Received: 1 December 2020 | Revised: 31 March 2021 | Accepted: 30 April 2021

DOI: 10.1002/jcop.22603

RESEARCH ARTICLE

JOURNAL of COMMUNITY PSYCHOLOGY WILEY

Examining the concept of equity in community psychology with natural language processing

Jonathan P. Scaccia PhD

Research and Evaluation, Dawn Chorus Group, Reading, Pennsylvania, USA

Correspondence
Jonathan P. Scaccia, PhD, Research and Evaluation, Dawn Chorus Group, 1014 Hartman Rd, Reading, PA 19606, USA.
Email: jon@dawnchorusgroup.com

Abstract

Large amounts of text-based data, like study abstracts, often go unanalyzed because the task is laborious. Natural language processing (NLP) uses computer-based algorithms not traditionally implemented in community psychology to effectively and efficiently process text. These methods include examining the frequency of words and phrases, the clustering of topics, and the interrelationships of words. This article applied NLP to explore the concept of equity in community psychology. The COVID-19 crisis has made pre-existing health equity gaps even more salient. Community psychology has a specific interest in working with organizations, systems, and communities to address social determinants that perpetuate inequities by refocusing interventions around achieving health and wellness for all. This article examines how community psychology has discussed equity thus far to identify strengths and gaps for future research and practice. The results showed the prominence of community-based participatory research and the diversity of settings researchers work in. However, the total number of abstracts with equity concepts was lower than expected, which suggests there is a need for a continued focus on equity.

KEYWORDS
community-based participatory research, equity, natural language processing, synthesis, text-mining

Research into the medical and social sciences indicates that it can take up to 17 years before a scientific finding makes its way into regular practice (Baker, 2001; Balas & Boren, 2000; Colditz & Emmons, 2012; Estabrooks et al., 2018; Morris et al., 2011). This fact is troubling. There are tens of thousands of academic journals filled with literally millions of great ideas and notable findings. So why are not these used? Furthermore, how does this hurt a public reliant on public health interventions?

The reality is that staying up to date with any topic can be fraught with intimidation and frustration. A specific set of skills and expertise are required to make sense of large swaths of research literature. Because of this barrier, many great ideas do not make it into practice. Consequently, community practitioners do not use the best strategies and do not know the most current knowledge.

Natural language processing (NLP) provides multiple methods to parse through large amounts of literature quickly. NLP is a related set of computer-based techniques that treat text as data. It extracts meaning and relationships from the text by analyzing the frequencies and interrelationships between words.

NLP can be beneficial because it can more effectively and efficiently process literature. However, these methods may be unfamiliar to community psychologists. Recent community-specific methodology books (Jason & Glenwick, 2016) make no mention of NLP, though they address text mining broadly. A cursory review of several journals (*The American Journal of Community Psychology*, *The Journal of Community Psychology*, *The Global Journal of Community Psychology Practice*, and *The Australian Community Psychologist)* yielded no articles that use NLP. Moreover, a review of the past 10 years of SCRA biennial programs showed no abstract using NLP. This is a methodological gap because these techniques are becoming more mainstream in fields adjacent to community psychology like health care (Raja et al., 2008).

The manuscript has two purposes. First, this manuscript demonstrates how NLP can be used to synthesize large amounts of literature. Although the methods discussed here may seem complicated, they can be implemented in an *out-of-the-box* format, meaning only slight modifications are needed to the underlying mathematics.

Second, this manuscript looked at how community psychology has discussed and researched equity as a use case. Even before *The Age of COVID*, data from multiple studies and sources have elucidated the enormous disparities in health and wellness in our society (Chambers, 2020; Radley et al., 2020). There are systemic challenges in communities that perpetuate the differences in outcomes (English et al., 2020; Plough, 2018). COVID has shown that these disparities show up in emergent conditions, not just chronic conditions (Berkowitz et al., 2020; Radley et al., 2020). These ongoing findings suggest that equity needs to be a central focus in all community-based research (Reid et al., 2019).

Equity is a broad term that encompasses both a process and a goal. Equity is a *process* when it refers to values like inclusiveness, participation, engagement, and outreach to describe how social service researchers and practitioners can approach an intervention or evaluation. Equity is also a *goal* involving achieving wellness outcomes regardless of race, zip code, and other social determinants that negatively impact health trajectories.

Community psychology has an ongoing goal of improving health and wellness for all. As such, the field is prepared to work toward equity and work equitably (Wolfe, 2014). Therefore this article takes stock of the community psychology-specific literature to see what equity research has already been conducted and what gaps remain. With community psychology values so aligned with equity, there was potentially a large amount of literature available. Any literature search can quickly become unfeasible and impractical when sorting through vast quantities of abstracts.

The specific research questions underneath this aim include:

1. What are the major areas of research around equity-related issues in the published community psychology literature?
2. How are key concepts related within the community psychology research space?

1 | METHODS

1.1 | Data

This inquiry focused on five journals: *The American Journal of Community Psychology*, *The Journal of Community Psychology*, *The Global Journal of Community Psychology Practice*, *The Australian Community Psychologist*, and *The American Journal of Orthopsychiatry*. Other community psychology-adjacent journals, like *Health Promotion and Practice*, were not included. The search was limited to journals indexed in PubMed, which excluded potential outlets like *The Community Psychologist*. PubMed's application programming interface (API) was used to query and extract all article abstracts and relevant metadata.

