Abstract

The paper aims to examine the visual exploration tool “Connected Papers” (www.connectedpapers.com), which identifies relevant literature based on content similarities and displays the result in the form of visual clusters. Connected Papers searches through Semantic Scholar literature corpus and discovers the most relevant related research papers using a specialized algorithm. The researcher will have to identify the most relevant paper as an “origin paper” among the retrieved papers. The origin paper acts as a base for the formation of a literature graph based on similarities. The unique features (i.e. “Prior” and “Derivate”) help a researcher to identify the most
relevant literature. In this paper, the topic “Use of drones in Agriculture” is selected to demonstrate the process of literature exploration using Connected Papers. A set of relevant papers is retrieved, out of which the researcher will have to select one of the most appropriate relevant papers. That paper is termed as the origin paper to create the visualization of literature. Connected Papers suggests the most relevant papers based on the search keywords, but the graph is solely based on a researcher's judgement when selecting an origin paper. As explained in the paper, the researcher of other domains may adopt the process to understand the literature mapping phenomena for their own discipline.

Keywords: Connected Papers, Literature visualization, Drones, Agriculture, Bibliographic coupling, Citation

Recommended Citation:


Introduction

The literature search is one of the essential parts of research. Finding the most relevant literature is done manually by using keywords to search relevant studies through different databases. There is a need for an automated system which will identify the most relevant articles for a topic. In the case of manual query-based searching in the traditional databases, there is a chance that relevant keywords may not be included in the search (i.e. a search for “Drones” and “Agriculture” may exclude “UAV” (Unmanned Aerial Vehicle)). Connected Papers (www.connectedpapers.com) has easily addressed this problem. The results identify the similarity and list a reasonable number of the most connected articles. This paper demonstrates the utility and effectiveness of using Connected Papers for the visual exploration of literature.

Background

About Connected Papers

Connected Papers defines itself as a unique, visual tool to help researchers and applied scientists find and explore papers relevant to their field of work (2023). Connected Papers is an easy to use and powerful visualization tool. The user has to select one of the most relevant papers from the search result, which is considered to be an “origin paper”. On the basis of the origin paper, the results will be displayed in the form of a graph. The graph indicates the co-relation among papers, not necessarily the number of times a paper has been cited.

History

Connected Papers was the brainchild of four friends named Alex Tarnavsky Eitan, Eddie Smolyansky, Itay Knaan Harpaz and Sahar Perets with the developer Ofer
Mustigman in 2020. The basic idea was to find a solution to the complexity of extensive literature exploration. The basic idea was to simplify the complexities of extensive literature exploration by selecting a relevant paper as the *origin paper*, and then Connected Papers discovers other papers connected to the same theme.

**Accessing Connected Papers**

Connected Papers can be accessed with or without logging in (for trial access, a maximum of two graphs is allowed). To log in, the user must register with an email address. There are two types of accounts: one is free; and the other is Premium (Academic and Business). The free version limits users to five graphs per month, whereas the premium account has no limitations.

**Workflow**

Once the researcher chooses an *origin paper*, Connected Papers analyses 50,000 publications, picks a few with strong links, and creates a graph based on similarity.

The graph is based on the concept of co-citations and bibliographic coupling. If the citations and references in two papers are very similar, it's likely that they are closely related. Connected Papers search strategy is based on the algorithm that results in a force-directed graph to arrange the relevant papers into visual clusters in a group by isolating the irrelevant papers. Within the graph, each node's shortest path to the *origin paper* is represented in the similarity space. Connected Papers accesses data from the Semantic Scholar Paper Corpus, which is licensed under ODC-BY.

**Topic of Research (Search Query): “Use of Drones in Agriculture.”**

In this paper, we have considered “Use of drones in agriculture and its future” by M. M. Özgüven, et al. as a base or origin paper and to demonstrate the process of literature exploration using Connected Papers. Drones or Unmanned Aerial Vehicles (UAV) play a vital role in crop management, such as crop monitoring and pest control, analysis of soil moisture, planting and harvesting.

**Exploration of Connected Papers**

When a query is entered in Connected Papers, it generates a list of the most relevant papers based on co-citation. Among them, one most desired paper will have to be selected, known as the *origin paper*. Based on the *origin paper*, Connected Papers generates a graph using a specialized algorithm, based on citation and bibliographic coupling. To build a graph, the researcher has to enter one of the following: a search string; a paper title; a DOI; or the URL of an arXiv, a Semantic scholar, or a PubMed paper. The researcher then selects *Build Graph* to generate a graph, as shown in Figure 1.
Search Results and the Selection of an Origin Paper to Build a Graph

An initial list of papers most relevant to the initial search is retrieved (based on the search “use of drones in agriculture”), as shown in Figure 2.

The researcher must then select the most relevant paper, termed *origin paper*, to build the literature visualization.
Connected Papers Graph Based on Origin Paper

Based on the sample *origin paper*, Connected Papers retrieved around 40 papers to build a graph based on the similarities and bibliographic coupling. The visual overview in the form of a graph is shown in Figure 3. The components of the literature graph page are elaborated as below:

**A:** The literature graph of similar clusters is based on the *origin paper*; for our study, the origin paper was “Use of drones in agriculture and its future”.

