The Collaborative Structure of Synthetic Biology Ethics

Brandon Sepulvado, sepulvado-brandon@norc.org, Department of Methodology and Quantitative Social Sciences, NORC at the University of Chicago
Jacob Jett, jjett2@illinois.edu, School of Information Sciences, University of Illinois at Urbana-Champaign
J. Stephen Downie, jdownie@illinois.edu, School of Information Sciences, University of Illinois at Urbana-Champaign

Abstract
Although synthetic biology is now an established field, ethical reflection surrounding it has only recently become a subject of sustained scientific inquiry. This poster displays the results from a series of network analyses examining how synthetic biology ethics is evolving. The studies gather ethics articles from 802 institutions and 2,179 authors and suggest that the fields' social structure has become democratized at the individual level but remains dominated by a handful of institutions at the organizational level.

Introduction
Since the emergence of synthetic biology in the early 20th century (Jacob & Monod 1961; Ludec 1912; Szybalski & Skalka 1978), the ethical issues posed by the field did not receive much attention outside bioethics. However, with the advent of large-scale gene analytics in the 1990s and the production of large genomics datasets, synthetic biology grew rapidly in the 2000s. By the late 2000s, synthetic biology ethics began to garner sustained attention from the scientific community (Sepulvado et al. 2020a, 2020b), yet most scientometric investigations of synthetic biology (Hu & Rousseau, 2008; Raimbault et al., 2016; Shapira et al., 2017) neglect a focused study of ethics. As such, we ask, what is the structure of collaborative activities (e.g., co-authorship) that sustains synthetic biology ethics?

This research takes place as part of a broader set of work focused on developing a repository and information retrieval system. When completed, the system will provide a single resource that permits synthetic biologists to model biological circuits (genetic pathways in biological organisms responsible for regulating cellular processes such as hormone manufacture, etc.), reference pertinent domain articles, and create new knowledge. One important goal is to also integrate the existing corpus of synthetic-biology-focused ethics literature, so that synthetic biologists can easily reference ethics issues pertinent to their work. A key initial task is to examine the field's social structure.

Data and Methods
In order to investigate the social structure of synthetic biology ethics, we consulted the Web of Science. We chose the Web of Science to maintain continuity with past bibliometric investigations of synthetic biology (e.g., Raimbault et al. 2016; Shapira et al., 2017) and because the Web of Science contains a good balance of humanities, which is important for ethics, and science (Falagas et al., 2008). Scientometricians have devoted considerable attention to keyword-based queries for synthetic biology literature (Hu & Rousseau, 2008; Raimbault et al., 2016; Shapira et al., 2017), and we used Shapira et al.'s (2017, p. 1450) proposed query—which we verified with practicing synthetic biologists—to generate the dataset used in the proceeding analyses. We retained only those records mentioning ethics, safety, dilemma, or security. This query resulted in a corpus of 15,172 articles, of which 572 specifically related to ethics.

83rd Annual Meeting of the Association for Information Science & Technology | October 24-28, 2020.

We construct two co-authorship networks: one in which individuals are connected when they co-author texts and one in which institutions are connected when individuals from each co-author texts together. We present below descriptive network analyses. Given the networks' sizes, we present them as heatmaps to more easily convey underlying structures. Individuals and institutions that are located close to each other are either directly connected through co-authorship or tend to be closely connected (e.g., their collaborators have collaborated); authors and institutions that are distant from one another are at most only very indirectly connected.

Results

Figure 1 presents a heatmap of the individual-level co-authorship network, which contains 2,179 authors and 7,297 collaborations, with red indicating frequent collaborations and blue indicating areas with few to no collaborations. Three to four clusters contain individuals who co-author very frequently. Outside those, one sees a handful of small orange clusters with periodic co-authorship. The remaining individuals tend to co-author texts very infrequently and with only a minimal number of others. Further analyses (not shown) suggest that authors of synthetic biology ethics texts tend to produce a relatively even number of texts, and citations among authors tend to be distributed evenly, as well.