There was a simple filter: all abstracts that mentioned *equity* or *disparities*, using the wildcards "equit|dispari." This search translates to "return all abstracts containing the strings 'equit' or 'dispari' regardless of what proceeds or follows those strings." There were no constraints on the publication date. Other adjacent terms could have been used to broaden the search. However, because research question two specifically targeted how community psychology potentially talks about related concepts, the search terms were short and simple.

This search process yielded 155 unique articles. All abstracts were retrieved on November 16, 2020. All the data is publicly available, either from the author, on the specific journal websites, or by replicating the PubMed search process.

1.2 | Analyses

The search results were in a database format and required a fair amount of preprocessing. Stopwords were removed first. These are words that add little informational value to the overall data set. Examples of stopwords are *about*, *of*, *this*, *these*, *and*, *so*, *on*. Next, the words were lemmatized. This process involves reducing a word down to its base form. In many cases, this involves making plural words singular. In others, it consists of transforming irregular English verbs. Lemmatization is a more sophisticated algorithm than simple word-stemming but is more computationally expensive. However, the data set was small enough that this tradeoff was not an issue.

For the first research question: *What are the major areas of research around equity-related issues in the published community psychology literature?*, there were four major analytic strategies. The most basic is *Bag of Words (BOW)*, which treats each word (or token) as informative in and of itself because it carries information. Results were computed for both unigram (one-word) and bigram (two-word) distributions. This type of approach is likely very familiar to readers in the form of *word clouds*, which use font size to demonstrate frequency. The core assumption of BOW is that if a word frequently occurs (except for stopwords), it is likely to be more important.

Topic Modeling determines clusters of words within these abstracts. The specific method, latent Dirichlet allocation (LDA), treats a *topic* as a cluster of words. Each *document* (the abstracts) contains multiple topics. Topic modeling takes a prespecified hyperparameter (k, the number of topics) as the initial input.

However, what is the correct number of topics? This particular problem demonstrates the importance of human judgment in interpreting NLP results. From an optimization perspective, the goal is to find the number of topics (k) that minimize the perplexity score. Perplexity represents how likely new data (text) is, given the data that the model has seen before (Kapadia, 2019). However, while higher ks may show continued perplexity improvements, the emergent topics may not be interpretable. LDA is not a natural language understanding (NLU) algorithm. It merely takes words and their relative frequency to one another and mathematically determines the most likely grouping. Therefore, human judgment is needed to determine whether a specific number of topics balances interpretability and specificity. This general process is very similar to "finding the elbow" in a scree plot following a principal components analysis (PCA).

An LDA returns a matrix, where each row is an abstract, and each column is a topic from 1 to k. The values for each topic are the γ-statistic (gamma): the likelihood that a particular article "belongs" to each topic. Each abstract was then assigned to a topic where the γ-statistic was the highest.

Next, the most representative words for each topic were generated. This analysis required two techniques: extractive summarization and term-frequency, inverse-document frequency. *Extractive summarization* is a text summarization method that determines the sentences that best represent a topic by measuring how similar sentences are to one another. For each topic, all the abstracts for that topic were condensed into a single block of text. The five most representative sentences for each topic were then computed. Extractive summarization approaches can sacrifice some readability but generally perform adequately, as indicated by their longevity in the field (Erkan & Radev, 2004).

The unique words for those topics were determined using the extractive summarization output using *term-frequency, inverse-document frequency* (*tf-idf*). *Term frequency (tf)* is how often a word occurs across a collection of documents. Under a general BOW assumption, the frequency might be helpful. However, some words may occur so much that they lose their distinguishing power. For example, the word *health* might show up a lot across these abstracts and appear in multiple topics.

Inverse document frequency (idf) is a statistic that measures the relevant frequency of a word across all documents. A common word would have an *idf* score around zero, while a rare word would score closer to one. By multiplying these two metrics together (*tf-idf*), researchers can see how important a word is to a particular set of documents. *Tf-idf* provides a more refined statistic than simple frequencies. It yields a word's relative frequency when correcting for spread across documents.

Word embeddings were used for the second research question, *How are key concepts related within the community psychology research space? Word Embeddings* are numerical representations of words encoded by a vector. Think of a document as a string of words that appear one after another in sequence. By examining the words that occur near one another in a sequential space, it is possible to capture semantic meaning by computing their co-occurrence with one another and then implementing a dimensionality reduction process. What makes this vectorization process so powerful is how it encodes information about the words that may not be apparent by just looking at a table of values.

There are many examples of "word algebra" that show how word embeddings preserve deep semantic encoding. There is a famous, almost cliched example in large pre-trained word embeddings. By taking the vector for *king*, subtracting the vector for *man*, then adding the vector for *woman*, the result is the vector for *queen* (Moody, 2015).

There are multiple methods to develop word vectors. As noted above, *transfer learning* models like *word2vec* and *GLoVE* use large, pre-trained vectors with an enormous number of parameters (Azunre, 2020). OpenAI's recently released GPT-3 has billions of parameters (Floridi & Chiriatti, 2020). However, one potential challenge in using transfer learning models is that domain-specific words and jargon may not be captured. For example, "culture" may refer to the community or society-level norms and values rather than the organizational-level factors that come up often in the implementation sciences.