**B:** Shows the visual presentation of the topic in the form of a graph

Similarity space is the context in which the relationship of the papers is positioned according to the content, themes, or other relevant factors embedded by them, even if the papers do not cite each other directly. Connected Papers utilizes a force-directed graph to arrange the papers in a manner that visually groups similar papers together and separates less similar ones. When a node is selected, the platform highlights the shortest path from that node to the *origin paper* within the similarity space.

**How to Read the Graph**

Each node in the graph represents an academic paper related to an *origin paper*, and the papers are arranged based on their similarity. The size of each node in the graph represents the number of citations the paper has received. The color of each node represents the year in which the paper was published, with brighter colors indicating more recent publications. Papers that are similar to each other are connected by strong lines and cluster together in the graph.
C: Expand

Expanding the list of papers display their citations, which contain the title, author/s, year of publication, citations, number of references, and percentage of similarity with the origin paper; fields can be sorted in the desired order. There is also an option to download the bibliographic references. The downloaded file will be in BibTeX format, and that can easily be imported to any reference management software to read. The expanded form of the list is shown in Figure 4. An abbreviated list of relevant papers is on the left side of the graph. This list can be scrolled up and down.

D: Bibliographic details

Any paper selected from the list (D), along with its bibliographic details (author, title, citations, abstract, full text link), will appear on the right side of the graph.

E: Full text link options

For any paper selected from the list, the paper details appear on the right side of the graph. Links are provided to the full text of the paper available from various providers (i.e. PDF, DOI, Google scholar page, PubMed, and Semantic scholar).

F: Prior and derivative works

Prior works and derivative works are shown in Figures 5 and 6, respectively.

Prior works refer to the papers that are most frequently cited by the papers in the graph. Typically, these papers are older works considered foundational, significant, and/or influential in the field, making it beneficial to become acquainted with them. “Selecting a prior work will highlight all graph papers referencing it, and selecting a
Derivative works are the papers that are frequently cited by many of the papers within the graph but are newer publications influenced by the origin paper. Typically, these papers serve as surveys of the field or recent works that draw inspiration from multiple papers in the graph. “Selecting a derived work will highlight all graph papers cited by it, and selecting a graph paper will highlight all derivative works citing it” (Connected Papers, 2023, “Derivative works” section).

Interest in the use of unmanned aerial systems (UAS) to estimate the aboveground biomass (ABB) of vegetation in agricultural and non-agricultural settings is growing rapidly but there is no standardized methodology for planning, collecting and analyzing UAS data for this purpose. We synthesized 46 studies from the peer-reviewed literature to provide the first-ever review on the subject. Our analysis showed that spectral and structural data from UAS imagery can accurately estimate vegetation biomass in a variety of settings, especially when both data types are combined.

Figure 5. Expanded form of “Prior works”

Figure 6. Expanded form of “Derivatives works”
As compared to the traditional literature search strategy, Connected Papers works well as it takes the most relevant paper identified by the researcher as origin paper and finds other Connected Papers of the searched domain. Connected Papers results into two distinctive sets of relevant papers. The set of most commonly cited papers in the graph (known as prior works) and the papers that cite many of the papers of the graph (known as derivative work) are the main outcomes of the Connected Papers. The outcome could be the parameter for identifying the most relevant literature in the domain of use of drones in agriculture and other fields. This can help to identify new collaborative research domains that are ongoing in agriculture and other fields.

**Conclusion**

Connected Papers is an innovative research tool that allows a researcher to explore and analyze literature in a specific field in an effective way. Some important features are visualization, clustering, citation generation, and auto paper recommendation. The exported data can be used in reference management software like Mendeley, Zotero, Endnote, etc. The algorithm used by Connected Papers makes finding relevant papers easy and efficient. An API to connect these platforms would add more value to it.

The tool's effectiveness depends on the comprehensiveness and diversity of the papers within the Semantic Scholar Paper Corpus. If certain fields, subfields, or regions are underrepresented in the corpus, the tool's results may not provide a complete or balanced view. Connected Papers is a relatively new platform and is still evolving, so it may not yet have all the features and capabilities that some researchers require. This can limit its usefulness for certain research works of researchers with specialized needs. Connected Papers can help researchers to scrutinize relevant papers by starting with the most relevant paper, and retrieving the most relevant papers based on co-citation and bibliographic coupling. Due to limited coverage of the Semantic Scholar Corpus, Connected Papers can consider wider literature coverage by utilizing open databases and social networks like Pubmed, Google Scholar, arXiv, ResearchGate, Academia, etc. The use of Connected Papers for mapping and visualizing networks can provide significant benefits for researchers in their research pursuits, enabling them to make resource discoveries in an efficient manner.

**References**

**Connected Papers.** (2023, April 9). Connected Papers | Find and explore academic papers. https://www.connectedpapers.com/about

This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).