

Using Bibliometric Data and Analysis Tools to Identify Emerging Research Opportunities



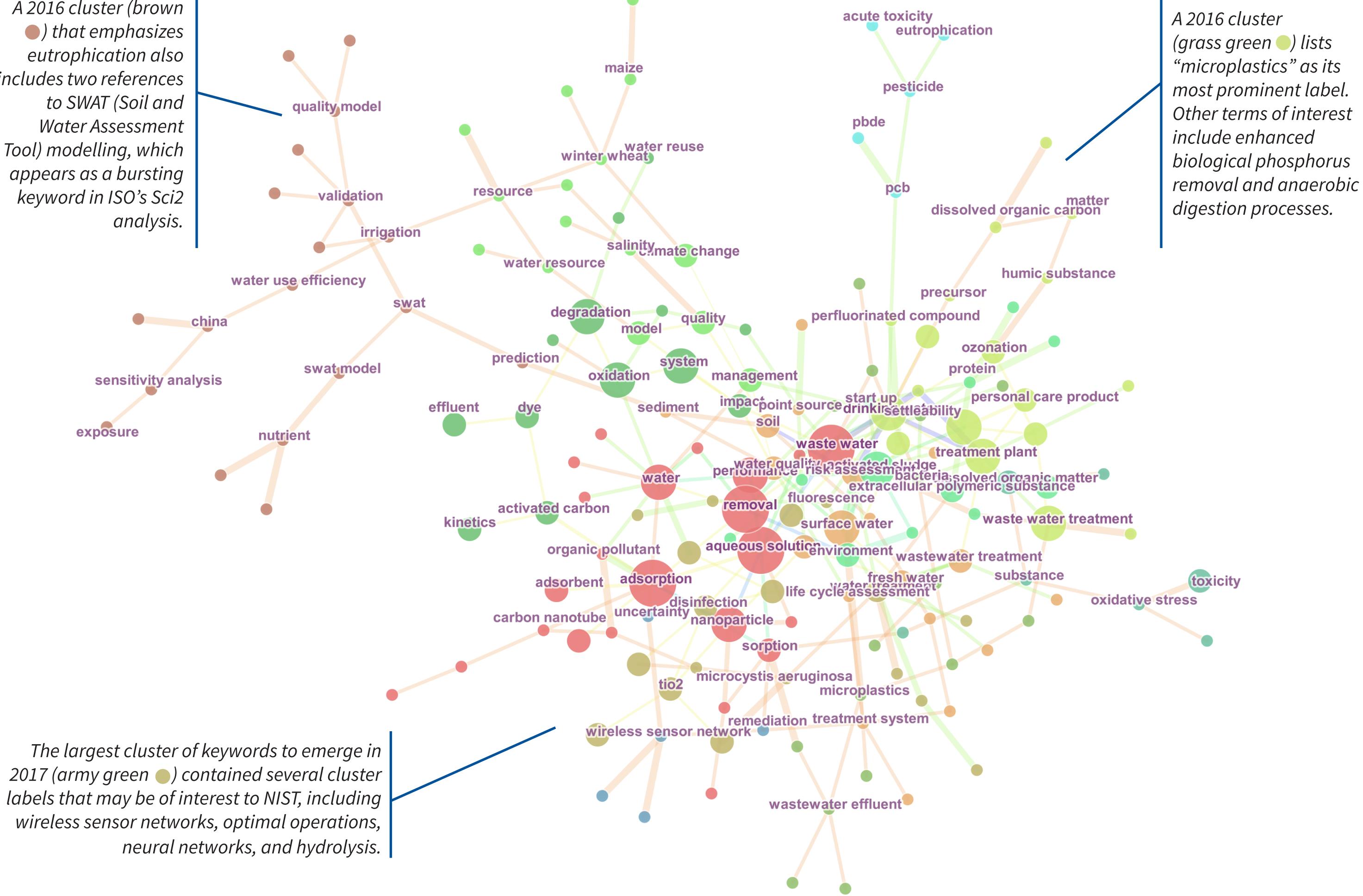
Amanda Malanowski Susan Makar **Amy Trost** susan.makar@nist.gov amanda.malanowski@nist.gov amy.trost@nist.gov

