del proyecto DTM-VIC [23].

3 Resultados

Antes de continuar, debemos advertir que se trata de una investigación recién iniciada y que los límites de la muestra nos fuerzan a que la interpretación de los resultados no pretenda más que elaborar hipótesis tentativas y conjeturas para futuros trabajos.

3.1 Ideas evocadas en la representación social del big data

Para explorar la estructura del campo representacional del big data, objeto de la **RQ1**, realizamos un “análisis prototípico” de las evocaciones [18, 19, 24], tendiente a distinguirlas entre centrales y periféricas. Para ello se ordenó el corpus de los términos evocados de acuerdo con la frecuencia de aparición y el orden medio de evocación, y luego se dividieron 4 segmentos, utilizando como criterio de corte las medianas.

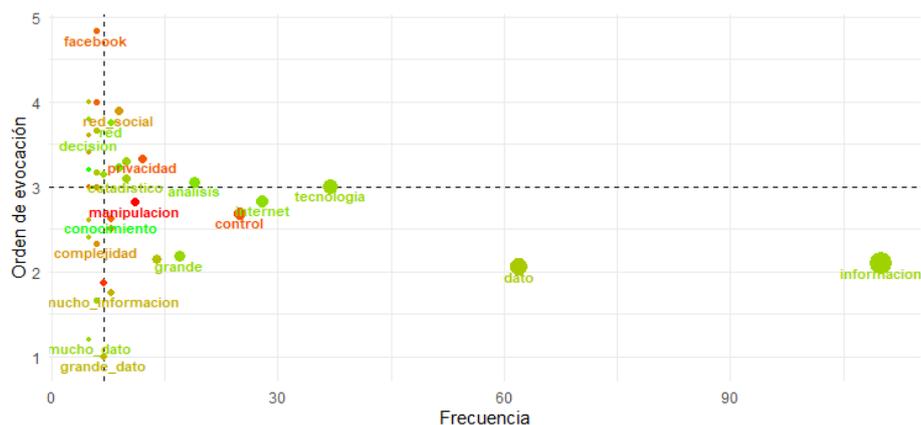


Fig. 2. Términos evocados para “big data”, por orden medio de evocación y frecuencia. El color denota la valoración media, en una escala de rojo (1/10) a verde (10/10).

El primer segmento comprende a las palabras enunciadas más frecuentemente y de más rápida evocación, es decir, las más comunes y accesibles. Estos son los tér-

minos que se puede decir que constituyen el núcleo de la representación: *información, dato, internet, control, grande, bases de datos, manipulación, conocimiento, marketing, poder, computadora, espionaje, y grandes datos*. Mayormente refieren a lo que denominamos la “premisa” del big data: la disponibilidad de grandes bases de datos, plausibles de otorgar información. Es interesante señalar que el término *grande* (lemma en el que convergen distintos términos, tales como *grandes, mucho*) es el único en el núcleo que remite a las famosas 3 V’s –volumen, velocidad y variedad– que popularizara Laney [25] y que se referencia en prácticamente todas las noticias que buscan introducir al big data a un público masivo por parte de la prensa digital argentina (Becerra, 2019). Sin embargo, también se debe aclarar que algunas de estas ideas están presentes en varias palabras compuestas de menor frecuencia, y que otro pre-procesamiento las podría haber puesto en un lugar más predominante.

Mayormente estos términos están asociados con una valoración positiva, como muestra la **fig. 3**, siendo *conocimiento* la palabra de mayor valor medio con 9.85/10, presente en este segmento. Sin embargo, también aparecen en este mismo segmento, las 3 palabras de menor valoración en todo el corpus: *manipulación* (1.4/10), *espionaje* (1.8/10) y *control* (2.4/10). También se observan otras palabras, tal vez más ambivalente, como *marketing* o *poder*. La distribución de estas valoraciones es indicativa de la polaridad que recubre el campo semántico del big data [26], en tanto la imagen de la “gran oportunidad” que ofrece su promesa epistémica, convive con una retórica de sospecha frente a los riesgos de su explotación para el sometimiento y el condicionamiento de otros, como un “gran hermano”.

