Received: 28 January 2019 | Accepted: 8 July 2019

DOI: 10.1111/2041-210X.13268

RESEARCH ARTICLE

Methods in Ecology and Evolution BRITISH ECOLOGICAL SOCIETY

An automated approach to identifying search terms for systematic reviews using keyword co-occurrence networks

Eliza M. Grames[1] | Andrew N. Stillman[1] | Morgan W. Tingley[1] | Chris S. Elphick[1,2]

[1]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut

[2]Center of Biological Risk, University of Connecticut, Storrs, Connecticut

Correspondence
Eliza Grames
Email: eliza.grames@uconn.edu

Funding information
University of Connecticut

Handling Editor: Robert Freckleton

Abstract

1. Systematic review, meta-analysis and other forms of evidence synthesis are critical to strengthen the evidence base concerning conservation issues and to answer ecological and evolutionary questions. Synthesis lags behind the pace of scientific publishing, however, due to time and resource costs which partial automation of evidence synthesis tasks could reduce. Additionally, current methods of retrieving evidence for synthesis are susceptible to bias towards studies with which researchers are familiar. In fields that lack standardized terminology encoded in an ontology, including ecology and evolution, research teams can unintentionally exclude articles from the review by omitting synonymous phrases in their search terms.
2. To combat these problems, we developed a quick, objective, reproducible method for generating search strategies that uses text mining and keyword co-occurrence networks to identify the most important terms for a review. The method reduces bias in search strategy development because it does not rely on a predetermined set of articles and can improve search recall by identifying synonymous terms that research teams might otherwise omit.
3. When tested against the search strategies used in published environmental systematic reviews, our method performs as well as the published searches and retrieves gold-standard hits that replicated versions of the original searches do not. Because the method is quasi-automated, the amount of time required to develop a search strategy, conduct searches, and assemble results is reduced from approximately 17–34 hr to under 2 hr.
4. To facilitate use of the method for environmental evidence synthesis, we implemented the method in the `R` package `litsearchr`, which also contains a suite of functions to improve efficiency of systematic reviews by automatically deduplicating and assembling results from separate databases.

KEYWORDS

evidence synthesis, keyword co-occurrence network, keyword identification, literature review, meta-analysis, systematic review, text mining

1 | INTRODUCTION

With an ever-growing body of scientific literature in ecology, evolution, conservation biology and related fields, there is an increasing need to summarize trends, identify emerging questions, clarify controversies, and explain conflicting results (Haddaway, Macura, Whaley, & Pullin, 2018; Sutherland & Wordley, 2018). Two central techniques to synthesize evidence are (a) systematic reviews, which search available literature for evidence with which to address a research question, and (b) meta-analyses, which quantitatively assess statistical evidence found through systematic reviews.

Modern approaches to evidence synthesis, whereby a formal strategy is used to study the way in which a topic has been studied, originated in clinical fields (Glass, 1976; Hunt, 1997; Smith & Glass, 1977), but have been co-opted by ecologists in recognition of the need for more rigorous methods of reviewing the literature (Pullin & Knight, 2001; Pullin & Stewart, 2006). A good meta-analysis remains dependent on a good sampling of the underlying universe of studies, which requires a careful and comprehensive systematic review (Côté, 2013). Formal approaches to systematic review in the field have often focused on applied questions, for example by leading to the development of the Collaboration for Environmental Evidence (Pullin & Knight, 2009), but are broadly applicable across ecology and evolutionary biology. Meta-analysis was introduced to ecology earlier, for example, in Järvinen's 1991 study of egg laying dates and clutch size in cavity-nesting birds. The approach has since become a standard tool for combining information from multiple studies with hundreds of studies across the discipline (Gurevitch, Curtis, & Jones, 2001; Koricheva, Gurevitch, & Mengersen, 2013).

Despite the important role research synthesis techniques play in building a stronger evidence base, their implementation is hampered in ecology, conservation biology and related fields by financial (as discussed in Sutherland & Wordley, 2018) and time (Haddaway & Westgate, 2019) costs. To reduce the time and resources needed to synthesize evidence, researchers have called for automation of the most time-intensive tasks while still maintaining the standards of conventional search methods (Delaney & Tams, 2018; O'Mara-Eves, Thomas, McNaught, Miwa, & Ananiadou, 2015; Paisley & Foster, 2018; Tsafnat et al., 2014; Tsertsvadze, Chen, Moher, Sutcliffe, & McCarthy, 2015). Technological advancements have enabled automation of key tasks such as removing duplicate articles, prioritizing articles for screening, and extracting data from tables and figures (Jonnalagadda, Goyal, & Huffman, 2015; Przybyla et al., 2018; Rathbone, Hoffmann, & Glasziou, 2015; Shemilt et al., 2014). These automation methods target the results of a systematic literature search after it has been conducted, ignoring the root problem—finding all relevant literature without retrieving excessive irrelevant evidence during the search process.

Ideally, search strategies for systematic reviews should be able to return all of the studies relevant to the review ('recall') without retrieving irrelevant studies ('precision'). Fields such as medicine—where systematic review originated (Cochrane, 1972)—and public health have institutional support and a standardized ontology (i.e. Medical Subject Headers, or MeSH) that facilitate search strategy development (Bramer, Rethlefsen, Mast, & Kleijnen, 2018). In ecology and evolutionary biology where systematic reviews have a less established history, there are no standardized ontologies like MeSH, leading researchers to use broad, non-specific keywords in their searches (Pullin & Stewart, 2006). The low precision of this approach means that only a small percent (0.473%; Haddaway and Westgate (2019), Supporting Information) of all search results in environmental systematic reviews are relevant, greatly increasing the amount of time required to screen articles. An alternative requires that researchers spend more time on search strategy development and iteratively test their search strategies against a set of known articles to select more specific, yet still comprehensive, keywords to maintain high precision without sacrificing the total number of suitable articles retrieved (O'Mara-Eves et al., 2015). Because researchers using this approach typically select keywords based on their own knowledge of the field and fail to specify how the search strategy was developed, it is also susceptible to selection bias (Haddaway, Woodcock, Macura, & Collins, 2015), irreproducibility, and low recall if researchers do not select a comprehensive set of terms. Increased standardization in search strategy development is necessary to improve the specificity, objectivity, and reproducibility of systematic reviews (Hausner, Waffenschmidt, Kaiser, & Simon, 2012; Stansfield, O'Mara-Eves, & Thomas, 2017).

