The State of Synthetic Biology Scholarship: A Case Study of Comparative Metrics and Citation Analysis

Jeffrey C. Cegan

Introduction

Synthetic biology is a growing field of study within scholarly literature. It is the deliberate engineering of existing biological systems or the creation of novel biological systems. Advancements in synthetic biology have been driven by key technologies that increase efficiency and reduce resource costs for DNA synthesis and sequencing (Raimbault et al. 2016). Such research generally seeks to empower greater design capacity and control over biological systems and phenotypic expression – the results of which have been posited to offer benefits in fields ranging from medicine to energy to environmental remediation (Khalil and Collins 2010; Church and Regis 2014).

While motivations of synthetic biology were described in the early twentieth century, the modern field has evolved through breakthroughs such as synthetic circuit engineering (Elowitz and Leibler 2000) and the creation of a bacterial cell with a chemically synthesized genome (Gibson et al. 2010). These and other related pieces of scholarship represent the iterative and evolutionary nature of synthetic biology, where scholars build from seminal works in the field to improve cellular design and control (Cameron et al. 2014). Such research was described in existing scholarship such as Raimbault et al. (2016), Cameron et al. (2014), and Oldham et al. (2012), yet these efforts focused upon reviewing the state of biological and computational sciences relevant to synthetic biology and generally did not describe narratives and developments within social sciences literature (Trump 2017). A more thorough and holistic quantification of synthetic biology literature will help visualize and describe the growth and development of the larger synthetic biology

J. C. Cegan (✉)
Environmental Laboratory, Engineer Research and Development Center,
US Army Corps of Engineers, Washington, DC, USA
e-mail: Jeffrey.C.Cegan@usace.army.mil

B. D. Trump et al. (eds.), *Synthetic Biology 2020: Frontiers in Risk Analysis and Governance*, Risk, Systems and Decisions, https://doi.org/10.1007/978-3-030-27264-7_4

community over time and will respond to foundational or seminal discussions noted within key publications (Trump et al. 2019).

With this inspiration, this chapter constructs and analyzes a citation network of synthetic biology scholarship in both technical and social sciences. The initial search of articles is described in Trump et al. (2019). Described herein as "communities of practice," synthetic biology literature is categorized into various paths of inquiry including (i) "state-of-science" literature detailing biological experimentation or computational models, (ii) "products" literature detailing the application of synthetic biology research to a potential product application, (iii) "risk" detailing literature that characterizes synthetic biology risks or provides metrics for a potential risk assessment, (iv) "governance" including literature that describes regulatory and governance needs, and (v) "ELSI" literature pertaining to ethical, legal, and moral implications that synthetic biology may incur. Using these five communities of practice, this chapter explored how a compilation of 712 publications related to synthetic biology acknowledge and discuss developments within and between their respective communities. We used network science to identify how articles within each community of practice cite one another. The connections among these communities of practice are quantitatively assessed through the construction of a citation network. Overall, such a citation network provides a holistic quantitative measure of the importance of different communities of practice within synthetic biology and identifies key performers with the greatest degree of influence upon the larger synthetic biology community. This chapter utilized multiple quantitative metrics to identify those publications with the greatest amount of network centrality or those that displayed a high degree of motivational influence upon synthetic biology's communities of practice from 2000 to 2017. Through such quantification, it may be possible to identify which publications and communities of practice have spurred the evolutionary development of synthetic biology's enabling technologies in particular and the larger discussion of its use and impacts in general.

The synthetic biology citation data were used to create a network, which is a graph that contains a number of nodes, a.k.a. vertices, connected by links, a.k.a. edges (Jacob et al. 2017). Each publication is represented by a vertex, and each citation is shown as an arrow into or out of the vertex, depending on whether the publication is being cited or citing another publication, i.e., the citation network is directed. This network is also disconnected because all nodes cannot be reached by all other nodes.

Developed by Euler in 1736, network science studies pairwise relations within a network. McPherson et al. (2001) introduced the concept of homophily, which suggests that similarity breeds connection in a network. Synthetic biology is an evolving field in which much of the discussion and research is steered by a few communities of practice. Their research is then picked up and recreated or furthered by universities, NGOs, and other labs on the periphery. Since the field is primarily driven by those few communities, the network analysis explored whether or not similar types of publications are more likely to cite within their own groupings.

Applying graph theory to citation networks was pioneered by Garfield (1955) and De Solla Price (1965) and has since been applied to several fields to analyze

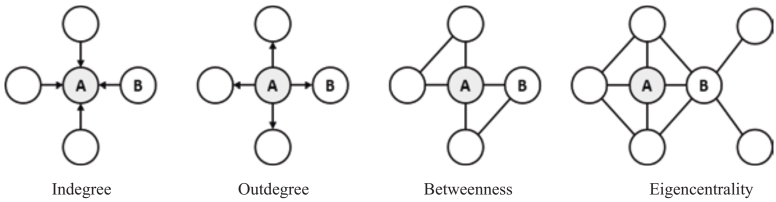

Fig. 1 Network centrality measures

networks and attempt to quantify the importance of a publication. Citation networks represent publications and their citations using nodes that are connected through links, or edges. Centrality measures are one method to quantify node analysis and seek the node(s) most central, or critical, to a network. The simplest measure is degree centrality, which is found by counting the edges connected to each node. Edges are further divided directionally, into and out of each vertex, called indegree and outdegree, respectively (Weisstein 2017). Betweenness centrality measures the extent that a vertex is positioned on the shortest path between other pairs of vertices in the same network (Leydesdorff 2007). Eigenvector centrality, or eigencentrality, uses the eigenvector of the largest eigenvalue and standardizes to the length of the eigenvalue (Bonacich 1972), giving greater weight to vertices connected to other highly connected vertices. In each of the networks shown in Fig. 1, node A has a greater centrality than node B according to the particular measure referenced. Such quantitative measures offer objective viewpoints regarding which publications might be driving discussion within larger synthetic biology literature and ultimately detail how the field has grown over time to incorporate various elements of biological and social sciences discussion.