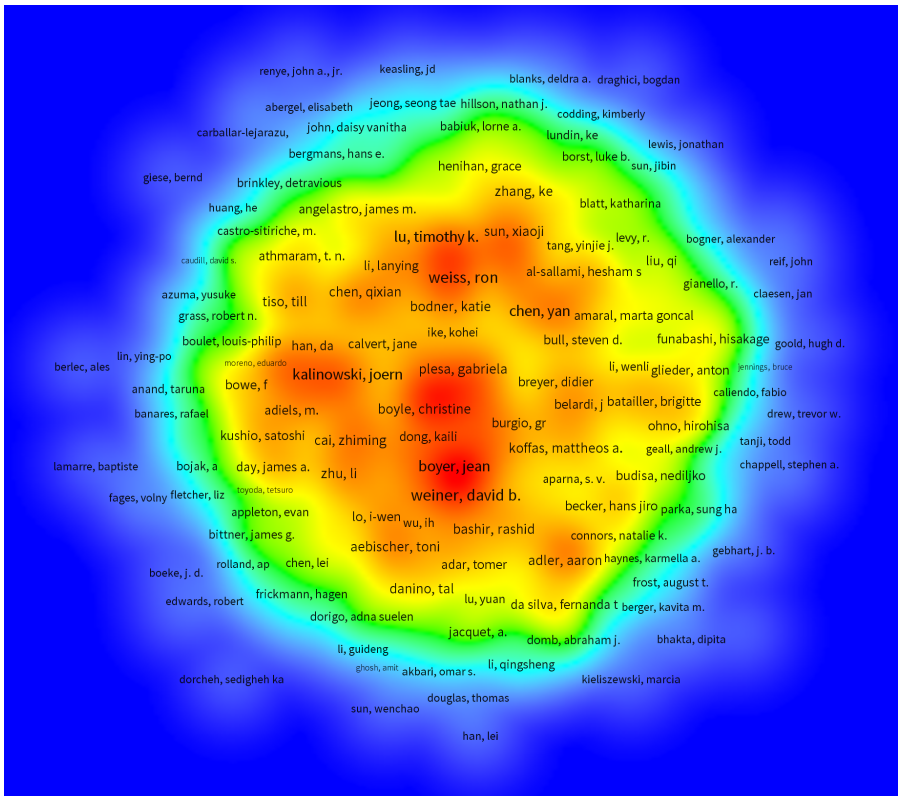

Figure 1: Heatmap of Co-Authorship Network. Shading reflects authors' collaborations.

Figure 2 presents the institution-level co-authorship network, comprising 802 institutions and 1,434 relationships. Red indicates the greatest density of institutional collaborations, whereas blue indicates few to no collaborations. A very different structure emerges at the institutional level. The clusters on the left-hand side suggest frequent collaborations between organizations in that space. The cluster with the densest collaborations tends to be dominated by private U.S. institutions, e.g., Harvard and MIT, and the second

densest cluster is dominated primarily by large public U.S. universities, though it also includes several Japanese institutions. The remaining cluster of the left side is composed of three somewhat distinct groups (i.e., one of primarily Chinese institutions and two of U.S. and European collaborations), and these left-hand groups cohere into one main connected component within the network. The remaining crescent shape suggests small sets of organizations that only infrequently collaborate, with the exception of two moderately connected groups on the right.

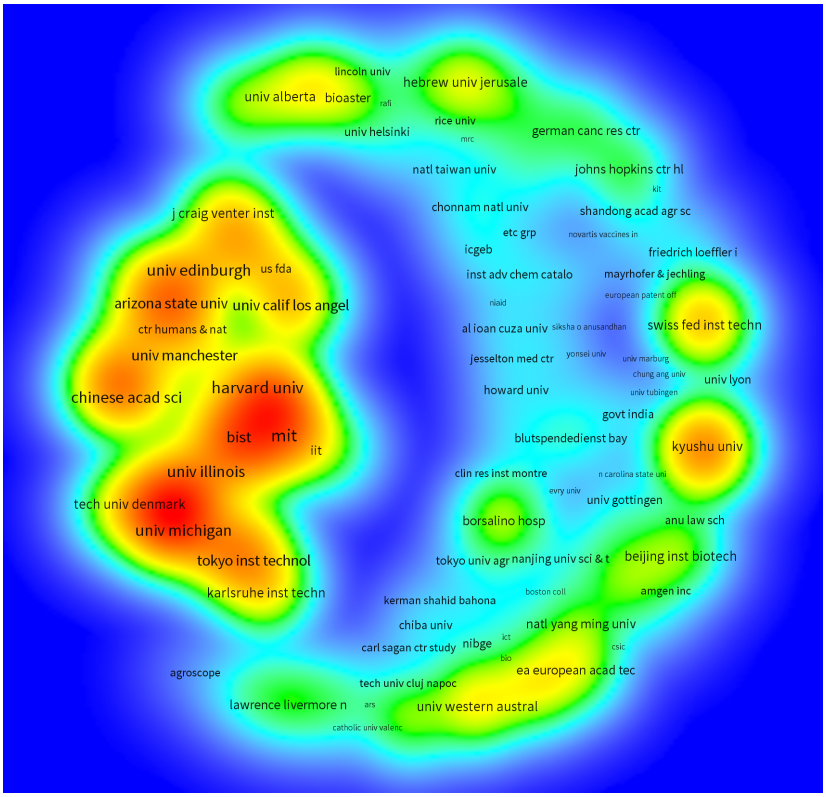

Figure 2: Heatmap of institutional collaborations. Shading reflects the number of links institutions have.

Figure 3 presents the same institutional-level co-authorship network, but the shading now reflects the productivity (i.e., the number of papers written by) of organizations in a given position. In this visualization, shading closer to red suggests higher productivity, and shading closer to blue suggests lower productivity. Institutions in the U.S. private university cluster and in one of the U.S.-Europe collaborative clusters (e.g., with Edinburgh, Manchester, and ASU) produce by far the most texts on synthetic biology ethics. Additional analyses (not shown) indicate that, within this institutional collaboration network, citations are focused in the main clusters and spread relatively evenly among the remaining institutions.