The two primary approaches to automating search strategy development are citation networks and text mining, both using a set of predetermined articles that researchers deem relevant to the review. Given a set of known articles, citation relationships between articles can identify related articles without using keywords. This approach has high precision, but low recall, and carries the risk of introducing selection bias, citation bias, publication bias, and other forms of bias because the starting set of articles influences what is eventually retrieved (Belter, 2016; Sarol, Liu, & Schneider, 2018). Text mining approaches typically identify potential keywords from a set of known articles based on their frequency in relation to baseline word frequencies, leading to difficulties with phrases composed of multiple generic words (Hausner et al., 2012; Zhang, Babar, Bai, Li, & Huang, 2011). For example, terms like 'species distribution model' are ignored because all three words on their own may be common, but in the proper order they convey a specific meaning. Although text mining approaches tend to have high recall, current methods require custom coding to select keywords for each review, require as much time to develop a search strategy as conventional methods (Hausner, Guddat, Hermanns, Lampert, & Waffenschmidt, 2015, 2016; Hausner et al., 2012; O'Mara-Eves et al., 2015) and are subject to selection bias similar to conventional methods. Current approaches to automating keyword selection require researchers to select a starting set of articles with which they are already familiar, predisposing citation networks and terms found with text mining towards familiar articles.

To combat these problems and make systematic reviews more accessible and comprehensive in ecology and evolutionary biology, we developed a quasi-automated method that is quick, reproducible objective and can be easily implemented in fields that lack standardized ontologies. We used text mining and keyword co-occurrence networks to efficiently identify potential keywords without relying on a potentially biased set of preselected articles. To facilitate reproducibility and transparency, we created the R package `litsearchr` (Grames, Stillman, Tingley, & Elphick, 2019a) to aid implementation of the method in a user-friendly format. To exemplify this method, we generated a search strategy for a review of ecological processes leading to declines in occupancy of black-backed woodpeckers *Picoides arcticus*—a post-fire specialist—with time since wildfire.

2 | MATERIALS AND METHODS

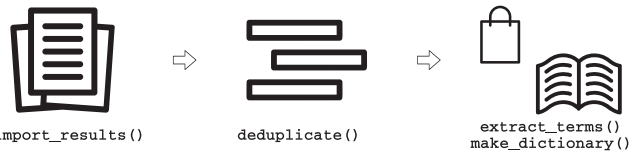
2.1 | Writing the naïve search and importing results

Our method (Figure 1) begins with a precise Boolean search that returns a set of highly relevant articles. This 'naïve' search should include only the most important search terms grouped into concept categories, or sets of synonymous terms, which are then combined into a Boolean search (see Supporting Information S1 for details on writing good naïve searches). The PICO (Population, Intervention, Control, Outcome; Richardson, Wilson, Nishikawa, & Hayward, 1995), PECO (Population, Exposure, Comparator, Outcome; Haddaway, Bernes, Jonsson, & Hedlund, 2016) or other variants (e.g. PICOTS; Samson & Schoelles, 2012) can be used if appropriate. If the naïve search is too imprecise, it will return many irrelevant articles and dilute the subsequent keyword selection process with the consequence that vague terms (e.g. 'positive effects') could be identified as important because they are the only terms shared by irrelevant articles.

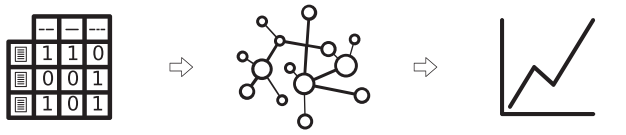
2.2 | Assembling and deduplicating results

Many articles are indexed in multiple databases, which can lead to over-representation of terms and themes in the results of a search because some articles appear more than once. To prevent over-representation when identifying potential terms, naïve search results need to be assembled and deduplicated in the second stage of the `litsearchr` workflow (Figure 1). Done manually, this is a time-intensive process because platforms and databases export results in different formats (Haddaway & Westgate, 2019). Given the path to a directory of search results, the `litsearchr` function `import_results` automatically identifies the file type and database from which each file originated, selects analogous columns (e.g. the Abstract field in Scopus results and the AB field in BIOSIS Citation Index), and binds them into a single dataset. The `deduplicate` function removes stopwords like 'and', 'while', 'through', etc. from the titles and abstracts and calculates similarity scores for all of the resulting tokenized abstracts and titles. The default settings in `litsearchr` remove exact

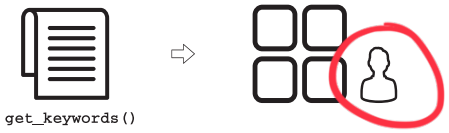
1. Research team writes a naïve search to capture a set of highly relevant articles. The naïve search should be conducted in at least two supported databases and results should be exported per `litsearchr` guidelines.



2. `litsearchr` imports search results in a standardized format and removes duplicates. Each article is turned into a "bag of words" before keywords are extracted from the full set of articles to create a dictionary of possible search terms.



3. `litsearchr` creates a document-feature matrix using each possible term from the dictionary as features; this matrix becomes a keyword co-occurrence network. `litsearchr` fits a model to the network to find a cutoff in keyword importance.



4. `litsearchr` suggests possible search terms central to the topic of the review. The research team manually considers these keywords and sorts them into concept groups. This process can be repeated with terms that share a stem.

5. `litsearchr` removes redundant terms and then translates terms into user-specified languages. It then writes Boolean search strings for each language which the review team uses to conduct searches.

FIGURE 1 Graphical representation of the litsearchr workflow. Icons with functions listed below them can be done automatically by litsearchr whereas other steps require manual input. An information specialist or librarian should be part of the review team, especially for steps indicated with a person icon. Icons created by Calvin Goodman, Meaghan Hendricks, Yu Luck, and Mun May Tee from the Noun Project

title duplicates, titles that are more than 95% similar, or abstracts that are more than 85% similar; these default title and abstract similarity levels can be changed by the user to alter the stringency of deduplication. In a sample of 1,083 article records from our woodpecker example, the default settings in the `deduplicate` function correctly classified 100% of the 308 duplicate articles identified by manual deduplication.