A novel set of vectors was trained for this study. Using methods detailed by Silge (2017a, 2017b), the pointwise mutual information between all terms was computed within a predefined *skip window* (10 words on either side of the target words). Using singular value decomposition, the number of dimensions was reduced to 128. This process yielded a set of word vectors, each with 128 dimensions. These relationships can be plotted with the Uniform Manifold Approximation and Projection (UMAP) algorithm (McInnes et al., 2018). UMAP is a dimensionality reduction process that takes the *k* number of dimensions. It reduces these dimensions down to two (or whatever number the researcher specifies.) This method allows visualization in two-dimensional space. UMAP was used because it preserves the global structure between topics better than other popular methods such as PCA or t-Distributed Stochastic Neighbor Embedding (tSNE; McInnes et al., 2018; Oskolkov, 2019).

The cosine similarity between key terms shows how closely words are related through this vectorization process to find conceptual similarities or gaps (Hvitfeldt & Silge, 2020).

All analysis was conducted in R 4.0.3 using open-source packages.

2 | RESULTS

The total number of abstracts returned by the search of “equit|dispari” was 155, with the first article published in 1976 (Cole & Pilisuk, 1976). This number was much lower than anticipated. Figure 1 shows the distribution of articles over time. There is a large jump in the occurrence of these abstracts over the past year. Two targeted journals, *GJCCP* and *Australian Community Psychologist*, did not return any abstracts with these terms. To reframe this finding, there have been 2580 articles in total published in these three journals since 2007, when the topic began to show up more regularly. With 140 equity articles published since 2007, roughly 5% of all abstracts mentioned equity or disparities.

2.1 | Bag of words

Figure 2 shows the frequency of the top 25 *unigrams* (or single words). *Health* and *community* far outnumber other terms. However, unigrams by themselves may not be informative. Therefore, Figure 3 shows the most frequently occurring *bigrams* (or two-word phrases). Because a word may appear at the beginning of one phrase and the end of another, the bigrams were plotted as a network. Each node represents a token, with the arrows representing the next token in the sequence. Each node's size represents its frequency, with a larger circle being a more frequently occurring word. The line thickness between the nodes represents a more frequently occurring relationship. There is a large word cluster around health (particularly mental health), along with some clusters around methods, particularly *CPBR* and *Searching the PsycInfo database* (or PubMed, which does not require an institutional subscription!).