Methodology

To populate the synthetic biology citation network, a topic search was conducted through the ISI Web of Knowledge. Specifically, 16 search terms were used to generate a curated list of 880 publications. Relevant criteria for inclusion into our citation network required that each publication (i) primarily focus upon the implications or research of synthetic biology in general or an enabling technology in particular, (ii) were indexed in a journal, book, or formal government document about synthetic biology, and (iii) were framed between the years 2000 and 2017 (see Cameron et al. 2014 for a brief description of biological developments within synthetic biology research). Each publication is classified into one of five “communities of practice,” including: (1) state of science, (2) products, (3) risk, (4) governance, or (5) ethical, legal, and social implications (ELSI).

State of science included publications that described the evolutionary or iterative advancement of a synthetic biology enabling technology within a laboratory setting or via computational modeling. Product publications focused upon applying synthetic biology scientific development into a commercial product. Risk publications sought to characterize or assess risks associated with synthetic biology research and development and governance publications identified and described strategies for the regulation and governance of processes or products of synthetic biology (Linkov et al. 2018). Finally, ELSI included publications examining social impact considerations (Merad and Trump 2019), legal challenges (Marchant et al. 2013), ethical concerns, and general economic impact discussion resulting from synthetic biology development and commercialization. Table 1 shows the number of publications in each community of practice.

Communities of practice were assigned manually to each publication consistent with Trump et al. (2019). Each publication is assigned to only one community of practice, meaning that the community of practice groupings are mutually exclusive. For the case where a publication is related to multiple community of practice areas, the author chose the most dominant community of practice.

Of the 880 returns, 168 were not linked to any other publication in the network for either having no references to other review publications or their files were not readable for text mining. This left a remaining 712 publications for analysis. A brief discussion of the removed isolated publications can be found in Appendix A. Publication dates range from 2000 to 2017, with every year represented in the ISI Web of Knowledge search. A histogram of publication years for the 712 publications in the network is shown in Fig. 2.

Another important methodological issue involves the use of text mining, an approach for extracting information from a body of text. The "tm" package in R was

Table 1 Number of publications in each community of practice

	State of science	Products	Risk	Governance	ELSI
Publications	445	122	33	42	70

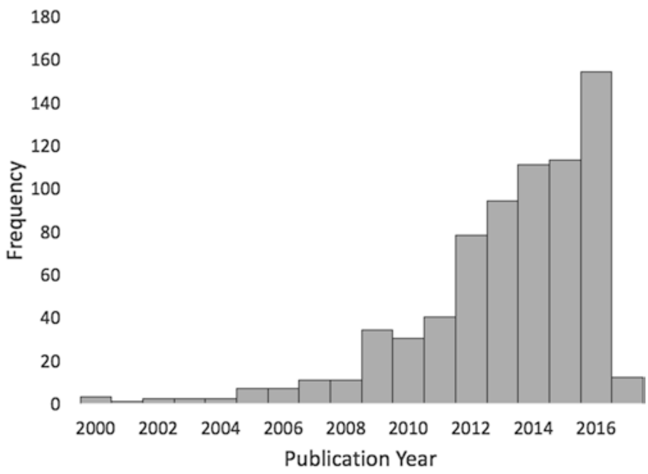

Fig. 2 Histogram of publication years

utilized to scan and identify patterns and frequencies of symbols within the PDF format of each publication. Using publication meta-data about the lead author's last name, publication year, and title information, we determined whether or not each publication was being referenced within another publication. This process produced an adjacency matrix of citations among all reviewed publications, and this quantitative data was converted into a citation network.

Results and Discussion

This section is divided into three parts. First, we show a graph of the network and cross pollination by community of practice. Second, we show top publications for each centrality and discuss pairwise correlations between measures. Finally, we examine the temporal changes in the network.

Network by Communities of Practice

In order to visualize the network, the graph shown in Fig. 3 was created to represent each of the 712 publications as a node, colored according to its respective community of practices. Links connecting one node to another are colored according to the community of practice of the terminal node, i.e., the publication being cited.

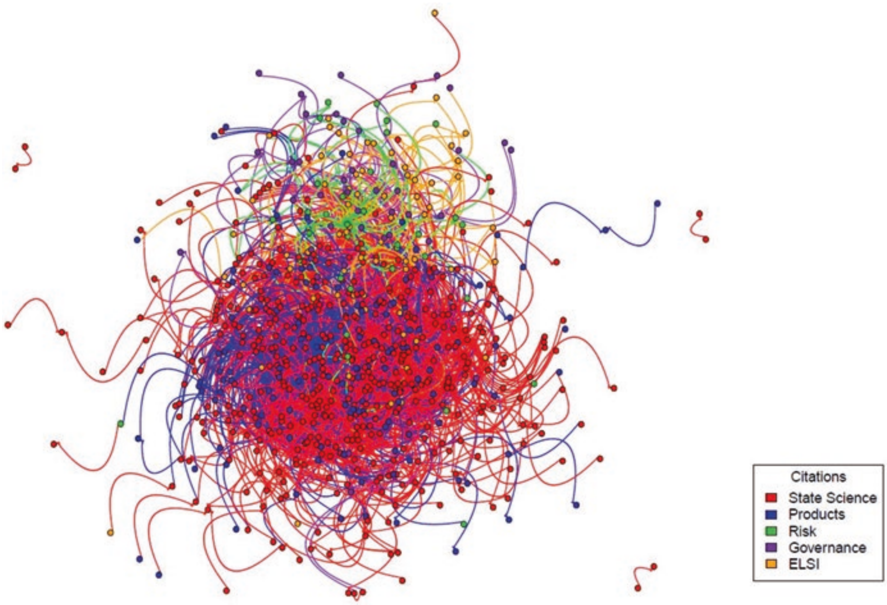

Fig. 3 Citation network of select synthetic biology literature (2000–2017)

Table 2 Cross pollination (citation counts)

	State of science	Products	Risk	Governance	ELSI	Total
State of science	1322	291	10	14	39	1676
Products	343	175	6	10	11	545
Risk	64	15	16	15	19	129
Governance	51	20	18	32	19	140
ELSI	113	39	10	19	72	253

Table 3 Cross pollination (citation percentages)

	State of science (%)	Products (%)	Risk (%)	Governance (%)	ELSI (%)	Total (%)
State of science	79	17	1	1	2	100
Products	63	32	1	2	2	100
Risk	50	12	12	12	15	100
Governance	36	14	13	23	14	100
ELSI	45	15	4	8	28	100

We analyze the interrelation of citations among communities of practice in Tables 2 and 3. The "Total" column sums to 2743, which represents the total number of links in the network.