2.3 | Extracting and identifying keywords

To extract potential keywords from the titles and abstracts of articles in the deduplicated dataset, `litsearchr` uses the Rapid Automatic Keyword Extraction (RAKE) algorithm, which is designed to identify keywords in scientific literature by selecting strings of words uninterrupted by stopwords or punctuation (Rose, Engel, Cramer, & Cowley, 2010). By default, the function `extract_terms` calls the RAKE algorithm from the `rapidraker` package (Baker, 2018), eliminates keywords that only appear in a single article, and excludes phrases with only one word from the list of potential keywords because *n*-grams greater than one are more specific and will result in a more precise search (Stansfield et al., 2017). Although these are the default options, `litsearchr` can also extract unigrams, use different keyword extraction algorithms, or report terms that occur in any number of articles above a user-specified threshold. Our method then combines these terms with the author- and database-tagged keywords as a dictionary object created with `create_dictionary` to define the universe of possible keywords. These possible keywords are passed to the function `create_dfm`, which wraps functions from the `quanteda` package (Benoit, 2018) to generate a document-feature matrix using the potential keywords as features and the combined titles, abstracts, and keywords of each article as the documents.

Rather than relying solely on frequency of keywords as an indicator of their relevance to a search strategy, the third stage of the `litsearchr` workflow (Figure 1) generates a keyword co-occurrence network (Su & Lee, 2010) with the function `create_network` to measure each term's importance and influence in relation to the topic being reviewed. In the keyword co-occurrence network, each node represents a potential search term and the edges are co-occurrences of two terms in the title, abstract, or tagged keywords of a study. Although multiple measures of node importance could be used, `litsearchr` defaults to using node strength, a weighted measure of the degrees of a network that indicates how well-connected a node is to other strong nodes (Radhakrishnan, Erbis, Isaacs, & Kamarthi, 2017). Because node strength of a network tends to follow a power law (Radhakrishnan et al., 2017), there are regions of rapid change where nodes become increasingly more important. To find these regions and identify tipping points beyond which keyword node strength increases dramatically, `find_cutoff` ranks nodes by strength and uses a genetic algorithm with migration (Spiriti, Eubank, Smith, & Young, 2013) to identify knots, or rapid change points, in keyword importance.

To determine a cutoff beyond which keywords will be manually considered for inclusion in the search terms, `litsearchr` takes the strength of the node at the first knot and retrieves the keywords associated with all nodes stronger than it. The `litsearchr` package suggests these terms to the review team, who must then manually review the terms to determine which terms are appropriate for the final search string and assign concept group(s) to each included term. After the review team makes decisions on the suggested terms, they can be passed back to the `litsearchr` function `get_similar`, which extracts keywords that share a unigram root word stem with the included keywords and suggests these for inclusion as well (e.g. 'nest site selection' would return terms like 'nesting success' because of the shared stem 'nest' or 'habitat selection' because of the stem 'select').

2.4 | Writing Boolean searches

Once the full suggested list of keywords is manually reviewed for inclusion in the search string and grouped into concept categories, `litsearchr` can write simple Boolean searches with the groups. At this stage of the workflow (Figure 1), research teams can add search terms not retrieved by `litsearchr` or unigrams that they deem important. `litsearchr` does not restrict keywords, but rather suggests new keywords that authors may not have considered.

When writing a search string for a systematic review, researchers will often want to use stemming to capture additional word forms because it makes the search string more efficient (Bramer, Jonge, Rethlefsen, Mast, & Kleijnen, 2018). For example, a search string to capture articles about fledglings should use the stemmed term `fledg*` because the asterisk will be replaced with alternative word endings (e.g. fledgling, fledglings, fledge, fledged, etc.). Consequently, all of the word forms do not need to be included in the search string. To stem terms and capture additional word forms when searching databases, `litsearchr` uses a stemming algorithm (Porter, 1980) to reduce each word to its root form if its stem is at least four characters long, which is the length suggested by information specialists to balance efficiency with recall (J. Livingston, pers. commun.).

The function `write_search` then removes redundant terms to make the search string more efficient and limit its length. It detects multi-word phrases that will be retrieved with shorter phrases that are also included in the search string (e.g. 'habitat suitability index' will be retrieved by 'habitat suitability' so it can be removed) or for which stemmed forms are identical (e.g. 'population density' and 'population densities' both reduce to 'popul* dens*' so only one is needed). Although the default option in `litsearchr` places search phrases in quotation marks to return exact phrases, this feature is optional. Within concept categories, `write_search` separates terms by the Boolean operator OR, then connects concept categories with the AND operator.

To facilitate access to non-English language sources, `litsearchr` can write searches in up to 53 languages (listed with `available_languages`), although stemming is not currently supported in non-English languages. The translations are done by accessing the Google Translate API, which requires independent registration for an API key. Given the scientific field of the review, `choose_languages` can prioritize non-English languages to search by using journal subject classifications. Searches written by `litsearchr` work in over twenty databases commonly used in ecology and evolution, which can be viewed with `usable_databases()`. Many other databases are likely also compatible with the search strings but have not been tested. Additionally, three open access thesis databases can be searched automatically with `scrape_results`.

TABLE 1 An example of naïve search terms in concept categories for a review answering the question: 'What ecological processes lead to declines in occupancy of black-backed woodpeckers (*Picoides arcticus*) with time since fire?' The concept categories using the PECO framework (Haddaway et al., 2016) are woodpeckers in post-fire forests (population), ecological processes (exposure) and changes in occupancy (outcome) with no comparator group. Terms are truncated with an asterisk to allow for stemming that captures additional word forms

Concept category	Terms
Population	(woodpecker* OR sapsucker* OR Veniliorn* OR Picoid* OR Dendropic* OR Melanerp* OR Sphyrapic*) AND (fire* OR burn* OR wildfire*)
Exposure	((nest* OR reproduct* OR breed* OR fledg*) W/3 (succe* OR fail* OR surviv*)) OR (surviv* OR mortalit* OR death*) OR ('food availab*' OR forag* OR provision*) OR (emigrat* OR immigrat* OR dispers*)
Comparator	[not applicable to research question]
Outcome	(occup* OR occur* OR presen* OR coloniz* OR colonis* OR abundan* OR 'population size' OR 'habitat suitability' OR 'habitat selection' OR persist*)

2.5 | Checking search strategy performance

Before conducting final searches for a systematic review, research teams should test their search strategy to confirm that articles known to be relevant to the review are found by the search terms. Given the search results and a character vector containing the titles of articles known to be relevant, `check_recall` determines if the search retrieved the known relevant articles. The function `search_performance` can then calculate performance metrics such as recall (percent of known relevant articles returned), precision (number of known articles returned out of the total number of search results), and number needed to process (NNP: number of articles that will need to be manually screened to find one relevant article).