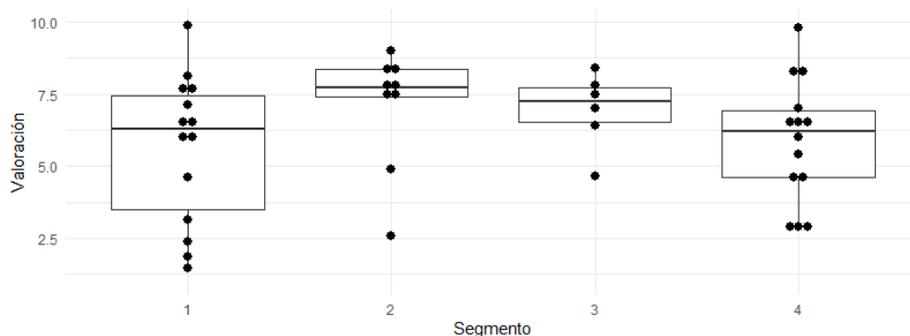


Fig. 3. Valoraciones por segmento.

El segundo segmento comprende a las palabras que se encuentran en la periferia del núcleo central, aquí definidas por tener una frecuencia superior a la mediana, pero también un orden de evocación mayor, indicando que fueron enunciados entre los últimos términos de cada respuesta. Aquí se cuenta con los términos *tecnología, análisis, privacidad, estadístico, informática, investigación, red social, red* y *almacenamiento*. Si bien son ideas muy generales, podemos conjeturar que refieren al anclaje del big data que lo vincula con su costado más mundano y material, y vinculado con su carácter de producto o servicio. La idea de *tecnología* podría referir al campo material y físico que soporta al big data, mientras que la de *análisis*, al proce-

so de transformación activa de los datos para darles sentido por parte de aquellos grupos abocados al big data. No casualmente, ambas ideas suelen enfrentarse en la retórica del big data con la idea de *magia*, que encubren esfuerzos y costos [27]. La valoración de estas palabras sigue, en líneas generales, la del primer segmento, aunque con una tendencia mayor hacia una connotación positiva.

El tercer y último segmento a considerar incluye palabras que se evocaron rápidamente pero que no tienen una frecuencia suficiente como para ser consideradas parte del sentido común. En términos de la teoría, podrían remitir a los sentidos propios de grupos específicos: *complejidad*, *nube*, *importante*, *mucho*, *mucho dato*, *número*.

3.1 Temas en la representación social del big data

El análisis anterior, tributario del modelo de núcleo y periferia, puede ser útil para dar una idea de la organización jerárquica de la representación, pero no nos permite aclarar de qué manera se relacionan entre sí las evocaciones. Así, uno podría preguntarse si acaso los términos de valoración más negativa en nuestro corpus (por ejemplo, *control*, *manipulación*, *espionaje*) se dan generalmente juntos o, si, por el contrario, se observan distribuidos con otros términos de valoración más positiva. Si algunos términos se dan juntos se podría conjeturar que esto responde a su pertenencia a uno o más “temas”, es decir, a una estructura recurrente e iterativa en la comunicación [28]. Este es el objeto de nuestra **RQ2**.

Para empezar a explorar esta pregunta se analizaron las correlaciones entre las palabras en el nivel de las respuestas, es decir, palabras que se evocaron juntas. Con este dato se construyó una red, y luego se detectaron automáticamente comunidades o agrupamientos (ilustradas con la forma del punto en la **fig. 4**) ($K=6$). Nos interesa evaluar si estos subgrupos se corresponden con términos de valoraciones similares, lo que podría ser indicador de la actitud hacia el big data.

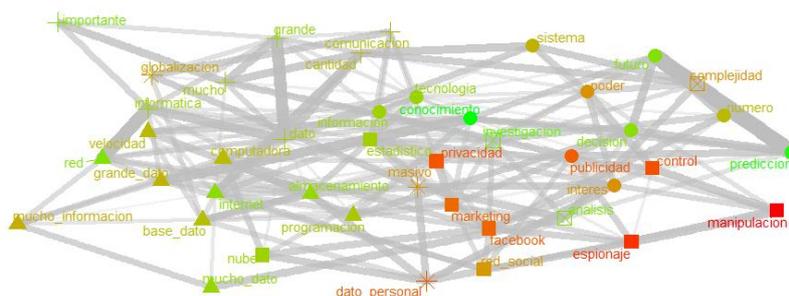


Fig. 4. Correlación entre palabras en la misma respuesta. La forma del punto remite a la comunidad. El color del punto y término denota la valoración media, en una escala de rojo (valoración baja de 1/10) a verde (valoración alta de 10/10).