The rows in Table 2 enumerate the citations that each community of practice made to each community column: of the 1676 network references identified in "state-of-science" publications, 1322 referenced other "state-of-science" publications, 291 referenced "product" publications, 10 referenced "risk" publications, and so on.

Table 3 shows that "state of science" is the largest share of cited publications, regardless of the publication category. This means that there is relatively little homophily among these categorical groupings because individual communities are not citing primarily within their own group. "State of science" seems to be at the forefront of this emerging field, and other synthetic biology communities are seemingly looking to "state-of-science" publications for guidance in their own field.

Network Centrality Measures

Measures of network centrality indicate the more important nodes and provide a greater understanding to key network structures. In this section, centrality concepts of degree, betweenness, and eigencentrality are used to understand the structure of the synthetic biology literature network.

Table 4 Sorted by decreasing indegree

Publication	Indegree	Outdegree	Betweenness	Eigen
Elowitz and Leibler (2000)	142	0	0	1
Gardner et al. (2000)	123	0	0	0.92766808
Endy (2005)	113	2	187.362	0.72251627
Ro et al. (2006)	95	0	0	0.47347666
Gibson et al. (2010)	91	0	0	0.38069174
Khalil and Collins (2010)	71	10	1657.26	0.63949588

Indegree

In a directed network, the indegree of a node is the number of links leading to that node. For publication citations, this means how often a publication is cited by other publications in the network. Indegree citations are the most common measure of a paper's relevance to the field, i.e., how many times it has been cited. The top six publications according to the indegree centrality measure are shown in Table 4. Table 4 also shows the profile publications' scores for the remaining three measures (outdegree, betweenness, and Eigen), which will be described in more detail in the following text.

Indicating the top-cited publications by indegree, several publications noted in Table 4 are described as seminal synthetic biology scholarship by Cameron et al. (2014). For example, Elowitz and Leibler (2000) and Gardner et al. (2000) published work describing the first synthetic circuits that were engineered to perform logical functions and increase control within and between cells. Further, Endy (2005) outlined the general principles of engineering biology, which served as a significant philosophical and scientific underpinning behind synthetic biology research moving forward. Finally, Gibson et al. (2010) described the creation of the first cell with a "synthetic genome," where synthesized DNA cassettes were assembled to recreate *M. mycoides* and inserted into a cell that would contain only the synthesized genome. While many other important scientific papers may be described here, this output signifies that key motivational or scientifically foundational works are among the top citation performers within synthetic biology scholarship.

Further, the citation network reveals that four of the six most cited publications do not cite any network publications themselves. For the top two publications (Elowitz and Leibler 2000; Gardner et al. 2000), this is due to the fact that they were both published in 2000 (the beginning year of the Web of Science search queries).

Outdegree

The outdegree of a node is the number of links leading away from that node. For publication citations, it shows how often a publication cites other publications in the network. Publications with high outdegree measures have a strong awareness of

Table 5 Sorted by decreasing outdegree

Publication	Indegree	Outdegree	Betweenness	Eigencentrality
MacDonald and Deans (2016)	0	31	0	0.36462184
Trosset and Carbonell (2015)	1	31	113.983	0.16192745
Brophy and Voigt (2014)	32	24	1669.1	0.43277875
Cameron et al. (2014)	16	21	493.538	0.40452803
Bradley et al. (2016)	1	21	66.2056	0.22901943
Xie and Fussenegger (2015)	1	18	26.7721	0.15101327

other publications in the field. The top six publications according to the outdegree centrality measure are shown in Table 5.

Each of the top six publications by outdegree was published in 2014 or later. Such articles reflect heavily upon engineering principles via enabling technologies of synthetic biology (Brophy and Voigt 2014; Bradley et al. 2016), applications of such enabling technologies to potential products (Trosset and Carbonell 2015; Xie and Fussenegger 2015), or review general progress within computational and engineering biology (MacDonald and Deans 2016; Cameron et al. 2014). Specifically, Brophy and Voigt (2014) describe principles of genetic circuit engineering, which Bradley et al. (2016) expand upon for microbial gene circuit engineering. Next, Trosset and Carbonell (2015) discuss the use of synthetic biology for pharmaceutical drug discovery, and Xie and Fussenegger (2015) discuss engineering principles for mammalian designer cells in biomedical applications. Finally, Cameron et al. (2014) offers a timeline of synthetic biology biological and computation sciences through 2014, while MacDonald and Deans (2016) further unpack the various tools, applications, and enabling technologies of synthetic biology. In general, these papers served as reviews or aggregations of synthetic biology research and heavily reference developments within the larger synthetic biology world.

Figure 2 shows that these publications were published after the majority of the network and therefore also had more literature to potentially cite. Four of these publications are cited one or fewer times by other network publications, likely due to the fact that all four were published after 2014 and have had little time to be cited. Of note, Brophy and Voigt (2014) is cited close to twice as often as the other five combined, suggesting that it both had a broad understanding of the field and quickly became an important publication.

Betweenness

Betweenness centrality measures the extent that a vertex is positioned on the shortest path between other pairs of vertices in the same network (Leydesdorff 2007). For publications within the network, a high betweenness measure indicates that the publication is playing a critical role to link publications to one another. In general, publications with high betweenness can be viewed as more interdisciplinary. The top six publications according to the betweenness centrality measure are shown in Table 6.

Table 6 Sorted by decreasing betweenness

Publication	Indegree	Outdegree	Betweenness	Eigencentrality
Brophy and Voigt (2014)	32	24	1669.1	0.43277875
Khalil and Collins (2010)	71	10	1657.26	0.63949588
Carr and Church (2009)	59	5	1437.99	0.36316962
Lu and Collins (2007)	30	9	1137.87	0.45722903
Mutalik et al. (2013)	31	7	1031.59	0.3009861
Kahl and Endy (2013)	9	7	902.154	0.12177155

Each publication in Table 6 was published between 2009 and 2014 and has non-zero indegree and outdegree centrality. This list excludes the majority of the publications from Tables 4 and 5. Brophy and Voigt (2014) repeats as a top performer here, which may derive from its review of circuit engineering techniques that are foundational to synthetic biology research. Similarly, Carr and Church (2009) discuss principles of genome engineering, while Kahl and Endy (2013) surveyed a general body of synthetic biology enabling technologies to report on the state of biological and computational research. Khalil and Collins (2010) more generally describe ongoing and potential future developments of both synthetic biology science and potential product applications. Mutalik et al. (2013) sought to address ongoing difficulties with synthetic biology research to reliably model and predict quantitative behaviors for novel genetic combinations and posed an approach to improve modeling and prediction capabilities.