Introduction

The Information Services Office (ISO) at the National Institute of Standards and Technology (NIST), a non-regulatory agency within the U.S. Department of Commerce, is working to identify emerging research areas where NIST can have the greatest impact on industry and society. ISO staff analyzed bibliometric data in order to reveal potential topics of interest in water quality metrology, a strategic priority area for NIST's Material Measurement

Emerging Keyword Clusters

A 2016 cluster (brown) that emphasizes eutrophication also includes two references



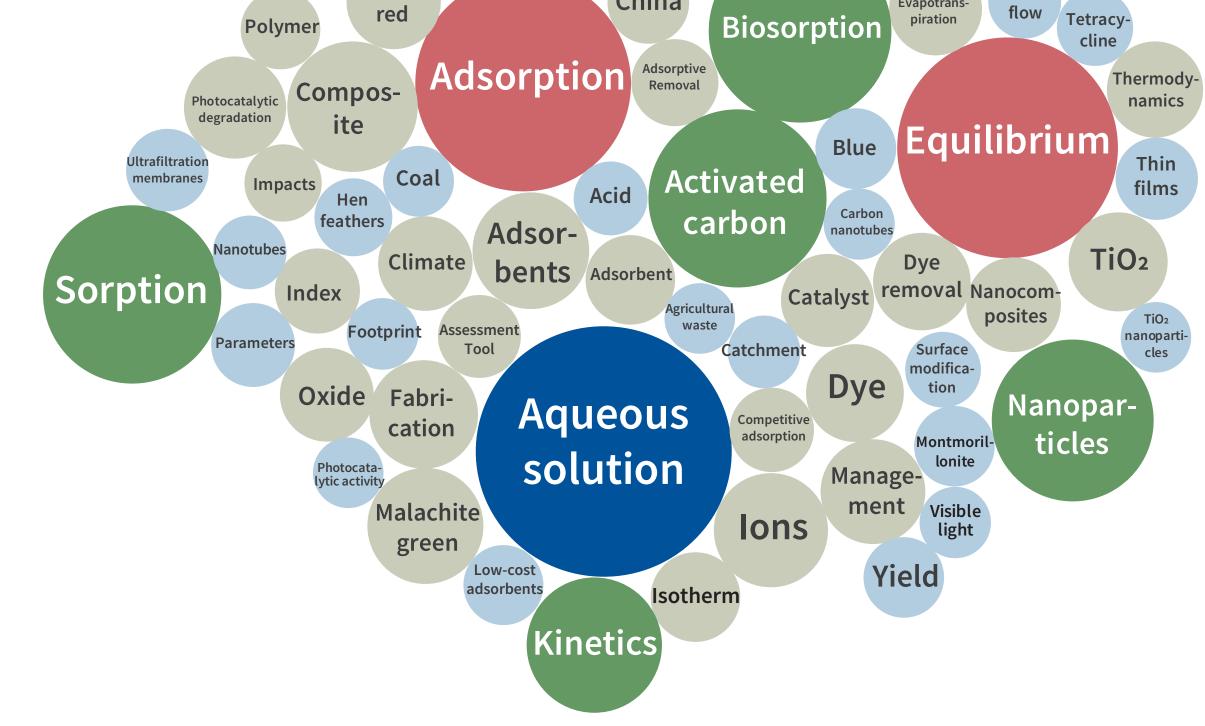
(grass green –) lists *"microplastics" as its*

Laboratory (MML).

Methodology

ISO's search strategy was created using measurement-related keywords identified from two MML water strategy documents and searching for those keywords within 31 water research journals. A *Web of Science* search identified 10,714 journal articles for the years 2012-2017. The full paper set was analyzed with the Sci2 tool. Analysis in R and CiteSpace focused on the 1,000 papers in this set that had received the most usage in *Web of Science* over the past 180 days. An analysis of *Essential Science Indicators* Research Fronts was used to evaluate the *CiteSpace* model.

Keyword Bursts SWA mode **1ethylene** Removal Chitosa Cu-Li Decay



Using Kleinburg's burst detection algorithm within the Sci2 tool and an approach identified by Guo et al. (2011), ISO identified keywords that have increased in frequency over time. An illustration of all bursts detected in 2017 provides clues about new research methods and instrumentation in water quality metrology.