2.6 | Worked example

To demonstrate the method, we developed a search strategy for a review answering the question: 'What ecological processes lead to declines in occupancy of black-backed woodpeckers *Picoides arcticus* with time since fire?' We identified our concept categories using the PECO framework (Haddaway et al., 2016) as woodpeckers in post-fire forests (population), ecological processes (exposure), and changes in occupancy (outcome) with no comparator (Table 1). To improve precision, we restricted our naïve search to titles, abstracts, and keywords and only conducted the naïve search in databases commonly used in ecology and evolutionary biology—BIOSIS Citation Index, Zoological Record, and Scopus. We conducted the searches in October 2018 with no further restrictions.

Our naïve search retrieved 1,083 articles, 308 of which were duplicates (Figure 2). From our deduplicated dataset, the RAKE algorithm identified 3,479 potential *n* –gram keywords and we extracted

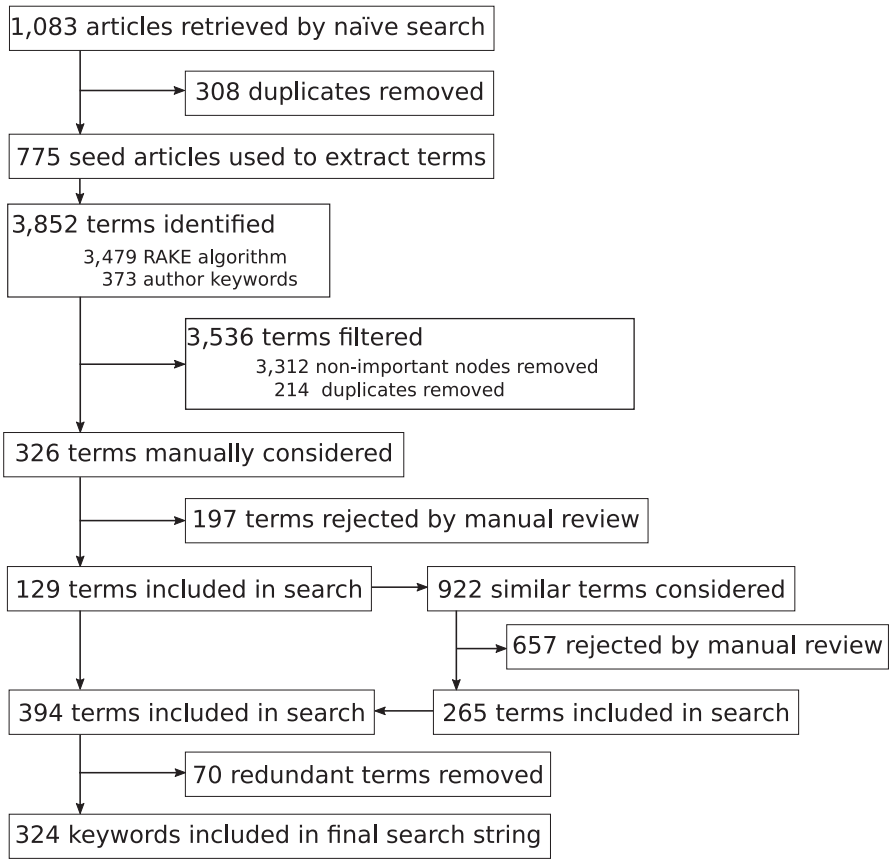


FIGURE 2 Flowchart of seed articles and term selection using `litsearchr` for an example review answering the question: 'What ecological processes lead to declines in occupancy of black-backed woodpeckers (*Picoides arcticus*) with time since fire?' See Supporting Information S2 for the search terms considered, rejected, and included at each step

373 author– and database–tagged keywords, of which 159 were not found by the RAKE algorithm. After fitting the spline model with three knots selected by the `freepsgen` algorithm (Spiriti, Smith, & Lecuyer, 2018), we manually considered 326 keywords in the first stage. We selected 129 terms and then considered an additional 922 terms that shared a unigram stem with the terms we selected, from which we retained 265 terms. After removing redundant terms, the final search string contained 324 unique keywords (Supporting Information S2).

2.7 | Method performance

To test our method, we compared its performance to the standard approach of manually creating a search string either de novo based on the review team's knowledge of the field or with iterative testing of search string combinations in databases used for the review. We compared the performance of search strings developed with `litsearchr` to the search strings reported in systematic reviews published in the journal *Environmental Evidence* (Figure 3). We selected a convenience sample of six systematic reviews that had clearly reported search strategies and for which we felt sufficiently knowledgeable about the topic to make decisions at the manual stages of the `litsearchr` workflow. For each published review, we wrote a naïve search string based on the title, abstract, and introduction and conducted the naïve search in three databases (Scopus, Zoological Record, and BIOSIS Citation Index). We put our naïve search results into `litsearchr` and used the default settings to generate a new search string for each review. Using the new search string developed with `litsearchr`, we conducted the search in eight to ten databases. We searched in the title, abstract, and keywords or equivalent search field and placed no filters or restrictions on the search other than the years searched. We restricted the end date for each search to match the date reported in the published review or our best approximation if the date was not reported. In order to control for access limitations or database changes since the review was published, we conducted replicated searches with the search string reported in the original review using the exact same set of databases and end dates as for the `litsearchr` searches. To test the recall and precision of the `litsearchr` and replicated searches, we checked their results against the list of included studies from the published review to determine which studies were retrieved. The list of included studies from the published review was treated as the gold standard; recall was measured as the percent of gold standard hits