High betweenness nodes have the potential to disconnect networks if they are removed. A publication with high betweenness and relatively low eigencentrality (e.g., Kahl and Endy 2013) may be an important gatekeeper between key clusters of publications. Kahl and Endy's survey of enabling technology serves as a link between the clusters surrounding the various enabling technologies of synthetic biology. Since it is a recent publication, it is not connected to many other highly connected publications. Because it brings together several technological clusters, many shortest paths pass through it and it would disrupt the network if removed, making it an important gatekeeper between clusters. The combination of many shortest paths passing through and a lack of connection to other highly connected publications give this survey a high betweenness centrality and a relatively low eigencentrality.

Eigencentrality

Eigencentrality measures the influence of a node in a network, giving greater weight to a node with more connections to other highly connected nodes. For publications within the network, a high eigencentrality measure indicates that the publication is playing an influential role to link important publications to one another while also being important itself. The top six publications according to the eigencentrality measure are shown in Table 7.

Table 7 Sorted by decreasing eigencentrality

Publication	Indegree	Outdegree	Betweenness	Eigencentrality
Elowitz and Leibler (2000)	142	0	0	1
Gardner et al. (2000)	123	0	0	0.92766808
Endy (2005)	113	2	187.362	0.72251627
Khalil and Collins (2010)	71	10	1657.26	0.63949588
Basu et al. (2005)	67	0	0	0.51157712
Ro et al. (2006)	95	0	0	0.47347666

The publications in Table 7 have a much greater range of publication years and centrality measures than the previous three tables. Four of the top six cite zero other publications in the network and have a betweenness centrality of zero, indicating that they are terminal nodes in the network. The other two, Endy 2005 and Khalil and Collins 2010, have a relatively low and high betweenness, respectively.

Within this range of papers, Elowitz and Leibler (2000) and Gardner et al. (2000) engineered the first synthetic genetic circuits that carried out specific design functions, which were described by Cameron et al. (2014) as foundational to synthetic biology research. Further, Endy (2005) describes engineering principles for biological systems, which are described by Raimbault et al. (2016) as seminal to synthetic biology's philosophical and scientific development. On a different note, Ro et al. (2006) describe how engineered yeast can be used to generate semi-synthetic artemisinic acid – an antimalarial precursor that was frequently discussed in future literature as a potential product application of synthetic biology in developing pharmaceutical applications (see also Paddon et al. 2013 for an update on this work).

Of the publications, five show up in the top six of multiple centrality measures. Elowitz and Leibler (2000), Gardner et al. (2000), Endy (2005), Khalil and Collins (2010), and Brophy and Voigt (2014) show up multiple times. In particular, Khalil and Collins (2010) shows up three times. These publications describe scientific research or perspectives on synthetic biology (Cameron et al. 2014) or have been further described as seminal or important publications in the field (Raimbault et al. 2016).

Pairwise Correlations

Table 8 shows the correlations between the centrality measures to determine how closely they are interrelated. This table compares how publications' different centrality measures generally relate to one another. Indegree and outdegree are not closely related, i.e., publications that are highly cited do not cite others in the network and publications that are cited frequently do not tend to cite others in the network. Furthermore, indegree is highly correlated with eigencentrality, demonstrating that publications which are cited the most are also the ones that are connected to other highly cited publications. This indicates high-density groupings around these publications.

Table 8 Pairwise correlations

	Indegree	Outdegree	Betweenness	Eigencentrality
Indegree	1.000	−0.025	0.392	0.774
Outdegree	−0.025	1.000	0.345	0.468
Betweenness	0.392	0.345	1.000	0.485
Eigencentrality	0.774	0.468	0.485	1.000

Evolution of the Network

In this section, we track the development of modern synthetic biology, where Cameron et al. (2014) defines 2000 as a critical starting point due to the first publications on synthetic circuit engineering (Elowitz and Leibler 2000; Gardner et al. 2000). We choose 34-year demarcations to view the growth in citation connectedness within the field and further allow observers to view growth during and immediately after important moments in modern synthetic biology history. These include the first synthetic circuit and toggle switch (Elowitz and Leibler 2000; Gardner et al. 2000), the first meeting of the Biobricks Conference Series in 2004, the description of an engineered bacteriophage for biofilm disposal, the creation of the first bacterial cell with an artificial genome (Gibson et al. 2010), and the commercial production of semi-synthetic artemisinic acid (Paddon et al. 2013). Data from each year includes all publications through that year (i.e., the data from 2004 includes publications from 2000, 2001, 2002, 2003, and 2004). Figure 4 shows what the network looked like in 2000, 2004, 2007, 2010, 2013, and 2016.

In 2000, the citation network contained only three disconnected nodes (Elowitz and Leibler 2000; Gardner et al. 2000; Ostergaard et al. 2000). Between 2007 and 2016, the number of publications in the network grows from 31 to 699 nodes, and the number of links grows from 75 to 2697. The number of links per node grows from just over two to almost four, showing that the network gets more connected as time goes on. The number of nodes in the center continues to increase, while the number of nodes on the outside remains relatively constant. The network becoming more connected suggests that synthetic biology is a unified field expanding in all directions.

As the network grows, its most central publications change over time. The top three publications for each year according to indegree, outdegree, betweenness, and eigencentrality are presented in the following figures. The evolution of top publications based on indegree centrality is shown in Fig. 5.