La primera comunidad de palabras corresponde a términos como *predicción, futuro, conocimiento, número, o decisión* cuyas valoraciones son de las más altas, junto con otras menos valoradas pero que de todas formas mantienen a la comunidad en una valoración positiva. Mayormente son términos que parecen ubicar al big data en su reclamo epistémico.

Con una valoración similar se encuentran otras 3 comunidades que incluyen, en el primero a los términos *red, grandes datos, internet, muchos datos, almacenamiento, bases de datos*, en el segundo a los términos *informática, grandes datos, cantidad, mucho, importante, y comunicación*, y finalmente a los términos *análisis, complejidad e investigación*. En estos grupos conviven nociones cercanas a las mencionadas 3 V's, con otros relativos a la computación, mixtura que no debería sorprendernos, en tanto que, como recuerda Diebold [29], la noción de "big data" se popularizó originalmente dentro del sector informático como un término para referir a los desafíos que se comenzaban a plantear en materia de almacenamiento y consulta de grandes volúmenes de información, y que luego se plasmaron en las 3 V's popularizadas por los medios de comunicación masivas.

Los últimos dos grupos incluyen términos que podrían referir al trasfondo "social" del big data. El primero incluye a los términos *datos personales, masivo y globalización*, los cuales denotan la naturaleza social y personal de los datos. Su valoración media tiende a una connotación negativa (4.3/10). Esta tendencia se profundiza más en la última comunidad (con una valoración media cercana a 3.8/10), donde los términos se vinculan directamente con los riesgos del big data, tanto en un sentido más inespecífico como en aplicaciones particulares: *espionaje, manipulación, control, marketing, privacidad, nube, Facebook, estadístico, red social*.

3.3 Fuentes de difusión y consulta del big data

De acuerdo con Moscovici [9], algunas representaciones sociales han tenido difusores que, con diferentes lógicas, han podido adaptar el objeto de representación a una cierta imagen de su público. Los medios de comunicación masivos como, por ejemplo, la prensa con sus líneas editoriales y secciones especializadas, suelen ser canales de difusión de estas representaciones. En otro trabajo, referido al tratamiento del big data por parte de la prensa digital argentina [11], describimos de qué manera su presentación difiere en contextos tan distintos como su presentación a un público masivo e inespecífico, su tematización dentro de las noticias relativas al trabajo, el empleo o la formación, o incluso a su incorporación en la agenda política.

En el presente trabajo nos limitamos a indagar acerca de los medios consultados para informarse acerca del big data, y nos preguntamos específicamente por el aporte de las carreras universitarias.

Con respecto a las fuentes, y como se observa en la siguiente tabla, los participantes de todas las carreras consultan mayormente sitios web; quienes más se informan del big data por medio de su carrera son los estudiantes y graduados de naturales y exactas, así como de las ingenierías y tecnologías, aunque muchos más consultan posteos y sitios; quienes más consultan diarios y sitios de noticias son los ciencias

médicas y de la salud, y sociales y empresariales, ambos en un porcentaje apenas menor que a la consulta de posteos y sitios.

Tabla 1. % de casos que afirmaron consultar un medio, por grupo de carreras. * Cursos refiere a formación por afuera de la carrera; ** Formación refiere a alguna instancia de la carrera. Hum = Humanidades; Ing. = Ingenierías y tecnologías; Med. = Médicas y de la salud; Nat. = Naturales y exactas; Soc. = Sociales y empresariales.