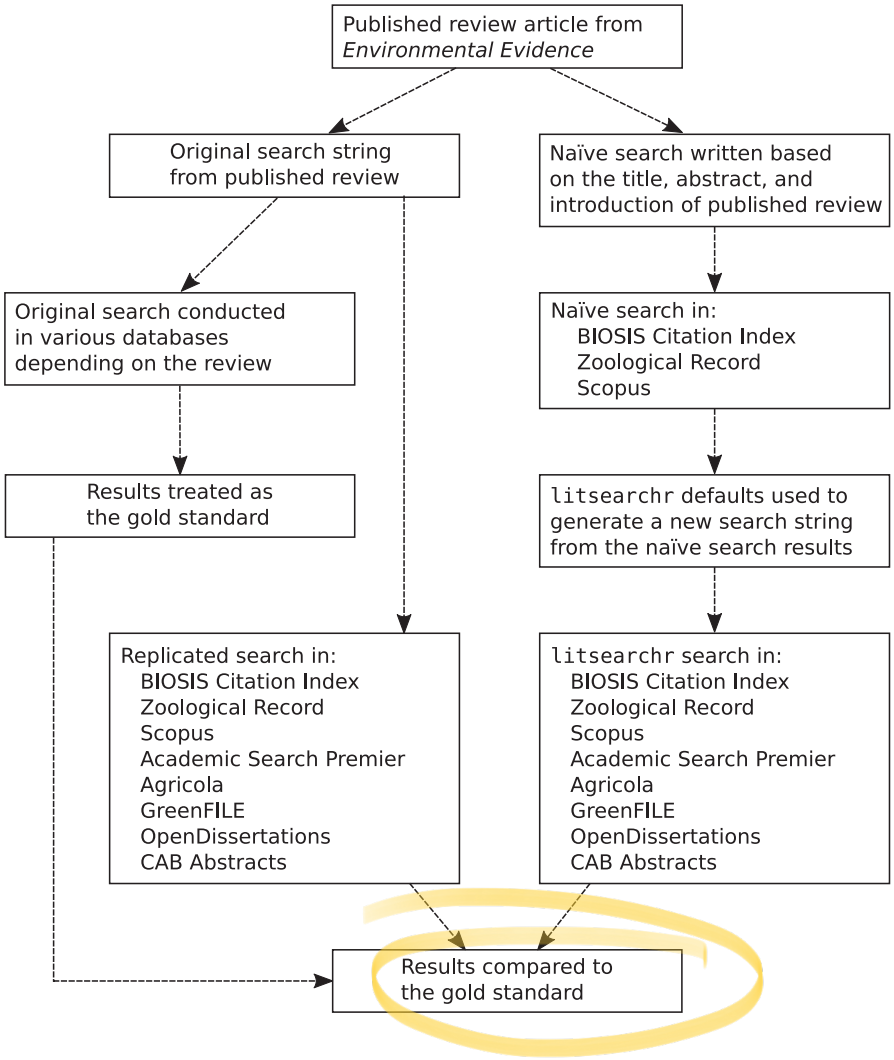


FIGURE 3 Representation of methods for testing `litsearchr` against search strategies from published reviews. For each published review included in the sample, a naïve search, `litsearchr` search, and replicated search were conducted and compared to the articles included in the original review

retrieved and precision was measured as the percent of all hits retrieved that were gold standard hits. We used two-tailed pairwise *t*-tests to compare our method to the replicated searches for both precision and recall. Because we wanted to know the minimum detectable difference in performance between our method and the replicated searches, we did power analyses with alpha of 0.050 and the standard deviation of the recall or precision to estimate the largest possible effect that could have been detected with power of 0.800. To compare the performance of our method to other quasi-automated techniques, we performed the same pairwise *t*-tests and power analyses on the raw data from three related studies that report precision and recall test results for citation network (Belter, 2016), text-mining (Hausner, Guddat, Hermanns, Lampert, & Waffenschmidt, 2015), and combination approaches (Sarol, Liu, & Schneider, 2018).

3 | RESULTS

We replicated searches from six systematic reviews published in *Environmental Evidence* (Figure 4). Although we do not know how long it took researchers to generate the published search strategies that we replicated, the complete process of generating and conducting the naïve search, creating a new search with our method,