From the first year of the network onward, Elowitz and Leibler (2000) and Gardner et al. (2000) are the most cited publications. Their staying power and recurring citation in synthetic biology literature likely stem from (i) their description of the first successful synthetic genetic circuits and (ii) their description with reviews of synthetic biology research and progress such as Raimbault et al. (2016) or Cameron et al. (2014). In this way, Cameron et al. (2014) describes research within these publications as foundational to the larger synthetic biology field, and their importance is reflected due to their status as the most cited publications within our

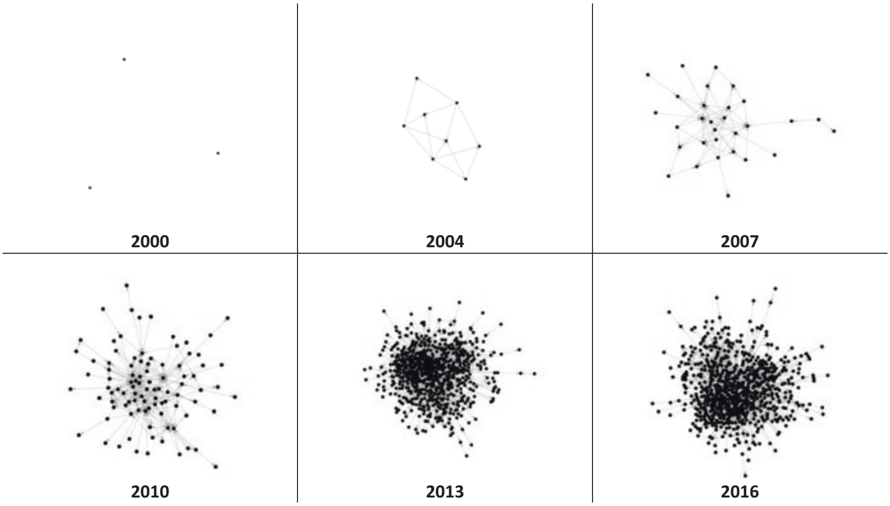

Fig. 4 Network evolution over key years

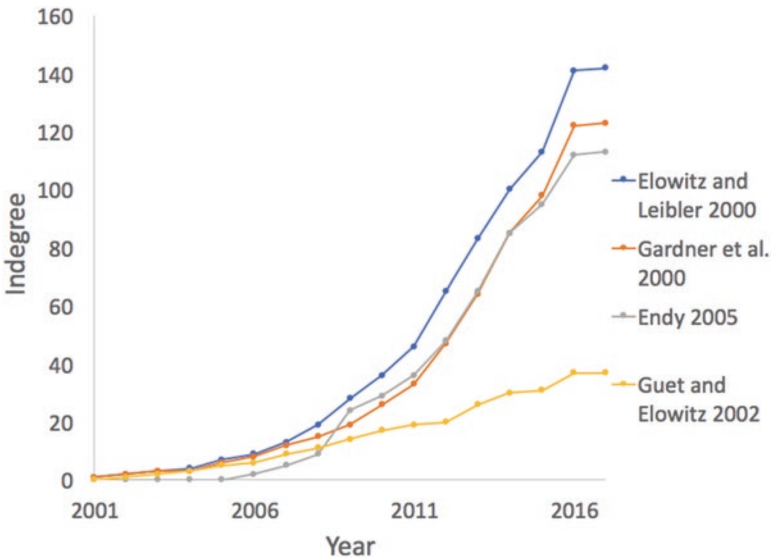

Fig. 5 Cumulative indegree of publications that appears in the top three at any point

dataset. Guet et al. (2002) was initially a high indegree performer, but did not experience the exponential growth seen in other papers circa 2008. The exponential growth of Elowitz and Leibler (2000), Gardner et al. (2000), and Endy (2005) show that these publications have had impact in shaping discourse within synthetic biology's biological, computational, and social sciences.

Unlike indegree, outdegree does not change over time, because a paper cannot cite future publications. Figure 6 shows the top publication by outdegree in each year.

In general, later publications cite more publications within the network. This is likely driven by the growth in scientific capacity and number of performers in the field (Raimbault et al. 2016; Oldham et al. 2012).

Like indegree, betweenness changes over time. The top three publications by betweenness were calculated for each year, and their full progressions are shown in Fig. 7.

Brophy and Voigt (2014), Khalil and Collins (2010), Carr and Church (2009), and Lu and Collins (2007) have significantly greater betweenness values than other publications. Each of them, particularly Brophy, became very central very quickly. Of note, Brophy and Voigt (2014) was an important "connector" in the field as soon as it was published, due to its pioneering work in genetic circuits.

Finally, the top three publications by eigencentrality were calculated for each year. Their respective eigencentralities were found over time and are shown in Fig. 8.

Elowitz and Leibler (2000) and Gardner et al. (2000) are at or near the top for the entire evolution of the network, indicating that their description and operation of synthetic genetic circuits had (and continues to have) substantial impact into various veins of synthetic biology research. Likewise, Endy (2005) maintains higher eigencentrality from 2005 to 2017, which is consistent with Raimbault et al.'s (2016) description of how the Endy lab has influenced the "programmatic discourses" of