	Hum.	Ing.	Med.	Nat.	Soc.
Cursos*	35%	43%	9%	39%	43%
Diarios y noticias	33%	29%	43%	34%	46%
Formación**	42%	50%	29%	45%	39%
Posteos y sitios	63%	79%	43%	71%	49%

Una forma posible de indagar la incidencia de estos medios en la representación social es por medio de un análisis de correspondencia que permita representar en forma simultánea los términos y los grupos de carrera.

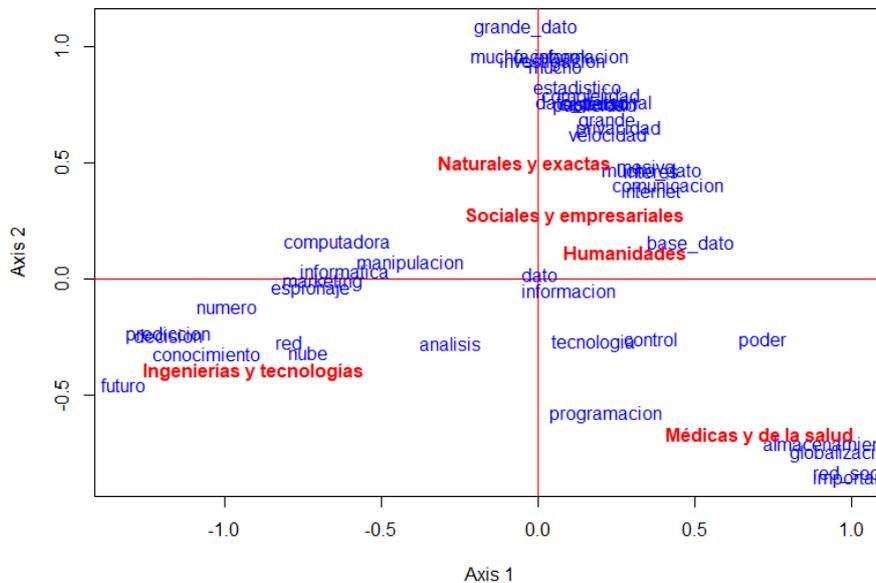


Fig. 5. Análisis factorial de términos y grupos de carreras.

En el centro del gráfico observamos a los términos más salientes del núcleo central (*dato e información*). El grupo de términos más cercano coincide con el grupo de carreras vinculadas a las humanidades, las ciencias sociales y empresariales, y las naturales y exactas. Las principales diferencias se encuentran entre las carreras vin-

culadas a las ingenierías y tecnologías, y las médicas y de la salud. Si volvemos a la tabla de fuentes consultadas, vemos que entre estas dos carreras se dan los contrastes más fuertes en tanto a la relevancia de la carrera para formarse en big data, o al interés dado a los diarios y sitios de noticias. Si observamos la superposición con los términos podemos ver además que las ingenierías y tecnológicas se vinculan mayormente con los términos del discurso de mayor valoración del big data, tales como *futuro*, *conocimiento* o *predicción*, mientras que las de las ciencias médicas y de la salud tienen una valoración menor y refieren mayormente a una de sus dimensiones sociales.

4 Conclusiones

En este trabajo presentamos resultados muy preliminares de una nueva línea de investigación, dentro de las coordenadas de la psicología social, que tiene por centro la representación social del big data. Dado el estado muy inicial del relevamiento, aquí nos limitamos a compartir algunos primeros análisis y algunas conjeturas para exploración más profunda en trabajos siguientes.

También hemos ensayado una primera combinación de análisis clásicos de esta tradición de psicología social, con algunas técnicas menos exploradas, como el uso de grafos para construir discursos, sin perder de vista el análisis estructural. Todos estos análisis se han implementado en lenguaje R, y se espera ponerlos a disposición de la comunidad por medio de un *package*.