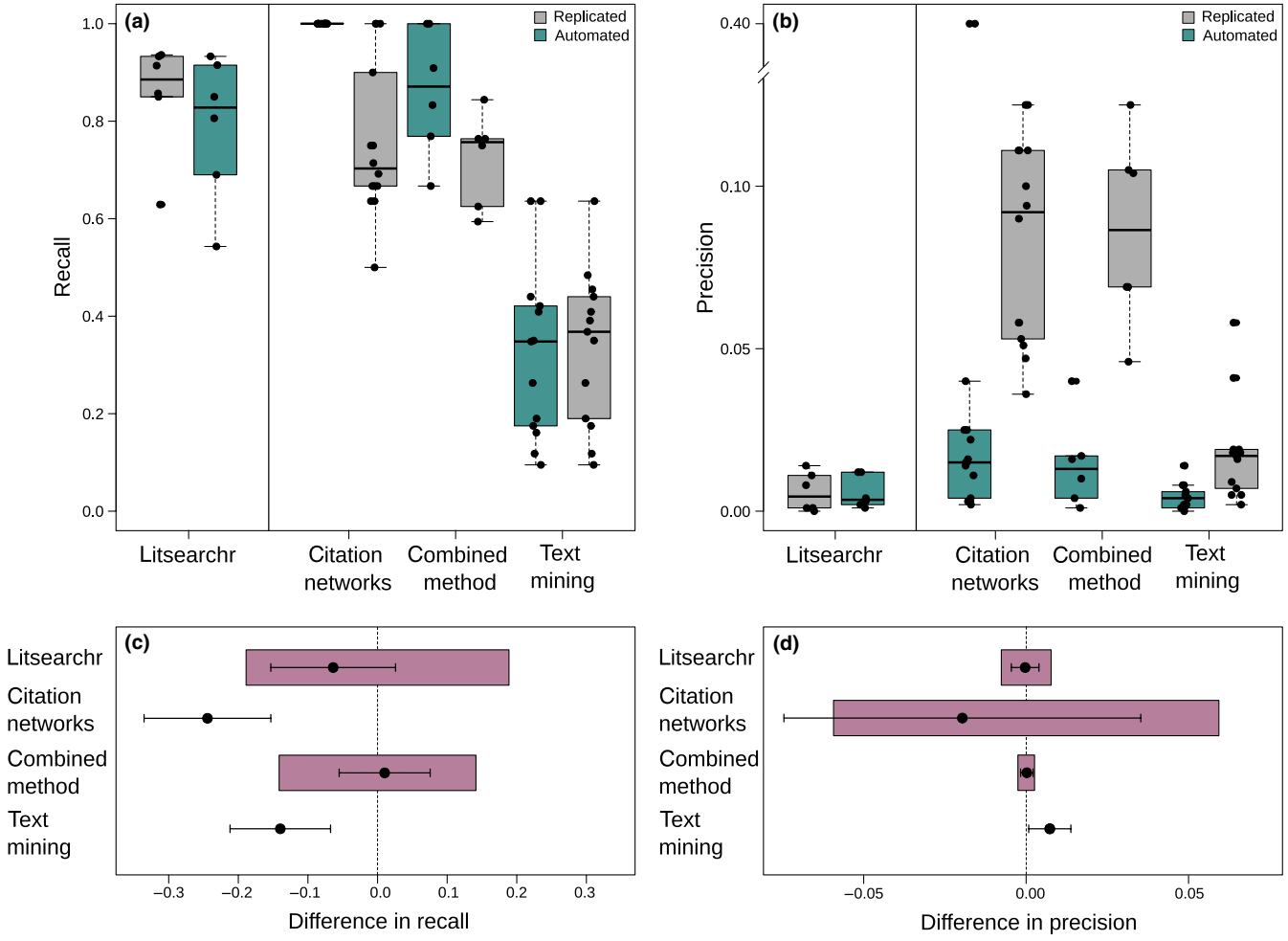


FIGURE 4 Performance of `litsearchr` in comparison to conventional search methods and identical comparisons reported for other partially automated approaches. The proportion of articles included in the original published review retrieved by the replicated and automated searches (recall) for each approach is shown in (a) and the precision of each search is shown in (b). Each pair of boxes compares conventionally developed searches (grey) to partially automated searches (teal) generated with `litsearchr` and other approaches. The results for citation networks (Belter, 2016), a combination of citation networks and text mining (Sarol et al., 2018), or text mining alone (Hausner et al., 2015) come from the raw data published in the respective studies for those methods. Black dots indicate recall and precision of each study included in the sample. The horizontal line of boxes shows the median recall and precision and whiskers give 95% confidence intervals. High recall indicates good sensitivity and retrieval of relevant articles; high precision indicates a narrow search with high relevancy of all retrieved articles. Differences between automated and replicated searches of the original conventional search for recall (c) and precision (d) are shown with the mean difference and confidence intervals from a pairwise *t*-test. Positive values favor the automated method. For methods with no significant differences ($p < .05$) between automated and replicated searches, pink boxes indicate the range within which differences would not be detectable based on a power analysis with alpha = 0.050 and power = 0.800. Data used to calculate precision, recall, and test statistics are available on Dryad (Grames et al., 2019b)

and checking the results for precision and recall took an average of 1.7 hr (*SD* = 0.7) per replicated review. By comparison, conventional methods take 8–23 hr for information specialists to develop search strategies (Hausner et al., 2012) and the assembly and deduplication process has been shown to average 1.37 days (Haddaway & Westgate, 2019). There were no significant differences in precision (t = −0.197; df = 5; p = 0.852; $\Delta\mu$ = −0.000) or recall (t = −1.827; df = 5; p = 0.127; $\Delta\mu$ = −0.064) between the replicated searches and our method. The power analyses indicated that we would have been able to detect a difference in means of at least −0.008 for precision and −0.189 for recall (Figure 4). These results indicate that our method is approximately 100% (at worst, 99.2%) as precise as conventional methods; there are no detectable differences in the percent of all search results retrieved that were gold standard hits. Our method is also approximately 93.6% (at worst, 81.1%) as good at recovering gold standard hits (recall) as manually-created searches. When the results of the naïve searches were combined with the results of the final `litsearchr` search strategy, the recall increased to 95.4% (t = −1.271; df = 5; p = 0.260; $\Delta\mu$ = −0.046); precision is not calculable due to overlap in databases accessed for the naïve and final searches. In five of the six replicated searches (Bernes, Bråthen, Forbes, Speed, & Moen, J., 2015; Haddaway et al., 2017; Land et al., 2016; Laverick et al., 2018; Villemey et al., 2018), `litsearchr` retrieved gold-standard articles not found by the replicated search terms in the same databases, meaning that the search string reported in the published review was incapable of retrieving all the articles that were ultimately included in the review. All four quasi-automated methods for developing search strategies were as precise as, or better than, the replicated searches, but only the text-mining approaches—Hausner et al. (2015) and `litsearchr`—matched replicated searches in terms of recall (Figure 4).

4 | DISCUSSION

The method we describe facilitates objective, reproducible search strategy development for systematic reviews and performs as well as conventional methods. As an unexpected benefit, `litsearchr` may also be able to capture articles relevant to published reviews that were not found with the published search strategies because it identifies synonymous phrases that were not included in the original search terms. `litsearchr` retrieved articles not returned by replicated searches from published search strategies, indicating that the published reviews may have found these articles by scanning references of included studies or with other manual search methods. This result suggests that our method can identify articles that are missed by conventionally-developed search strategies, even though the naïve searches were written by non-experts on the topic. Further testing is needed to determine if search results uniquely retrieved by our method meet the inclusion criteria of the published reviews but were not found as part of the original systematic review. Combining expert knowledge with our quasi-automated method could lead to improvements in search recall, especially for fields with non-standardized or nuanced terminology that lack formal ontologies. Additionally, litsearchr greatly reduces the amount of time required to conduct a systematic review by decreasing time spent on search strategy development and administrative tasks like assembly and deduplication. By reducing the time needed to develop a search strategy and assemble and deduplicate the results, our method makes large systematic reviews and meta-analyses more feasible than doing them with conventional approaches and is also well-suited for rapid reviews.

In future versions of `litsearchr`, we plan to add support for conducting the naïve search and word stemming in languages other than English. To facilitate searches of gray literature, we also plan to add the ability to write searches using an automatically-identified critical subset of keywords for web searches or databases with strict character limits (e.g. Google, Google Scholar, and JSTOR). Finally, we plan to add functions that export complete search strategies and search strategy development for easy reporting and reproducibility.

ACKNOWLEDGMENTS

Many thanks go to E. Hennessey, B. Johnson, J. Livingston and C. Mills for insight into conventional search strategy development, and to N. Haddaway for helpful comments on an earlier version of this manuscript. Thanks to R. Bagchi for the suggestion to use co-occurrence networks to identify important nodes and to T. Wisneskie and T. Woerfel for identifying bugs in the package. E.M.G. was funded by a University of Connecticut Outstanding Scholars Fellowship.

AUTHORS' CONTRIBUTIONS

E.M.G., M.W.T. and C.S.E. conceived the project. E.M.G. and A.N.S. developed and tested the method. E.M.G. wrote the code for `litsearchr`. E.M.G. drafted the manuscript and all co-authors contributed critically. All authors approved the final manuscript before submission.

DATA AVAILABILITY STATEMENT

The `litsearchr` package v0.1.0 (Grames et al., 2019a) release is at https://doi.org/10.5281/zenodo.2551701, and the unstable version is hosted at https://github.com/elizagrames/litsearchr. Additional documentation is at https://elizagrames.github.io/litsearchr. Data are deposited in the Dryad Digital Repository https://doi.org/10.5061/dryad.n1kv40m (Grames, Stillman, Tingley, & Elphick, 2019b).