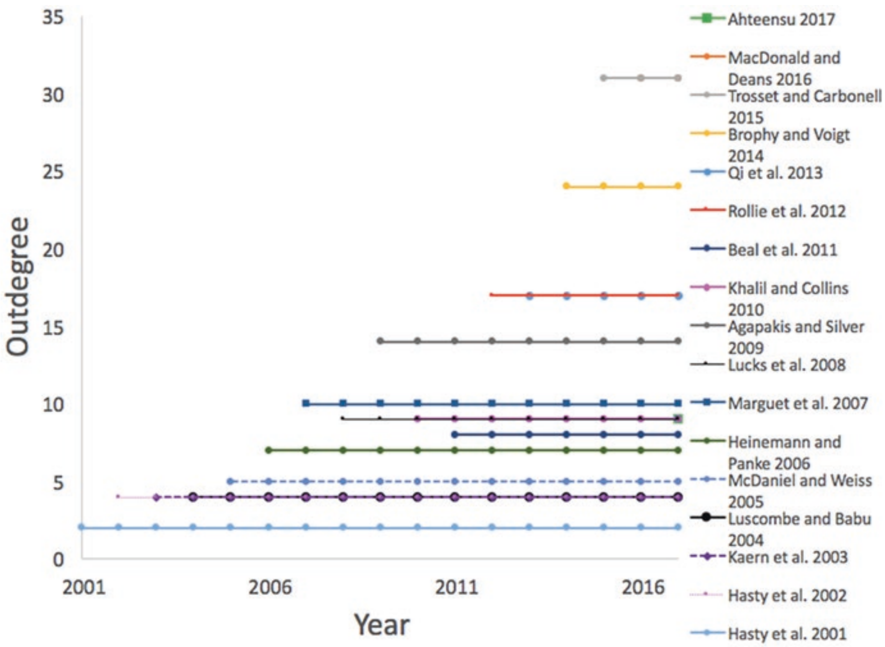

Fig. 6 Outdegree of top publication each year

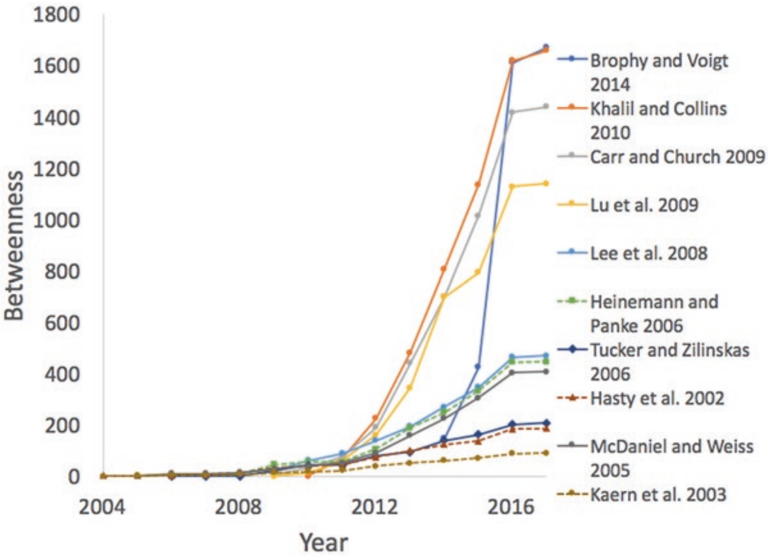

Fig. 7 Cumulative betweenness of publications to appear in the top three at any point

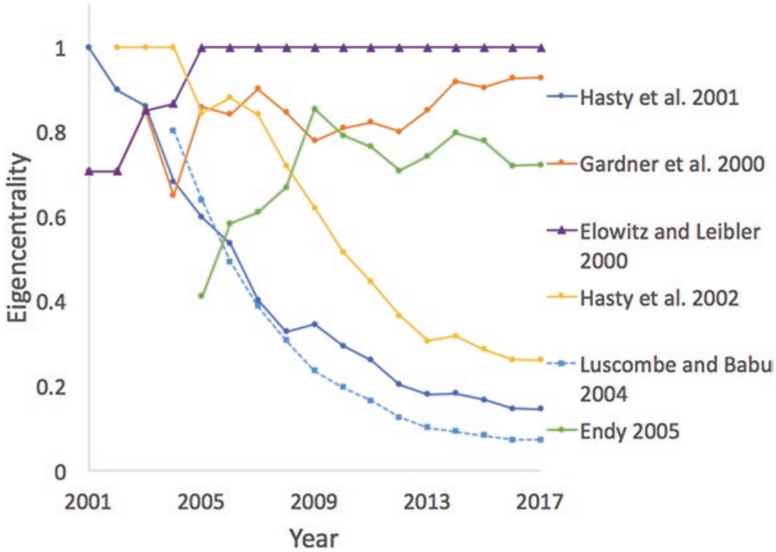

Fig. 8 Cumulative eigencentrality of publications to appear in the top three at any point

synthetic biology research by suggesting principles of engineering biology that have shaped inquiries into modeling, parts creation, and techniques for genomic assembly. These concepts are further noted as foundational to driving modern synthetic biology science by Cameron et al. (2014) and Trump (2016).

Other publications initially expressed higher eigencentrality yet declined in this metric over time. Hasty et al. (2001) and Hasty et al. (2002) build from work described in Gardner et al. (2000) regarding the design of synthetic circuits and thereby may be described as iterative rather than evolutionary developments in synthetic biology science.

Although excluded from the analysis of the network, isolates are still peer-reviewed publications and part of the synthetic biology network as a whole. These publications do not connect to any of the other publications in this particular network for a variety of reasons including corrupted or unreadable files, which do not allow the publication to be properly formatted for text mining. The numbers of publications and isolates over time are shown in Fig. 9, and the percentages of the publications that are isolates are shown in Fig. 10.

The number of publications in the network has been growing exponentially over the last 17 years. The number of isolates is as well, but not nearly as quickly. Figure 10 shows that over the last decade between 20% and 30% of publications have been isolates in any given year. Due to the small size of the network in the beginning, the percent varied greatly, particularly when there were three unconnected publications in 2000. Over the last 5 years, the percent of isolated publications has stabilized at roughly 20%, showing that the vast majority of new literature is adding to the existing network each year.

