

Network Evolution of Scholarly Communication on CRISPR Gene Editing

Hong Chen, Hirokazu Shirado
Carnegie Mellon University

Abstract

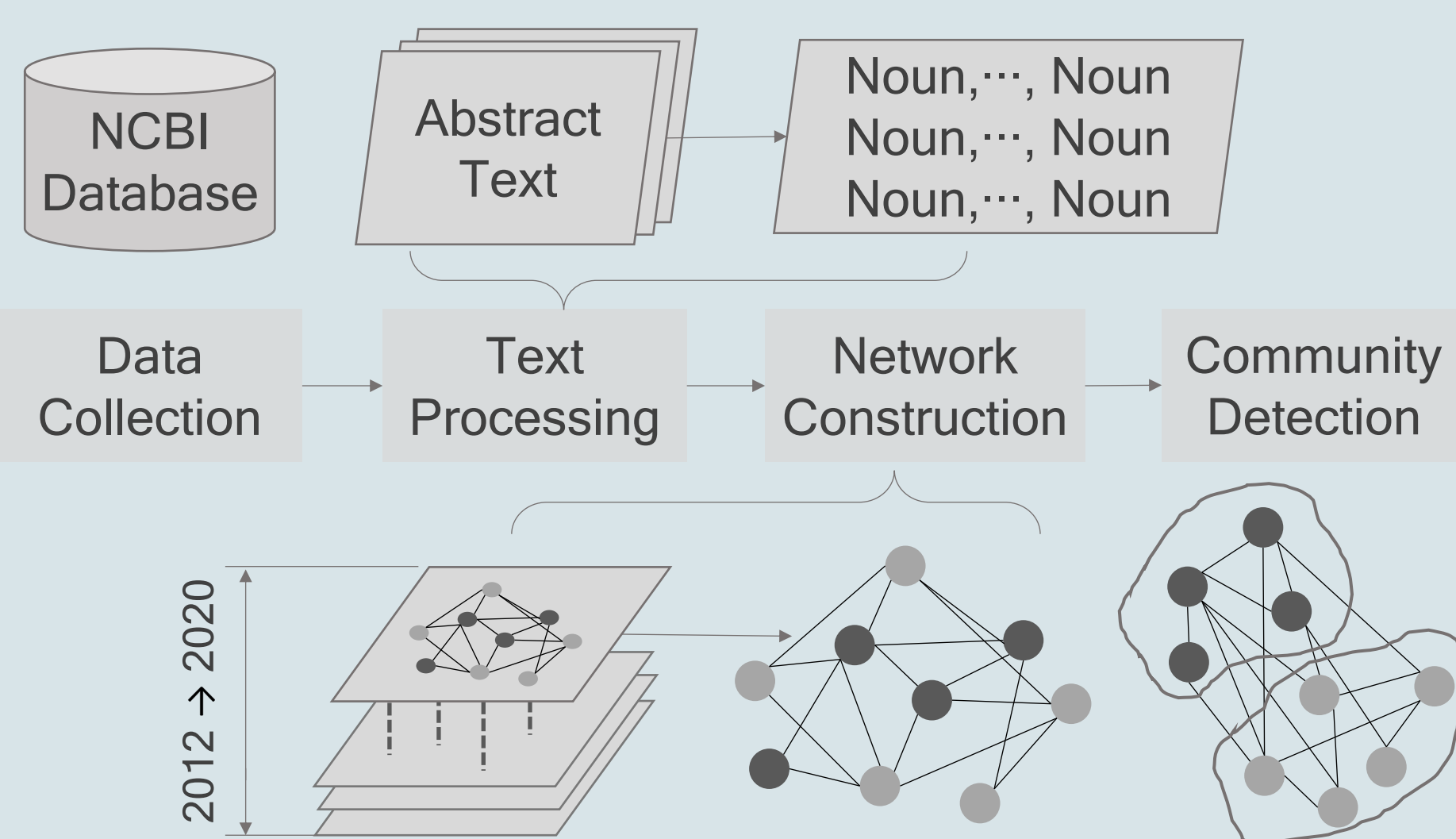
This research examines how the scientific community has evolved its semantic structure about the gene-editing technology CRISPR-Cas9 by analyzing over 23,000 CRISPR-related scientific papers from 2012 through 2020 with a text network approach.