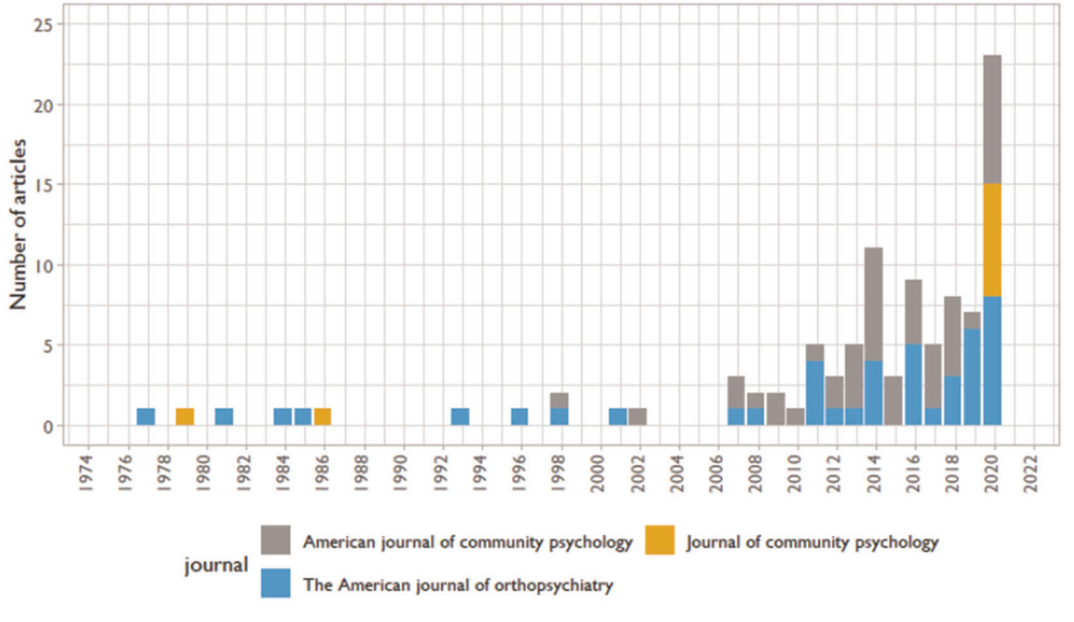

FIGURE 1 Distribution of equity-related abstracts over time

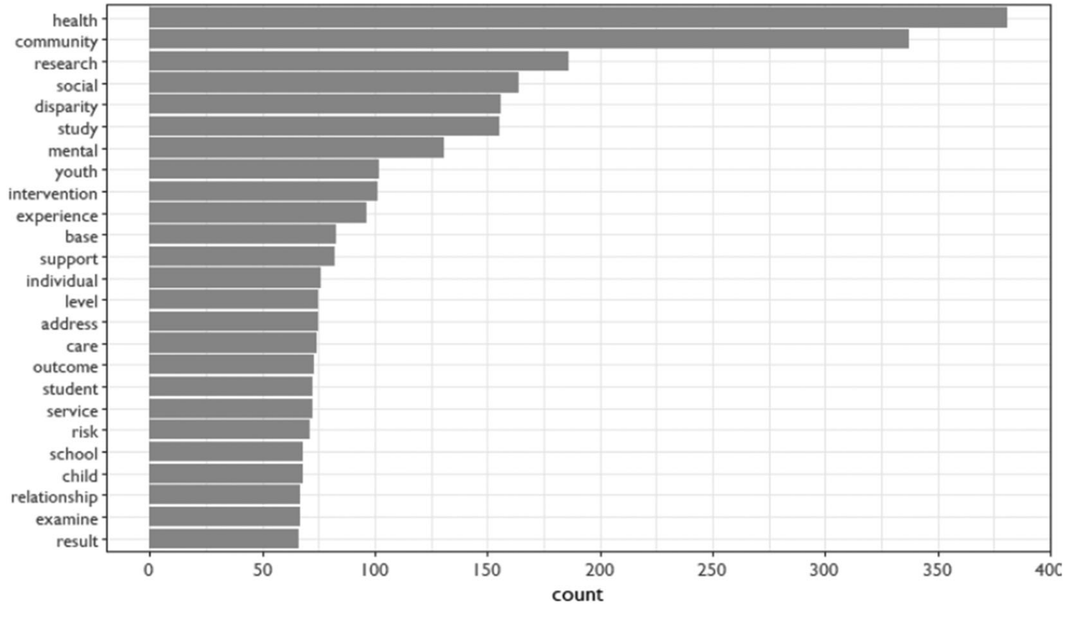

FIGURE 2 Most frequently occurring words in the database

2.2 | Topic modeling

To determine the minimum perplexity score, different configurations of topics from 2 to 30 were used. Although the absolute perplexity minimum was reached at $k = 30$, there began to be diminishing gains at $k = 20$ topics, a.k.a., the bend in the plot. Therefore because there were not many abstracts in this analysis, the more parsimonious $k = 20$ was chosen.

2.2.1 | Number of articles per unique topic

Figure 4 shows how many abstracts each topic had ranked from highest to lowest. The words on the *y*-axis were determined using *tf-idf*, which distinguished between topics by extracting what is unique to each topic compared to all other topics. When there was a tie in *tf-idf* score, more than three words are displayed (e.g., topic 8, in which *conversation*, *collective*, and *circle* had the same score.) This figure shows a high-level description of the different clusters, such as criminal justice (topic 10), school climate (topic 11), and CBPR (topic 16). The largest occurring topic overall concerned *community-based participatory research*.

Finally, Table 1 shows what article was most representative of each topic. The articles in this table had the highest γ-statistic for each topic, meaning that they had the highest probability of belonging to that topic. There is a relatively even spread across AJO and AJCP, with JCP only appearing once.

2.2.2 | Word embeddings

Word Embeddings were used to example relationships between keywords in the abstracts. Figure 5 applies the UMAP algorithm to the overall word vector space, reducing 128 dimensions to two for visualization purposes. The top 40 most frequently occurring words are plotted to show how different terms are related, with each dot