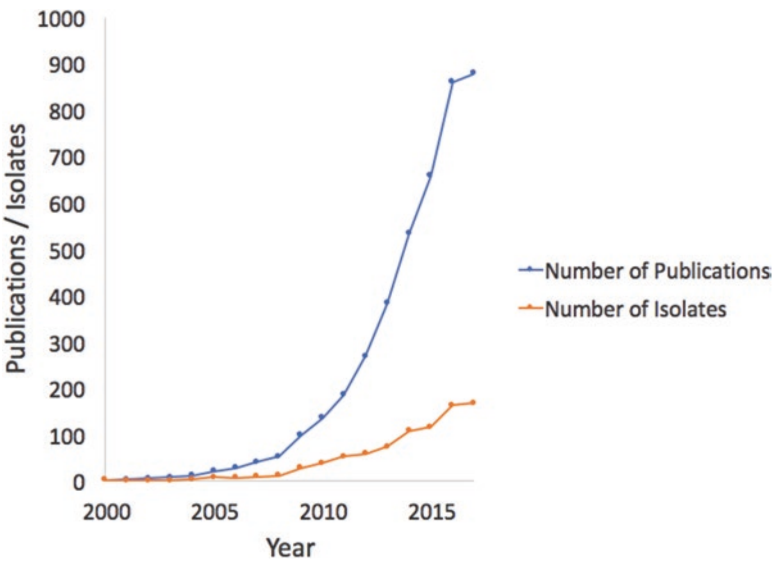

Fig. 9 Number of publications and isolates over time

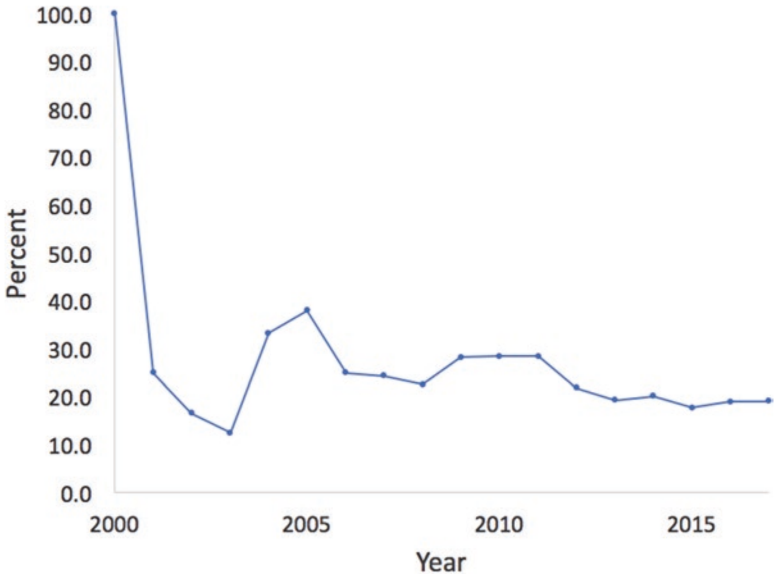

Fig. 10 Percent of publications that are isolates over time

Conclusion

The emerging field of synthetic biology is strongly linked and has significant publications that stand out among their peers, playing a critical role in connecting the network. The field is strongly linked because 80% of the 880 publications reviewed are a part of one large connected subgraph. Several publications appear multiple times in the final centrality measures and are then shown to rise to prominence rather quickly. These publications are either early seminal publications (Elowitz and Leibler 2000; Gardner et al. 2000), connecting publications (Brophy and Voigt 2014), or made important discoveries (Gibson et al. 2010; Paddon et al. 2013), showing that there are in fact several key publications connecting the field of synthetic biology. From a social sciences–oriented perspective, other publications have significant outdegree connection with state-of-sciences work (Seager et al. 2017; Cummings and Kuzma 2017) or use a mixture of natural and social sciences perspectives to inform regulatory gaps and opportunities (Oye et al. 2014; Bates et al. 2015).

As one of the first efforts to holistically characterize both biological and social sciences literature, this chapter sought to provide quantitative metrics to indicate how five individual communities of practice have developed over 2000–2017 and thereby identify key performers which influenced discourse in the field. Based upon quantitative metrics of cumulative indegree and eigencentrality, these top performers included breakthroughs in circuit engineering (Gardner et al. 2000; Elowitz and Leibler 2000; Brophy and Voigt 2014; Bradley et al. 2016), guidelines and philosophies for engineering biology (Endy 2005), reviews and progress reports of the field (Cameron et al. 2014; MacDonald and Deans 2016), and descriptions of product applications of synthetic biology (Trosset and Carbonell 2015; Xie and

Fussenegger 2015). Other important publications by cumulative indegree include important discoveries, such as Gibson et al.'s (2010) description of work to create the first cell with a synthetic genome and Paddon et al. (2013) which described the semi-synthetic production of artemisinic acid for antimalarial applications.

Furthermore, the broader synthetic biology community is not populating into various disparate groups but instead coalescing and commenting upon iterative developments and improvements to the field. More specifically, state of science and products literature appear to build directly from recent developments in synthetic biology science, in the form of either theoretical or applied research in biological or modeling research (Trump et al. 2018a; b). Risk, governance, and ELSI discussion *also* comment upon developments in "state of science" and "product" development, where they seek to respond to developments in the field and describe emerging benefits and challenges that may arise from synthetic biology's growing maturation and worldwide access (see examples such as with Tucker and Zilinskas 2006; Trump et al. 2017; Oye et al. 2014; Malloy et al. 2016; Schmidt et al. 2009). Rather than a collection of several unconnected and separated communities, synthetic biology's various communities of practice in physical and social sciences have at least acknowledged a small number of key developments in the field and may be using such works to develop a shared understanding of synthetic biology's physical and social sciences.

Such a citation network can also be helpful to identify those publications with the greatest degree of impact upon the synthetic biology community. Specifically, this chapter reviewed publication citation counts by indegree, where top performers here were generally described as motivational or seminal papers within reviews such as with Cameron et al. (2014) or Church et al. (2014). Such a citation network will help new entrants to the synthetic biology community identify key streams of research to follow (see Raimbault et al. 2016) or review possible motivations for future biological, computational, and social sciences research in the field. Such networks should be continually updated over time to account for adaptations in the synthetic biology literature, such as new trends in risk assessment protocols and testing procedures (Finkel et al. 2018; Linkov et al. 2017). Overall, network analysis and citation networks can help quantify and illustrate the growth trajectories, key performers, and disruptive events that spur growth in scholarly literature and may help the larger synthetic biology community to respond to such widely read and cited papers within and between their individual communities of practice.

References

Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H., & Weiss, R. (2005). A synthetic multicellular system for programmed pattern formation. *Nature, 434*, 1130–1134.

Bates, M. E., Grieger, K. D., Trump, B. D., Keisler, J. M., Plourde, K. J., & Linkov, I. (2015). Emerging technologies for environmental remediation: Integrating data and judgment. *Environmental Science & Technology, 50*(1), 349–358.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *The Journal of Mathematical Sociology, 2*(1), 113–120.