Introduction

In 2012, two researchers reported their findings that a genetic engineering technique could be programmed with RNA to edit genomic DNA. After the discovery, the technique CRISPR-Cas9 has been gaining widespread adaptation in biology but also causing much scientific and ethical uncertainty including human germline modification. Although much public controversy exists over the potential benefits and perils, few studies have explored how the scientific community has been responding to this revolutionary technology.

Here we apply a text network approach to CRISPR-related scientific papers and examine the network dynamic of scholarly communication.

Methodology



With the noun terms from CRISPR-related scientific papers, we generated a multi-layer network based on the adjacency matrix weighted by the TF-IDF value for overlapping terms. Then we applied a generalized Louvain method to detect cohesive clusters.

Conclusions

Using the text network approach, we find that a stable semantic structure has emerged from scholarly communication over years, and a characteristic move of ethically sensitive words along with network dynamics.

A powerful technology such as CRISPR-Cas9 can challenge social norms and ethics. This study shows potential of network approach for understanding the macro dynamics of scholarly communication led by the technology revolution.

Figure 1. Text networks of the CRISPR-related papers. The top figure shows the overall semantical structure of nine years. To examine the clearly cut topology, we filtered the networks to conserve significant edges. The semantically coherent clusters are assigned with the same colors.

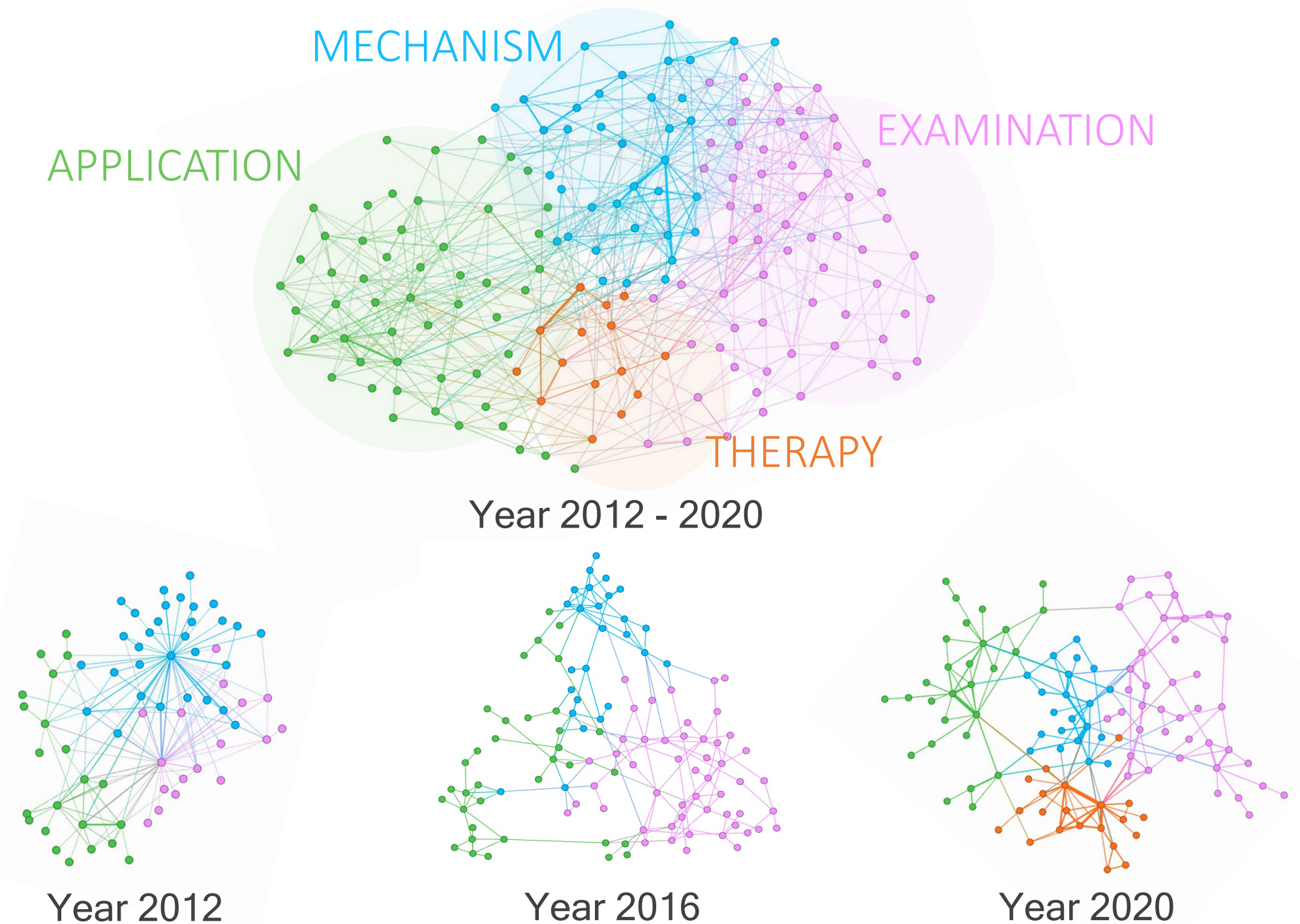
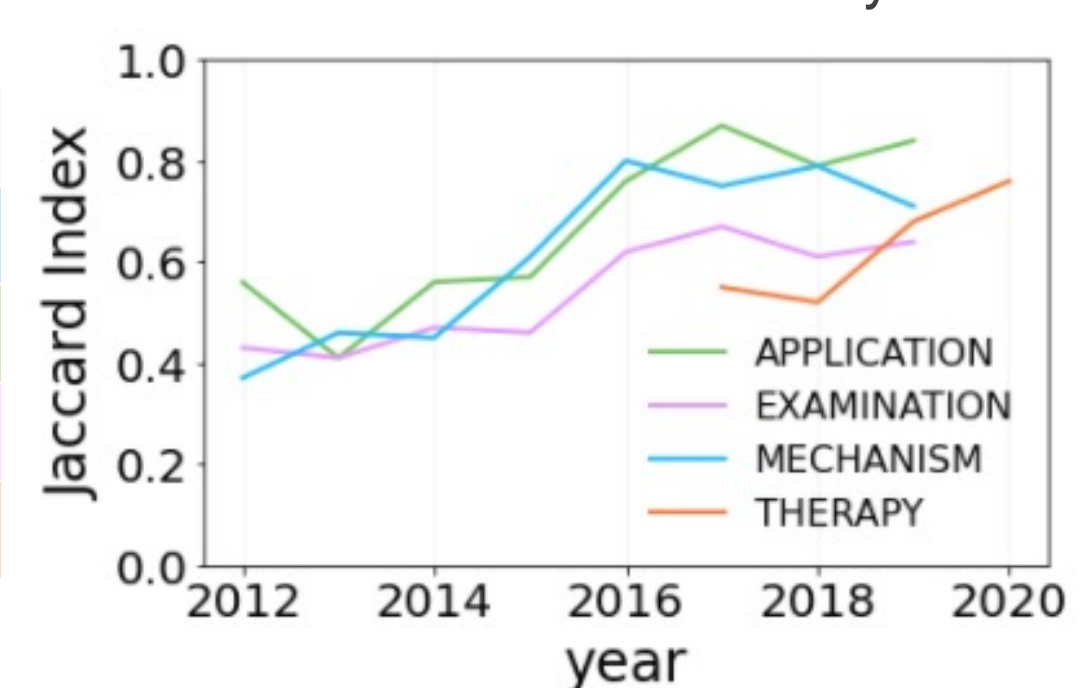


Table 1. Topics and Key terms of each cluster with corresponding color

TOPIC	Key terms
MECHANISM	DNA, bacteria, locus
APPLICATION	technology, system, review
EXAMINATION	expression, study, protein
THERAPY	human, immune cell

Figure 2. Jaccard similar coefficient of terms within each cluster over years



Results

The text network rapidly grows after the technology discovery. This suggests that scientists use more various words in their publications of CRISPR along with the scientific development.

The semantic structure also shifts over years. As the research field becomes established, the semantic structure also becomes stable. The increasing size of “application” cluster denotes more discussion on the advantages and challenges about this technology. The cohesive cluster of “therapy” has emerged and expanded after year 2017, suggesting the focus shift from biological understanding to therapeutic applications.

Our text network analysis captures the shift of ethically sensitive words in the scholarly communication. For example, the move of terms “human” and “embryo” around position and clusters suggests the changing context of researchers’ discussion on the ethically sensitive topic.