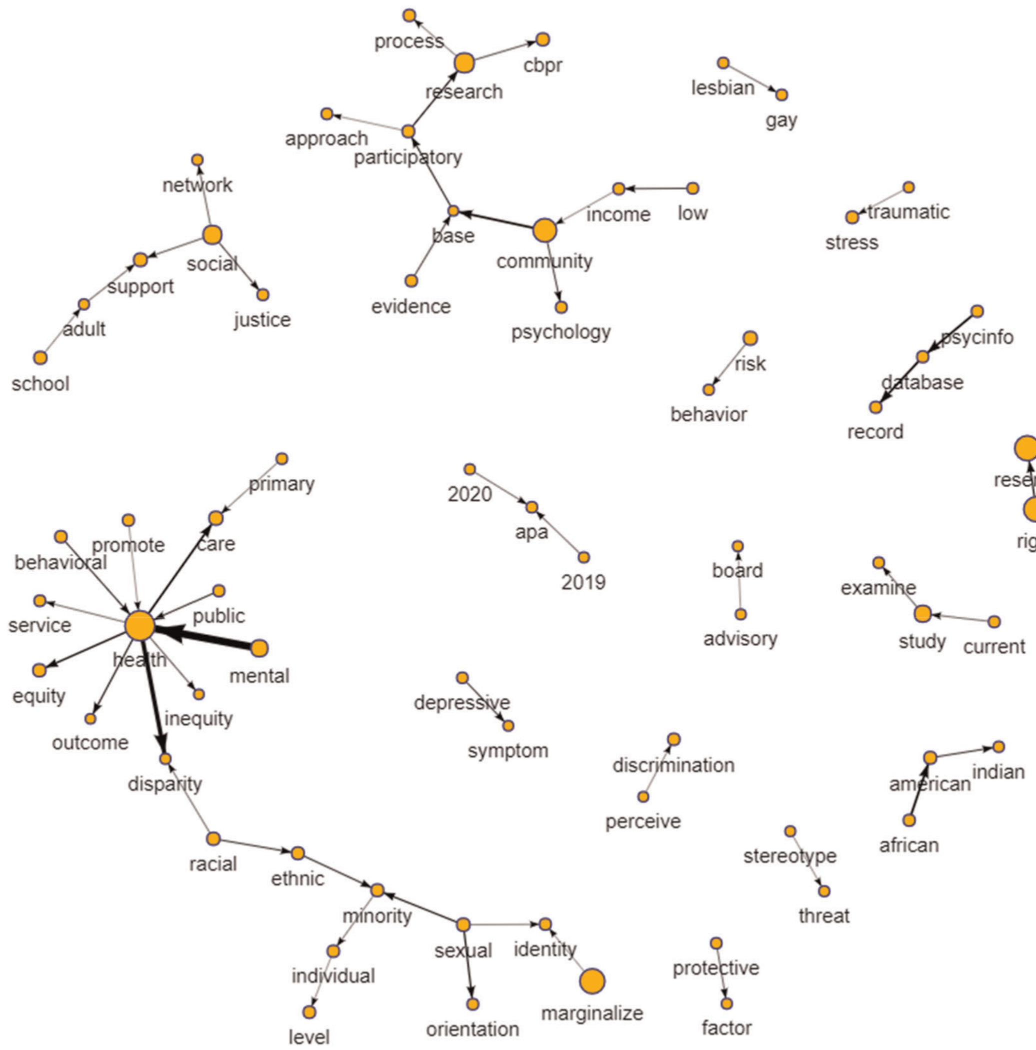

FIGURE 3 Network plot of word associations in abstracts

representing a specific word. Forty terms were chosen to balance between comprehensiveness and clarity for the visualization.

What this figure shows is that certain words are more closely associated with each other within these abstracts. For example, *racial* is closely spaced to *school*, *student*, *population*, and *American*. *Equity* is most closely related to *base* (a lemmatization of *evidence-based)*, *process*, and *research*. *Disparity*, however, appears in the lower right-hand corner in a cluster by itself. Strangely, *child* and *youth* did not appear closely related.

Apart from this visualization, cosine similarity was used to find the top 10 most closely related words in our key abstract searches. For equity, the most closely related words were, in descending order: *depression*, *power*, *sample*, *future*, *stand*, *inclusion*, *improve*, *finding*, and *incorporate*. For disparity, the most closely related words were, in descending order: *white*, *experience*, *affect*, *level*, *colonialism*, *underserved*, *inequity*, *expand*, and *city*.

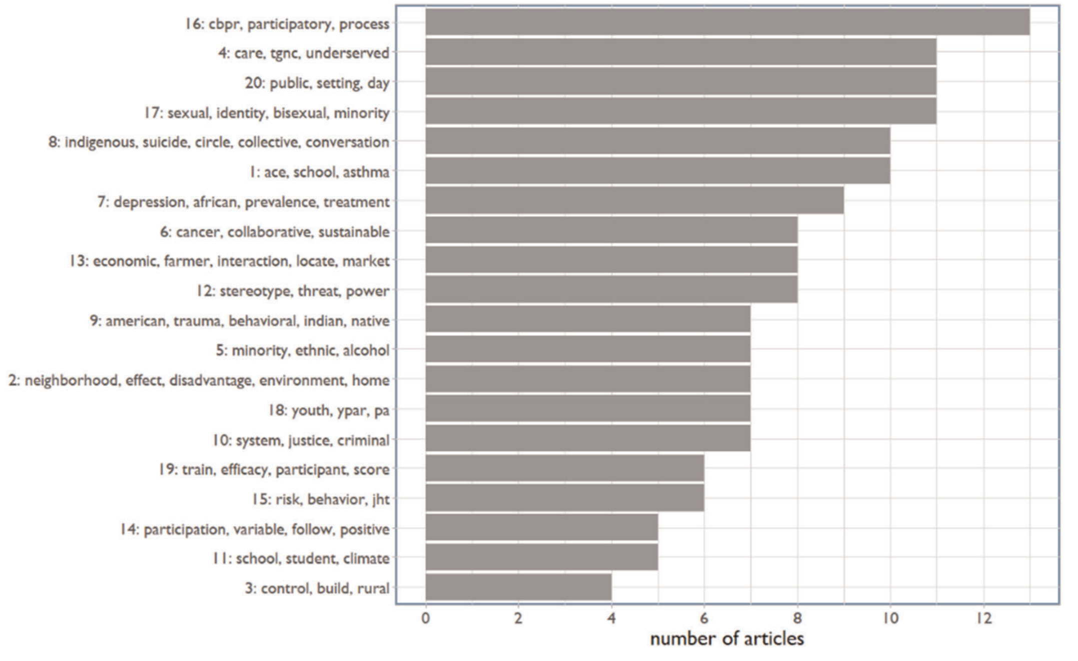

FIGURE 4 Number of abstracts per topic. The *y*-axis contains the topic number followed by tf-idf words

3 | DISCUSSION

3.1 | Usage of NLP in equity research

With some much data available online, the process of synthesis and translation can be laborious. Although community psychology has a long and rich qualitative research tradition, recent trends in computer-aided synthesis available through NLP have not been utilized. There is great potential to accelerate the synthesis process by leveraging these advances.

Of course, working with machine learning methods can open up all sorts of unintended (or intended bias) and must be implemented carefully and judiciously (Bergstrom & West, 2020; O'neil, 2016). NLP is an example of a narrow-AI paradigm, meaning that the algorithms do not understand the model outputs. The algorithms only cluster and weigh them. Therefore, there is still a need for human interpretation of results. This reflective process is also consistent with a participatory approach. Rather than just handing over results to communities, there can be a reflective process whereby all stakeholders can make meaning together (Scott et al., 2020). NLP is not human-independent and can fit in well with research on equity-related topics.

There can be a learning curve for all the methods discussed in this article, which is true of any new technique or process. However, there are highly active online communities (such as *StackOverflow* and *Kaggle*) and thousands and thousands of resources (such as *Medium*) that can support the implementation (and debugging) of these methods. In many ways, these materials are more accessible to the practitioner because they focus on implementing these methods rather than the underlying mathematics (which are important but intimidating).

