

Network Evolution of Scholarly Communication on CRISPR Gene Editing

Keywords: *CRISPR, text network analysis, semantic structure, network evolution, the science of science*

Extended Abstract

In 2012, two researchers reported their findings that a genetic engineering technique could be programmed with RNA to edit genomic DNA. Since the discovery, the technique CRISPR-Cas9 has been gaining widespread adaptation in biology, given its promise for a wide range of medical and commercial applications. Because of its high application potentiality, the technology has also caused much scientific and ethical uncertainty including human germline modification [1]. Although much public controversy exists over the potential benefits and perils, few studies have explored how the scientific community who engages in the development has been responding to this revolutionary technology. Here we apply a text network approach [2] to CRISPR-related scientific papers and examine the network dynamics of scholarly communication on the gene editing technology.

Text networks represent relationships between words, as social networks represent relationships between people [3]. To apply the network-based approach, we collected 23,187 scientific articles using the query term “CRISPR” from year 2012 through year 2020 from the National Center for Biotechnology Information (NCBI) database. We only used the abstract texts as corpus to capture the articles’ basics.

After text preprocessing, in keeping with prior work [2], we extracted the 1,000 noun terms of the most frequent occurrence in the corpus (comprising of 192,847 terms) as the nodes of text networks. To examine the network evolution, we separately generated the network of each year based on the paper’s publication year, i.e., nine networks in total from 2012 through 2020. For each network, we computed the sum of the term frequency-inverse document frequency (TF-IDF) for the overlapping terms in a paper’s abstract as edge weight. We then applied the Louvain community detection algorithm to the text networks to identify cohesive clusters in the semantic space.

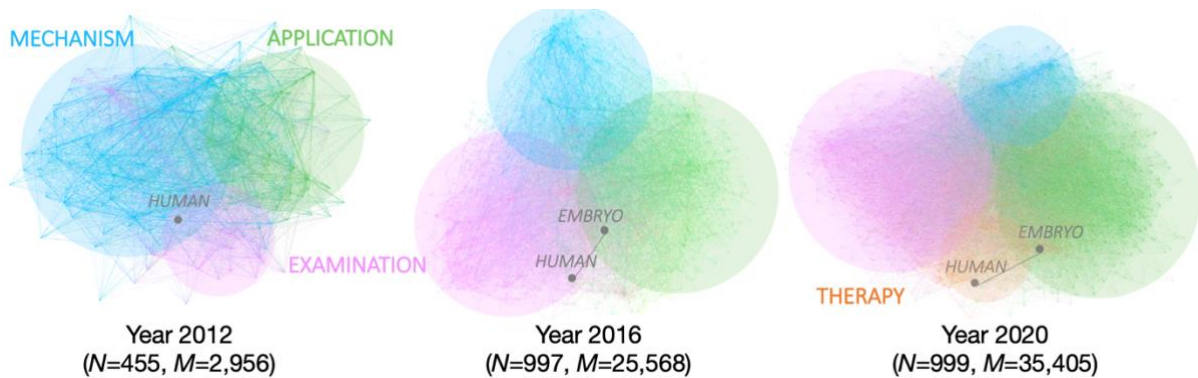


Figure 1. Text networks of CRISPR papers published in 2012, 2016, and 2020.

Figure 1 shows the text networks of year 2012, 2016, and 2020. To visualize and examine the clearly cut topology, we filtered the networks to conserve edges above 0.01 of edge weight. We interpreted each coherent cluster with high degree terms and identified the

topic. We labeled the cluster of the terms such as “DNA,” “bacteria,” and “locus” as “mechanism” (colored with blue in Fig. 1); the one of “gene expression,” “study,” and “protein” as “examination” (pink); the one of “technology,” “system,” and “review” as “application” (green); and the one of “human” and “immune cell” as “therapy” (orange). In Fig. 1, the semantically coherent clusters are assigned with the same colors while the node position is not fixed across years.

We find that the text network rapidly grows after the technology discovery (Fig. 1). This suggests that scientists use more various words in their publication about CRISPR along with the scientific development. The semantic structure also shifts over years. In the text network of year 2012, three semantically cohesive clusters are detected, and the “mechanism” cluster is the largest. As the research field becomes established, the semantic structure also becomes stable. The “examination” and “application” clusters are placed across the network and the “mechanism” cluster are between them with decreasing cluster size. The increasing size of the “application” cluster denotes that more general reviews have been published to discuss the advantages, current challenges, and concerns about the gene-editing technology. We also find another cohesive cluster of “therapy” has emerged and expanded after year 2017. The findings suggest that researchers have changed the focus more from biological understanding to therapeutic applications such as gene therapy over years.

Text network analysis also allows us to investigate the orientation of specific terms in the evolution of semantic structure. Since the discovery of CRISPR-Cas9, technology-related ethical issues have been discussed in the context of possible applications. Our text network analysis captures the shift of ethically sensitive words in the scholarly communication. For example, the terms “human” and “embryo,” highlighted with large nodes in Figure 1, move around the positions and clusters in the text network, suggesting the changing context of researchers’ discussion on the ethically sensitive topic. In 2012, the term “human” is located at the periphery of the “mechanism” cluster. After moving closer to the cluster of “therapy”, it takes a position bridging between clusters with 36% neighbors from the “examination” cluster and 41% neighbors from the “application” cluster as of 2020. In addition, the term “embryo” appears in 2013 and connects with the term “human” in the text network. In the text network of year 2020, more triangles are observed with the “human”-“embryo” dyad and the terms in the “application” cluster such as the term “clinical application,” indicating more discussions on, for example, possible applications on human embryo.

We examine how the science community has evolved the semantic structure of scholarly communication about CRISPR-Cas9. Using the text network approach, we find that a stable semantic structure has emerged from scholarly communication over eight years. We also find a characteristic move of the ethically sensitive words along with the text network’s dynamics. A powerful technology such as CRISPR-Cas9 can challenge social norms and ethics [1]. We cannot discuss such a challenge without a thorough understanding of its historical context and social inertia. This study shows potential of network approach for understanding the macro dynamics of scholarly communication led by the technology revolution.

References

- [1] Davis, Kevin. *Editing Humanity*. Pegasus Books (2020).
- [2] Rule, Alix, Jean-Philippe Cointet, and Peter S. Bearman. "Lexical shifts, substantive changes, and continuity in State of the Union discourse, 1790–2014." *PNAS* 112.35 (2015): 10837-10844.
- [3] Bail, Christopher A. "Combining natural language processing and network analysis to examine how advocacy organizations stimulate conversation on social media." *PNAS* 113.42 (2016): 11823-11828.