ORCID

Eliza M. Grames https://orcid.org/0000-0003-1743-6815
Andrew N. Stillman https://orcid.org/0000-0001-6692-380X
Morgan W. Tingley https://orcid.org/0000-0002-1477-2218

REFERENCES

Baker, C. (2018). rapidraker: Rapid Automatic Keyword Extraction (RAKE) alrogithm. R package version 0.1.0.

Belter, C. W. (2016). Citation analysis as a literature search method for systematic reviews. *Journal of the Association for Information Science and Technology*, *67*, 2766–2777.

Benoit, K. (2018). quanteda: Quantitative analysis of textual data. R package version 1.3.4.

Bernes, C., Bråthen, K. A., Forbes, B. C., Speed, J. D., & Moen, J. (2015). What are the impacts of reindeer/caribou (*Rangifer tarandus* L.) on arctic and alpine vegetation? A systematic review. *Environmental Evidence*, *4*, 4.

Bramer, W. M., de Jonge, G. B., Rethlefsen, M. L., Mast, F., & Kleijnen, J. (2018). A systematic approach to searching: An efficient and complete method to develop literature searches. *Journal of the Medical Library Association*, *106*, 531.

Bramer, W. M., Rethlefsen, M. L., Mast, F., & Kleijnen, J. (2018). Evaluation of a new method for librarian-mediated literature searches for systematic reviews. *Research Synthesis Methods*, *9*, 510–520.

Cochrane, A. L. (1972). *Effectiveness and efficiency: random reflections on realth services*. London: Nuffield Provincial Hospitals Trust.

Côté, I. M., Curtis, P. S., Rothstein, H. R., & Stewart, G. B. (2013). Gathering data: Searching literature and selection criteria. In J. Koricheva, J. Gurevitch, & K. Mengersen (Eds.), *Handbook of meta-analysis in ecology and evolution* (pp. 37–51). Princeton, NJ: Princeton University Press.

Delaney, A., & Tams, P. A. (2018). Searching for evidence or approval? A commentary on database search in systematic reviews and alternative information retrieval methodologies. *Research Synthesis Methods*, *9*, 124–131.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*, 3–8.

Grames, E. M., Stillman, A. N., Tingley, M. W., & Elphick, C. S. (2019a). litsearchr: Automated search term selection and search strategy for systematic reviews. R package version 0.1.0. https://doi.org/10.5281/zenodo.2551701.

Grames, E. M., Stillman, A. N., Tingley, M. W., & Elphick, C. S. (2019b). Data from: An automated approach to identifying search terms for systematic reviews using keyword co-occurrence networks. *Dryad Digital Repository*, https://doi.org/10.5061/dryad.n1kv40m

Gurevitch, J., Curtis, P. S., & Jones, M. H. (2001). Meta-analysis in ecology. *Advances in Ecological Research*, *32*, 199–247.

Haddaway, N. R., Bernes, C., Jonsson, B. G., & Hedlund, K. (2016). The benefits of systematic mapping to evidence-based environmental management. *Ambio*, *45*, 613–620. https://doi.org/10.1007/s13280-016-0773-x

Haddaway, N. R., Hedlund, K., Jackson, L. E., Kätterer, T., Lugato, E., Thomsen, I. K., … Isberg, P.-E. (2017). How does tillage intensity affect soil organic carbon? *A Systematic Review. Environmental Evidence*, *6*, 30. https://doi.org/10.1186/s13750-017-0108-9

Haddaway, N. R., Macura, B., Whaley, P., & Pullin, A. S. (2018). ROSES RepOrting standards for Systematic Evidence Syntheses: Pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. *Environmental Evidence*, *7*, 7. https://doi.org/10.1186/s13750-018-0121-7

Haddaway, N. R., & Westgate, M. J. (2019). Predicting the time needed for environmental systematic reviews and systematic maps. *Conservation Biology*, *33*, 434–443. https://doi.org/10.1111/cobi.13231

Haddaway, N., Woodcock, P., Macura, B., & Collins, A. (2015). Making literature reviews more reliable through application of lessons from systematic reviews. *Conservation Biology*, *29*, 1596–1605. https://doi.org/10.1111/cobi.12541

Hausner, E., Guddat, C., Hermanns, T., Lampert, U., & Waffenschmidt, S. (2015). Development of search strategies for systematic reviews: Validation showed the noninferiority of the objective approach. *Journal of Clinical Epidemiology*, *68*, 191–199. https://doi.org/10.1016/j.jclinepi.2014.09.016

Hausner, E., Guddat, C., Hermanns, T., Lampert, U., & Waffenschmidt, S. (2016). Prospective comparison of search strategies for systematic reviews: An objective approach yielded higher sensitivity than a conceptual one. *Journal of Clinical Epidemiology*, *77*, 118–124. https://doi.org/10.1016/j.jclinepi.2016.05.002

Hausner, E., Waffenschmidt, S., Kaiser, T., & Simon, M. (2012). Routine development of objectively derived search strategies. *Systematic Reviews*, *1*, 19. https://doi.org/10.1186/2046-4053-1-19

Hunt, M. (1997). *How science takes stock: The story of meta-analysis*. New York City, NY: Russell Sage Foundation.

Järvinen, A. (1991). A meta-analytic study of the effects of female age on laying-date and clutch-size in the great tit *Parus major* and the pied flycatcher *Ficedula hypoleuca*. *Ibis*, *133*, 62–67. https://doi.org/10.1111/j.1474-919X.1991.tb04811.x

Jonnalagadda, S. R., Goyal, P., & Huffman, M. D. (2015). Automating data extraction in systematic reviews: A systematic review. *Systematic Reviews*, *4*, 78. https://doi.org/10.1186/s13643-015-0066-7

Koricheva, J., Gurevitch, J., & Mengersen, K. (2013). *Handbook of Meta-analysis in Ecology and Evolution*. Princeton, NJ: Princeton University Press.

Przybyla, P., Brockmeier, A. J., Kontonatsios, G., Le Pogam, M. A., McNaught, J., Elm, E., … Ananiadou, S. (2018). Prioritising references for systematic reviews with RobotAnalyst: A user study. *Research Synthesis Methods*, *9*, 470–488.