Once learned, the type and number of documents available for efficient analysis increase exponentially: *school board minutes*, *newspaper articles*, *Twitter feeds*, *conference abstracts*, *blog posts*, anything that is text-based. This data-gathering process can itself be equitable because a much broader diversity of sources can be collected to help understand and tell the story of community-based settings. Researchers who combine methods like those

TABLE 1 Most representative article for each topic

Topic & tf-idf words	Most representative article	Authors	Date	Journal
1: ace, school, asthma	Mastery motivation and school readiness among young children experiencing homelessness.	Ramakrishnan, Masten	10/26/2020	AJO
3: control, build, rural	Multi-level intervention to prevent influenza infections in older low income and minority adults.	Schensul, Radda, Coman, Vazquez	8/17/2009	AJCP
4: care, tgnc, underserved	The provider perspective on behavioral health care for transgender and gender non-conforming individuals in the Central Great Plains: A qualitative study of approaches and needs.	Holt, Hope, Mocarski, Meyer, King, Woodruff	10/5/2020	AJO
5: minority, ethnic, alcohol	Epilogue for the special issue on sociocultural factors and mechanisms in alcohol use: Epidemiology, prevention, and intervention among ethnic minority groups: Lessons learned.	Sue	1/27/2020	AJO
7: depression, African, prevalence, treatment	Measurement and impacts of intersectionality on obsessive-compulsive disorder symptoms across intensive treatment.	Wadsworth, Potluri, Schreck, Hernandez-Vallant	7/3/2020	AJO
8: indigenous, suicide, circle, collective, conversation	Engaging critical community resilience praxis: A qualitative study with Mapuche communities in Chile facing structural racism and disasters.	Atallah, Contreras Painemal, Albornoz, Salgado, Pilquil Lizama	11/28/2019	JCP
9: american, trauma, behavioral, indian, native	The Urban American Indian Traditional Spirituality Program: Promoting Indigenous Spiritual Practices for Health Equity.	Gone, Tuomi, Fox	6/29/2020	AJCP
10: system, justice, criminal	Mattering at the Intersection of Psychology, Philosophy, and Politics.	Prilleltensky	4/22/2020	AJCP
11: school, student, climate	The Racial School Climate Gap: Within-School Disparities in Students' Experiences of Safety, Support, and Connectedness.	Voight, Hanson, O'Malley, Adekanye	9/1/2016	AJCP
13: economic, farmer, interaction, locate, market	Classroom peer relationships and behavioral engagement in elementary school: the role of social network equity.	Cappella, Kim, Neal, Jackson	6/20/2014	AJCP

TABLE 1 (Continued)

Topic & tf-idf words	Most representative article	Authors	Date	Journal
14: participation, variable, follow, positive	What's love got to do with it: Relationship functioning and mental and physical quality of life among pregnant adolescent couples.	Kershaw, Murphy, Divney, Magriples, Niccolai, Gordon	6/20/2014	AJCP
15: risk, behavior, jht	No youth left behind to human trafficking: Exploring profiles of risk.	Reid, Baglivio, Piquero, Greenwald, Epps	4/3/2020	AJO
16: cbpr, participatory, process	A Participatory, Mixed Methods Approach to Define and Measure Partnership Synergy in Long-standing Equity-focused CBPR Partnerships.	Coombe, Chandanabhumma, Bhardwaj, Brush, Greene-Moton, et al.	8/3/2020	AJCP
17: sexual, identity, bisexual, minority	Differences in sexual identity dimensions between bisexual and other sexual minority individuals: Implications for minority stress and mental health.	la Roi, Meyer, Frost	5/1/2019	AJO
18: youth, ypar, pa	Youth Participatory Action Research for Health Equity: Increasing Youth Empowerment and Decreasing Physical Activity Access Inequities in Under-resourced Programs and Schools.	Abraczinskas, Zarrett	6/26/2020	AJCP
19: train, efficacy, participant, score	Mentoring the next generation of behavioral health scientists to promote health equity.	Milburn, Hamilton, Lopez, Wyatt	9/2/2019	AJO
20: public, setting, day	Identifying and addressing mental health risks and problems in primary care pediatric settings: a model to promote developmental and cultural competence.	Godoy, Carter	6/26/2013	AJO

Note: AJCP, American Journal of Community Psychology; AJO, American Journal of Orthopsychiatry; JCP, The Journal of Community Psychology.