Referencias

1. Becerra G (2018) Big data como objeto de estudio y método para la investigación empírica en sociología y psicología social. In: 47 Jornadas Argentinas de Informática & Simposio Argentino de Tecnología y Sociedad. Sociedad Argentina de Informática, Buenos Aires
2. Becerra G (2018) Interpelaciones entre el Big data y la Teoría de los sistemas sociales. Propuestas para un programa de investigación. *Hipertextos* 6:41–62
3. Wagner W, Hayes N (2011) El discurso de lo cotidiano y el sentido común. La teoría de las representaciones sociales. *Anthropos*, Barcelona
4. Moscovici S (2001) Why a theory of social representations?
5. Jodelet D (1985) La representación social: fenómenos, conceptos y teoría. In: Moscovici S (ed) *Psicología Social II*. Paidós, Barcelona, pp 17–40
6. Marková I (2017) The making of the theory of social representations. *Cad da Pesqui* 47:358–374
7. Wagner W (1998) Social Representations and beyond: Brute Facts, Symbolic Coping and Domesticated Worlds. *Cult Psychol* 4:297–329. <https://doi.org/10.1177/1354067X9800400302>
8. Howarth C (2006) A social representation is not a quiet thing: Exploring the critical potential of social representations theory. *Br J Soc Psychol* 45:65–86. <https://doi.org/10.1348/014466605X43777>

9. Moscovici S (1979) El psicoanálisis, su imagen y su público. Huemul, Buenos Aires
10. Abric J-C (1993) Central System, Peripheral System: Their functions and roles in the dynamics of social representations. *Pap Soc Represent* 2:75–78
11. Becerra G (2019) La construcción del big data en la prensa digital argentina. In: XIII Jornadas de Sociología. Universidad de Buenos Aires, Buenos Aires
12. boyd D, Crawford K (2012) Critical Questions for Big Data. *Information, Commun Soc* 15:662–679
13. Kitchin R (2014) Big Data, new epistemologies and paradigm shifts. *Big Data Soc* 1:
14. van Dijck J (2014) Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveill Soc* 12:197–208
15. Byung-Chul H (2014) Psicopolítica. Neoliberalismo y nuevas técnicas de poder. Herder, Barcelona
16. Portmess L, Tower S (2015) Data barns, ambient intelligence and cloud computing: the tacit epistemology and linguistic representation of Big Data. *Ethics Inf Technol* 17:1–9. <https://doi.org/10.1007/s10676-014-9357-2>
17. Puschmann C, Burgess J (2014) Metaphors of Big Data. *Int J Commun* 8:1690–1709
18. Dany L, Urdapilleta I, Lo Monaco G (2014) Free associations and social representations: some reflections on rank-frequency and importance-frequency methods. *Qual Quant* 49:489–507. <https://doi.org/10.1007/s11135-014-0005-z>
19. Lo Monaco G, Piermattéo A, Rateau P, Tavani JL (2017) Methods for Studying the Structure of Social Representations: A Critical Review and Agenda for Future Research. *J Theory Soc Behav* 47:306–331. <https://doi.org/10.1111/jtsb.12124>
20. Barreiro A, Gaudio G, Mayor J, et al (2014) La justicia como representación social: Difusión y posicionamientos diferenciales. *Rev Psicol Soc* 29:319–345. <https://doi.org/10.1080/02134748.2014.918821>
21. Team R Core (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
22. Wickham H (2019) Welcome to the tidyverse. *J Open Source Softw* 4:. <https://doi.org/https://doi.org/10.21105/joss.01686>
23. Lebart L, Piron M (2013) Práctica del análisis de los datos numéricos y textuales con Dtm-Vic. INDEC - Instituto Nacional de Estadística y Censos, Buenos Aires
24. Wachelke J, Wolter R (2011) Critérios de construção e relato da análise prototípica para representações sociais. *Psicol Teor e Pesqui* 27:521–526. <https://doi.org/10.1590/S0102-37722011000400017>
25. Laney D (2001) 3D Data Management: Controlling Data Volume, Velocity, and Variety
26. Paganoni MC (2019) Framing big data : a linguistic and discursive approach. palgrave macmillan, Cham
27. Elish MC, Boyd D (2018) Situating methods in the magic of Big Data and AI. *Commun Monogr* 85:57–80. <https://doi.org/10.1080/03637751.2017.1375130>
28. Luhmann N (2000) The reality of the mass media. Stanford University Press, California
29. Diebold FX (2012) The Origin(s) and development of “ Big Data ”: the phenomenon , the term , and the discipline. *Penn Econ Work Pap*. <https://doi.org/10.2139/ssrn.2202843>