Land, M., Granéli, W., Grimvall, A., Hoffmann, C. C., Mitsch, W. J., Tonderski, K. S., & Verhoeven, J. T. (2016). How effective are created or restored freshwater wetlands for nitrogen and phosphorus removal? *A Systematic Review. Environmental Evidence*, *5*, 9.

Laverick, J., Piango, S., Andradi-Brown, D., Exton, D., Bongaerts, P., Bridge, T., … Rogers, A. (2018). To what extent do mesophotic coral ecosystems and shallow reefs share species of conservation interest? *A Systematic Review. Environmental Evidence*, *7*, 16. https://doi.org/10.1186/s13750-018-0127-1

O'Mara-Eves, A., Thomas, J., McNaught, J., Miwa, M., & Ananiadou, S. (2015). Using text mining for study identification in systematic reviews: A systematic review of current approaches. *Systematic Reviews*, *4*, 5. https://doi.org/10.1186/2046-4053-4-5

Paisley, S., & Foster, M. J. (2018). Innovation in information retrieval methods for evidence synthesis studies. *Research Synthesis Methods*, *9*, 506–509.

Porter, M. (1980). An algorithm for suffix stripping. *Program*, *14*, 130–137. https://doi.org/10.1108/eb046814

Pullin, A. S., & Knight, T. M. (2001). Effectiveness in conservation practice: Pointers from medicine and public health. *Conservation Biology*, *15*, 50–54. https://doi.org/10.1111/j.1523-1739.2001.99499.x

Pullin, A. S., & Knight, T. M. (2009). Doing more good than harm–building an evidence-base for conservation and environmental management. *Biological Conservation*, *142*, 931–934. https://doi.org/10.1016/j.biocon.2009.01.010

Pullin, A. S., & Stewart, G. B. (2006). Guidelines for systematic review in conservation and environmental management. *Conservation Biology*, *20*, 1647–1656. https://doi.org/10.1111/j.1523-1739.2006.00485.x

Radhakrishnan, S., Erbis, S., Isaacs, J. A., & Kamarthi, S. (2017). Novel keyword co-occurrence network-based methods to foster systematic reviews of scientific literature. *PLoS ONE*, *12*, e0172778.

Rathbone, J., Hoffmann, T., & Glasziou, P. (2015). Faster title and abstract screening? Evaluating Abstrackr, a semi-automated online screening program for systematic reviewers. *Systematic Reviews*, *4*, 80. https://doi.org/10.1186/s13643-015-0067-6

Richardson, W. S., Wilson, M. C., Nishikawa, J., & Hayward, R. S. (1995). The well-built clinical question: A key to evidence-based decisions. *ACP Journal Club*, *123*, A12–A12.

Rose, S., Engel, D., Cramer, N., & Cowley, W. (2010). Automatic keyword extraction from individual documents. In M. W. Berry, & J. Kogan (Eds.), *Text mining* (pp. 1–20). Chichester, UK: John Wiley & Sons Ltd.

Samson, D., & Schoelles, K. M. (2012). Chapter 2: Medical tests guidance (2) developing the topic and structuring systematic reviews of medical tests: Utility of PICOT S, analytic frameworks, decision trees, and other frameworks. *Journal of General Internal Medicine*, *27* , 11–19. https://doi.org/10.1007/s11606-012-2007-7

Sarol, J., Liu, L., & Schneider, J. (2018). Testing a citation and text-based framework for retrieving publications for literature reviews. *Proceedings of the 7th International Workshop on Bibliometric-enhanced Information Retrieval (BIR 2018)*, pp. 22–33.

Shemilt, I., Simon, A., Hollands, G. J., Marteau, T. M., Ogilvie, D., O'Mara-Eves, A., … Thomas, J. (2014). Pinpointing needles in giant haystacks: Use of text mining to reduce impractical screening workload in extremely large scoping reviews. *Research Synthesis Methods*, *5*, 31–49. https://doi.org/10.1002/jrsm.1093

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, *32*, 752. https://doi.org/10.1037/0003-066X.32.9.752

Spiriti, S., Eubank, R., Smith, P. W., & Young, D. (2013). Knot selection for least-squares and penalized splines. *Journal of Statistical Computation and Simulation*, *83*, 1020–1036. https://doi.org/10.1080/00949655.2011.647317

Spiriti, S., Smith, P., & Lecuyer, P. (2018). freeknotsplines: Algorithms for implementing free-knot splines. R package version 1.0.1.

Stansfield, C., O'Mara-Eves, A., & Thomas, J. (2017). Text mining for search term development in systematic reviewing: A discussion of some methods and challenges. *Research Synthesis Methods*, *8*, 355–365. https://doi.org/10.1002/jrsm.1250

Su, H. N., & Lee, P. C. (2010). Mapping knowledge structure by keyword co-occurrence: A first look at journal papers in technology foresight. *Scientometrics*, *85*, 65–79. https://doi.org/10.1007/s11192-010-0259-8

Sutherland, W. J., & Wordley, C. F. (2018). A fresh approach to evidence synthesis. *Nature*, *558*, 364–366. https://doi.org/10.1038/d41586-018-05472-8

Tsafnat, G., Glasziou, P., Choong, M. K., Dunn, A., Galgani, F., & Coiera, E. (2014). *Systematic Review Automation Technologies*. *Systematic Reviews*, *3*, 74. https://doi.org/10.1186/2046-4053-3-74

Tsertsvadze, A., Chen, Y. F., Moher, D., Sutcliffe, P., & McCarthy, N. (2015). How to conduct systematic reviews more expeditiously? *Systematic Reviews*, *4*, 160. https://doi.org/10.1186/s13643-015-0147-7

Villemey, A., Jeusset, A., Vargac, M., Bertheau, Y., Coulon, A., Touroult, J., … Sordello, R. (2018). Can linear transportation infrastructure verges constitute a habitat and/or a corridor for insects in temperate landscapes? *A Systematic Review. Environmental Evidence*, *7*, 5. https://doi.org/10.1186/s13750-018-0117-3

Zhang, H., Babar, M. A., Bai, X., Li, J., & Huang, L. (2011). An empirical assessment of a systematic search process for systematic reviews. *15th Annual Conference on Evaluation Assessment in Software Engineering (EASE 2011)*.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Grames EM, Stillman AN, Tingley MW, Elphick CS. An automated approach to identifying search terms for systematic reviews using keyword co-occurrence networks. *Methods Ecol Evol*. 2019;10:1645–1654. https://doi.org/10.1111/2041-210X.13268