Clustered Abstract Terms

— infrared

2017 (army green) contained several cluster labels that may be of interest to NIST, including wireless sensor networks, optimal operations,

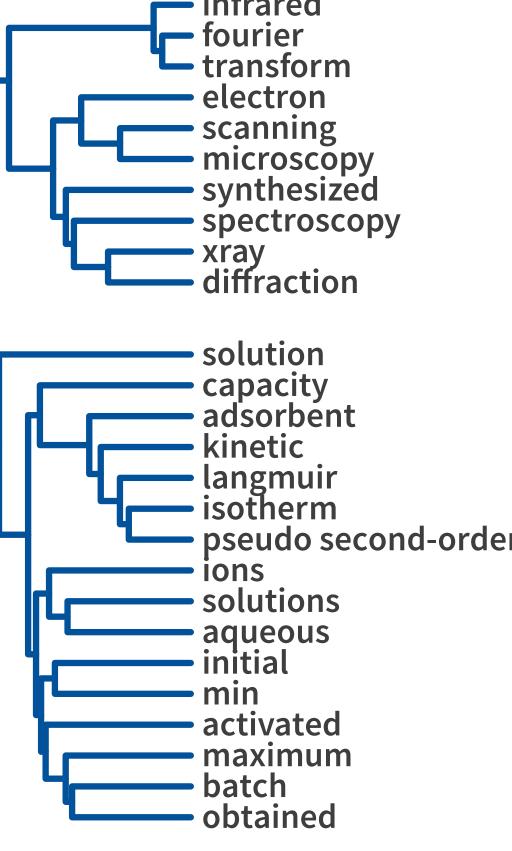
With the CiteSpace program, ISO created a network model that groups co-occurring keywords. Keywords that appear in a larger number of articles have larger nodes. A spectral clustering algorithm (Chen et al. 2010) identified 12 groups, or clusters, of keywords. Five of these clusters emerged in 2016 or 2017. Several clusters, defined by color, are described above.

Model Comparison

Of the three approaches shown here, the emerging keywords model (pictured above) provided the most potential for further exploration; large clusters of keywords can be connected to pertinent papers and linked to source material in the *CiteSpace* interface. ISO compared the results of this model to 22 groups of papers found in *Essential Science Indicators (ESI)* Research Fronts. An *ESI* Research Front is a group of highly cited papers over a five-year period in a specialized topic defined by a cluster analysis.

Conclusions

ISO's analyses revealed several potential topics of interest to NIST researchers, including wireless sensor networks, neural networks, SWAT modeling, microplastics, biological phosphorus removal, and anaerobic digestion processes.



A cluster analysis of abstract terms in R with 300 stop words removed—groups like terms in a hierarchical fashion. This type of analysis can indicate term linkages. In ISO's analysis, however, this method did not provide much useful information due to its reliance on single words (rather than phrases) and more commonly used terms.

Three of the most recent paper groups identified in the ESI search are all mirrored by a slightly broader *CiteSpace* cluster. For example, a Research Front for silver nanoparticles mirrored one of the larger *CiteSpace* clusters, Nanoparticles.

The Research Fronts paper groups seemed to be more closely related, while *CiteSpace* groupings often included apparently disparate topics. However, ISO's *CiteSpace* analysis was able to uncover several recent cluster topics, including wireless sensor networks and microplastics, that could not be retrieved in a search of Research Fronts.

Future efforts will focus on smaller sets of papers identified with the help of subject matter experts in order to conduct a deep analysis of the topics that are most pertinent to the strategic planning process at NIST.

References

Chen, Chaomei, Fidelia Ibekwe-SanJuan, and Jianhua Hou. "The structure and dynamics of cocitation clusters: A multiple-perspective co-citation analysis." *Journal of the American Society for Information Science and Technology* 61, no. 7 (2010): 1386-1409.

Fang, Yuqing. "Visualizing the structure and the evolving of digital medicine: a scientometrics review." *Scientometrics* 105, no. 1 (2015): 5-21. doi:10.1007/s11192-015-1696-1.

Guo, H. N., S. Weingart, and K. Börner. "Mixed-indicators model for identifying emerging research areas." Scientometrics 89, no. 1 (2011): 421-35. doi:10.1007/s11192-011-0433-7.

NOTE: Identification of commercial entities in this poster is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology.

Special Libraries Association 2017 Conference, Phoenix, Arizona, June 2017