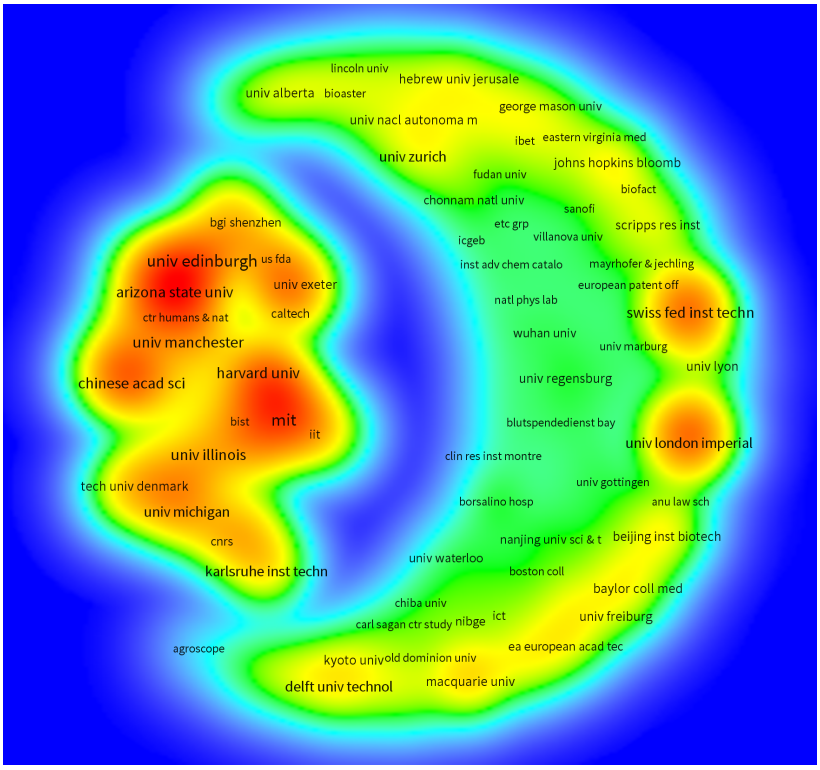

Figure 3: Heatmap of institutional collaborations. Shading reflects institutions' research output.

Conclusion

On the individual level, synthetic biology ethics does not seem to be dominated by a centralized group of scholars, frequently called an invisible college (Crane, 1975; Price, 1965). However, the social structure is still organized around a handful of frequently collaborating communities. This makes sense given that synthetic biology ethics publications started growing at an exponential rate only around 2010 (Sepulvado et al. 2020a, 2020b). On the institutional level, the production of texts, citations, and collaboration patterns suggest that the field is still grounded by a small number of institutions. Although there is no single invisible college in either network, the clusters might indicate emerging schools of thought. This would be consequential for synthetic biology information retrieval systems because they would need to ensure that all positions on a given ethical issue are represented in results returned to users.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Nos. NSF-1939929 and NSF-1939887.

References

Crane, D. (1975). *Invisible colleges: Diffusion of knowledge in scientific communities*. University of Chicago Press.

Falagas, M.E., Pitsouni, E.I., Malietzis, G.A., & Pappas, G. (2008). Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB Journal* 22(2), 338-342.

Hu, X. & Rousseau, R. (2015). From a word to a world: the current situation in the interdisciplinary field of synthetic biology. *PeerJ 3*, e728.

Jacob, F.ß. & Monod, J. (1961). On the regulation of gene activity. *Cold Spring Harbor Symposia on Quantitative Biology 26*, pp 193-211.

Ludec, S. (1912). *La biologie synthétique, étude de biophysique*. Paris: A. Poinat.

Price, D. d. S. (1965). *Little science, big science*. Columbia University Press.

Raimbault, B., Cointet, J.P., & Joly, P.B. (2016). Mapping the emergence of synthetic biology. *PLoS ONE 11*(9), e0161522.

Sepulvado, B., Jett, J. & Downie, J. S. (2020a). The Social and Conceptual Structure of Synthetic Biology and its Ethical Discourse. Paper presented at *Sunbelt 2020 Virtual Conference* (13-17 July 2020).

Sepulvado, B., Jett, J., & Downie, J. S. (2020b). The Social and Conceptual Organization of Synthetic Biology Ethics. Poster presented at *IWDBA 2020: 12th International Workshop on Bio-Design Automation* (Online, 3-7 August 2020).

Shapira, P., Kwon, S., & Youtube, J. (2017). Tracking the emergence of synthetic biology. *Scientometrics 112*, 1439-1469.

Szybalski, W. & Skalka, A. (1978). Nobel prizes and restriction enzymes. *Gene 4*(3), pp 181-2.