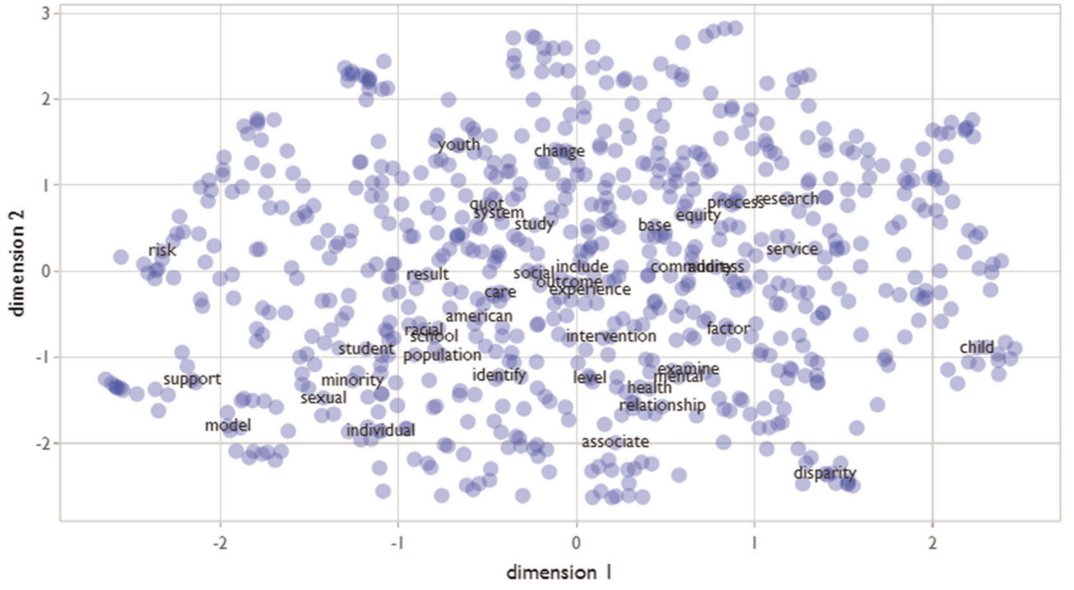

FIGURE 5 UMAP figure of word vectors. *Community* and *Address* are the words running on top of each other

discussed in this text with other, traditional methods may simultaneously reduce the research/evaluation burden of participants, effectively analyze more data, and devote the residual resources to more in-depth techniques like CBPR.

3.2 | Specific results

Equity is critical in achieving the fundamental goals of community psychology (Ozer et al., 2020). Anything less puts our work at risk of becoming exclusionary. The work and research of community psychologists are defined by specific values and competencies (Dalton & Wolfe, 2012). Therefore, there is an important need to take stock of how community psychologists have researched this area thus far so the field can further advance toward the goal of health and wellness for all. However, it is highly notable that more abstracts (and two entire journals) did not include an explicit equity component given the broad search criteria that were used.

Supplementary *google trends* search for *health equity* (searching for *equity* by itself brings in financial results) and *disparities* were conducted to see if the terminology used was a recent phenomenon. According to these results, these terms' use has slowly and steadily grown over time, not exponentially. At least on the surface, it does not appear that the spike in research in 2020 is due to changes in word use in the broader population. This finding reflects a *prima facie* gap; thus far, the published community psychology work has not comprehensively addressed equity.

Within the abstracts with an equity lens, the topic modeling showed that a core community psychology research strategy leads the way: Community-Based Participatory Research (Kral & Allen, 2016; Wallerstein et al., 2017). The field is well accustomed to participatory processes that can help engage end-users and stakeholders throughout a research process. CBPR satisfies the condition that equitable research focuses on both process and outcomes. This finding was also confirmed in the word embeddings, whereby equity was closely related to both the terms of *process* and *research*.

Consistent with community psychology being a setting-diverse field, there were topic clusters around schools, transgender and gender non-conforming groups, Native Americans, and youth communities. This finding shows that within these different settings, researchers are approaching their questions with an equity-informed lens.

These findings, taken together, suggest that there is breadth to community psychology equity research. However, with the emerging primacy of equity-focused interventions (Brownson et al., 2021), should community psychologists be striving to address equity centrally in their work more explicitly? Such a focus would certainly be consistent with our social justice values (Wolfe, 2014).

3.3 | Additional limitations

The NLP methods used here provide a high-level overview. By relying on the abstracts, critical information that is likely contained in the full text was not considered or "read" by the algorithms. Therefore, consider these results exploratory. A further and more in-depth reading of the text would provide a richer picture of how and where community psychology has implemented equity-informed.

This was a very simple search, focusing on a narrow range of search terms in a limited selection of journals. Using adjacent concepts like *equality* and *justice* would have pulled in additional articles, though it may have opened up the number of topics considerably. Also, many other journals may be a more suitable home for work on equity, such as *Health Promotion and Practice*, *Health Education and Behavior*, and, of course, *Health Equity*. Researchers working in these areas may simply go to these outlets first.

A more critical gap concerns what articles are published in the first place. Academic publications have a specific set of purposes that may not align with the goals of public health departments and nongovernmental organizations devoted to pursuing equity. As such, an unknowable amount of innovation and progress may be occurring that is not captured by the narrow set of publications listed in this article. In a way, this further emphasizes the need to continue to use CBPR methods that can help engage community members in the research dialogue. NLP methods can do some things very well, but they can not do all things.

4 | CONCLUSION

NLP methods can be a powerful tool in the community-based researcher's toolbox. This article has shown how a broad set of methods can be applied to academic literature to understand overarching themes and constructs. Specifically, this article has shown that community psychology's equity-related work has been more recent yet diverse in application. The emphasis on CBPR appears to be a critical ingredient in this study and should continue to be relied on to foster progress.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/jcop.22603

ORCID

Jonathan P. Scaccia https://orcid.org/0000-0001-6800-1286

REFERENCES

Azunre, P. (2020). *Transfer learning for natural language processing*. Manning Press.

Baker, A. (2001). Crossing the quality chasm: A new health system for the 21st century. *British Medical Journal*, *323*(7322), 1192.

Balas, E. A., & Boren, S. A. (2000). Managing clinical knowledge for health care improvement. *Yearbook of Medical Informatics*, (1), 65–70.

Bergstrom, C. T., & West, J. D. (2020). *Calling bullshit: The art of skepticism in a data-driven*. Penguin UK.

Berkowitz, S. A., Cené, C. W., & Chatterjee, A. (2020). Covid-19 and health equity—Time to think big. *New England Journal of Medicine*, *383*(12):e76.

Brownson, R. C., Kumanyika, S. K., Kreuter, M. W., & Haire-Joshu, D. (2021). Implementation science should give higher priority to health equity. *Implementation Science*, *16*(1), 1–16.

Chambers, D. A. (2020). Covid19: Considering the intersection between implementation science and COVID-19. *Implementation Research and Practice*, *1*. https://doi.org/10.1177/0020764020925994

Colditz, G. A., & Emmons, K. M. (2012). The promise and challenges of dissemination and implementation research. *Dissemination and Implementation Research In Health: Translating Science to Practice*, 3–22.

Cole, J., & Pilisuk, M. (1976). Differences in the provision of mental health services by race. *American Journal of Orthopsychiatry*, *46*(3), 510–525.

Dalton, J., & Wolfe, S. M. (2012). Joint Column. Education connection and the community practitioner: Competencies for community psychology practice. *The Community Psychologist*, *45*(4), 7–14.

Estabrooks, P. A., Brownson, R. C., & Pronk, N. P. (2018). Dissemination and implementation science for public health professionals: An overview and call to action. *Preventing Chronic Disease*, *15*, 162.

English, M., Moshabela, M., Nzinga, J., Barasa, E., Tsofa, B., Marchal, B., & Kruk, M. E. (2020). Systems and implementation science should be part of the COVID-19 response in low resource settings. *BMC Medicine*, *18*(1), 1–2.

Erkan, G., & Radev, D. R. (2004). Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research*, *22*, 457–479.

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, *30*, 1–14.

Hvitfeldt, E., & Silge, J. (2020) Supervised machine learning for text analysis in R. https://smltar.com/

Jason, L., & Glenwick, D. (Eds.). (2016). *Handbook of methodological approaches to community-based research: Qualitative, quantitative, and mixed methods*. Oxford University Press.

Kapadia, S. (2019). Evaluate topic models: Latent Dirichlet Allocation (LDA). *Towards Data Science*.

Kral, M. J., & Allen, J. (2016). Community-based participatory action research. *Handbook of methodological approaches to community-based research: Qualitative, quantitative, and mixed methods*, 253–262.

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *ArXiv preprint arXiv*, *1802*, 03426.

Moody, C. (2015). A word is worth a thousand vectors. *Mulithreaded*. https://multithreaded.stitchfix.com/blog/2015/03/11/word-is-worth-a-thousand-vectors/

Morris, Z. S., Wooding, S., & Grant, J. (2011). The answer is 17 years, what is the question: Understanding time lags in translational research. *Journal of the Royal Society of Medicine*, *104*(12), 510–520.

O'neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books.

Oskolkov, N. (2019) How exactly UMAP works—And why exactly it is better than tSNE. *Towards Data Science*. https://towardsdatascience.com/how-exactly-umap-works-13e3040e1668

Ozer, E. J., Abraczinskas, M., Duarte, C., Mathur, R., Ballard, P. J., Gibbs, L., Olivas, E. T., Bewa, M. J., & Afifi, R. (2020). Youth participatory approaches and health equity: Conceptualization and integrative review. *American Journal of Community Psychology*, *66*, 267–278.

Plough, A. L. (Ed.). (2018). *Advancing health and well-being: Using evidence and collaboration to achieve health equity*. Oxford University Press.

Radley, D. C., Collins, S. R., & Baumgartner, J. (2020). 2020 scorecard on state health system performance. *The Commonwealth Fund*.

Raja, U., Mitchell, T., Day, T., & Hardin, J. M. (2008). Text mining in healthcare. Applications and opportunities. *Journal of Healthcare Information Management*, *22*(3), 52–56.

Reid, A., Abraczinskas, M., Scott, V., Stanzler, M., Parry, G., Scaccia, J., Wandersman, A., & Ramaswamy, R. (2019). Using collaborative coalition processes to advance community health, well-being, and equity: A multiple-case study analysis from a national community transformation initiative. *Health Education & Behavior*, *46*(1_suppl), 100S–109S.

Scott, V. C., Alia, K., Scaccia, J., Ramaswamy, R., Saha, S., Leviton, L., & Wandersman, A. (2020). Formative evaluation and complex health improvement initiatives: A learning system to improve theory, implementation, support, and evaluation. *American Journal of Evaluation*, *41*(1), 89–106.

Silge, J. (2017a). Word vectors with tidy data principles. https://juliasilge.com/blog/tidy-word-vectors/

Silge, J. (2017b). Tidy word vectors, take 2!. https://juliasilge.com/blog/word-vectors-take-two/

Wallerstein, N., Duran, B., Oetzel, J. G., & Minkler, M. (Eds.). (2017). *Community-based participatory research for health: Advancing social and health equity*. John Wiley & Sons.

Wolfe, S. M. (2014). The application of community psychology practice competencies to reduce health disparities. *American Journal of Community Psychology*, *53*(1–2), 231–234.

How to cite this article: Scaccia, J. P. (2021). Examining the concept of equity in community psychology with natural language processing. *J Community Psychol*, 49, 1718–1731. https://doi.org/10.1002/jcop.22603