Bradley, R. W., Buck, M., & Wang, B. (2016). Tools and principles for microbial gene circuit engineering. *Journal of Molecular Biology, 428*(5), 862–888.
Brophy, J. A. N., & Voigt, C. A. (2014). Principles of genetic circuit design. *Nature Methods, 11*(5), 508–520.
Cameron, D. E., Bashor, C. J., & Collins, J. J. (2014). A brief history of synthetic biology. *Nature Reviews Microbiology, 12*(5), 381–390.
Carr, P. A., & Church, G. M. (2009). Genome engineering. *Nature Biotechnology, 27*(12), 1151–1162.
Church, G. M., & Regis, E. (2014). *Regenesis: How synthetic biology will reinvent nature and ourselves*. Basic Books.
Church, G. M., Elowitz, M. B., Smolke, C. D., Voigt, C. A., & Weiss, R. (2014). Realizing the potential of synthetic biology. *Nature Reviews Molecular Cell Biology, 15*(4), 289–294.
Cummings, C. L., & Kuzma, J. (2017). Societal risk evaluation scheme (SRES): Scenario-based multi-criteria evaluation of synthetic biology applications. *PLoS One, 12*(1), e0168564.
De Solla Price, D. J. (1965). Networks of scientific papers. *Science, 149*(3683), 510–515.
Elowitz, M. B., & Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature, 403*(6767), 335–338.
Endy, D. (2005). Foundations for engineering biology. *Nature, 438*(7067), 449–453.
Finkel, A. M., Trump, B. D., Bowman, D., & Maynard, A. (2018). A "solution-focused" comparative risk assessment of conventional and synthetic biology approaches to control mosquitoes carrying the dengue fever virus. *Environment Systems and Decisions, 38*(2), 177–197.
Gardner, T. S., Cantor, C. R., & Collins, J. J. (2000). Construction of a genetic toggle switch in Escherichia coli. *Nature, 403*(6767), 339–342.
Garfield, E. (1955). Citation indexes for science: A new dimension in documentation through Association of Ideas. *Science, 122*(3159), 108–101.
Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., et al. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science, 329*(5987), 52–56.
Guet, C. C., Elowitz, M. B., Hsing, W., & Leibler, S. (2002). Combinatorial synthesis of genetic networks. *Science, 296*(5572), 1466–1470.
Hasty, J., Mcmillen, D., Isaacs, F., & Collins, J. J. (2001). Computational studies of gene regulatory networks: In numero molecular biology. *Nature Reviews Genetics, 2*(4), 268–279.
Hasty, J., Mcmillen, D., & Collins, J. J. (2002). Engineered gene circuits. *Nature, 420*(6912), 224–230.
Jacob, R., Harikrishnan, K. P., Misra, R., & Ambika, G. (2017). Measure for degree heterogeneity in complex networks and its application to recurrence network analysis. *Royal Society Open Science, 4*, 160757. https://doi.org/10.1098/rsos.160757.
Kahl, L. J., & Endy, D. (2013). A survey of enabling technologies in synthetic biology. *Journal of Biological Engineering, 7*(1), 13.
Khalil, A. S., & Collins, J. J. (2010). Synthetic biology: Applications come of age. *Nature Reviews Genetics, 11*(5), 367–379.
Leydesdorff, L. (2007). Betweenness centrality as an indicator of the interdisciplinary of scientific journals. *Journal of the American Society for Information Science and Technology, 58*(9), 1303–1319.
Linkov, I., Trump, B. D., Wender, B. A., Seager, T. P., Kennedy, A. J., & Keisler, J. M. (2017). Integrate life-cycle assessment and risk analysis results, not methods. *Nature Nanotechnology, 12*(8), 740.
Linkov, I., Trump, B. D., Anklam, E., Berube, D., Boisseasu, P., Cummings, C., et al. (2018). Comparative, collaborative, and integrative risk governance for emerging technologies. *Environment Systems and Decisions, 38*(2), 170–176.
Lu, T. K., & Collins, J. J. (2007). Dispersing biofilms with engineered enzymatic bacteriophage. *Proceedings of the National Academy of Sciences, 104*(27), 11197–11202.
MacDonald, I. C., & Deans, T. L. (2016). Tools and applications in synthetic biology. *Advanced Drug Delivery Reviews, 105*, 20–34.

Malloy, T., Trump, B. D., & Linkov, I. (2016). Risk-based and prevention-based governance for emerging materials. *Environmental Science and Technology., 50*, 6822.

Marchant, G. E., Abbot, K. W., & Allenby, B. (Eds.). (2013). *Innovative governance models for emerging technologies*. Cheltenham: Edward Elgar Publishing.

McPherson, M., Smith-Lovin, L., & Cook, J. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*, 415–444.

Merad, M., & Trump, B. D. (2019). *Expertise under scrutiny: 21st century decision making for environmental health and safety*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-20532-4.

Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q.-A., Tran, A. B., Paull, M., Keasling, J. D., Arkin, A. P., & Endy, D. (2013). Precise and reliable gene expression via standard transcription and translation initiation elements. *Nature Methods, 10*(4), 354–360.

Oldham, P., Hall, S., & Burton, G. (2012). Synthetic biology: Mapping the scientific landscape. *PLoS One, 7*(4), e34368.

Ostergaard, S., Olsson, L., & Nielsen, J. (2000). Metabolic engineering of Saccharomyces cerevisiae. *Microbiology and Molecular Biology Reviews, 64*(1), 34–50.

Oye, K. A., Esvelt, K., Appleton, E., Catteruccia, F., Church, G., Kuiken, T., et al. (2014). Regulating gene drives. *Science, 345*(6197), 626–628.

Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., et al. (2013). High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature, 496*(7446), 528–532.

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

Ro, D.-K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., Chang, M. C. Y., Withers, S. T., Shiba, Y., Sarpong, R., & Keasling, J. D. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature, 440*(7086), 940–943.

Schmidt, M., Kelle, A., Ganguli-Mitra, A., & de Vriend, H. (Eds.). (2009). *Synthetic biology: The technoscience and its societal consequences*. Dordrecht: Springer Science and Business Media.

Seager, T. P., Trump, B. D., Poinsatte-Jones, K., & Linkov, I. (2017). Why life cycle assessment does not work for synthetic biology. *Environmental Science and Technology, 51*, 5861.

Trosset, J.-Y., & Carbonell, P. (2015). Synthetic biology for pharmaceutical drug discovery. *Drug Design, Development and Therapy, 9*, 6285.

Trump, B. (2016). *A comparative analysis of variations in synthetic biology regulation*. University of Michigan.

Trump, B. D. (2017). Synthetic biology regulation and governance: Lessons from TAPIC for the United States, European Union, and Singapore. *Health Policy, 121*(11), 1139–1146.

Trump, B., Cummings, C., Kuzma, J., & Linkov, I. (2017). A decision analytic model to guide early-stage government regulatory action: Applications for synthetic biology. Regulation and Governance.

Trump, B. D., Cegan, J., Wells, E., Keisler, J., & Linkov, I. (2018a). A critical juncture for for synthetic biology. *EMBO Reports., 19*(7), e46153.

Trump, B. D., Foran, C., Rycroft, T., Wood, M. D., Bandolin, N., Cains, M., et al. (2018b). Development of community of practice to support quantitative risk assessment for synthetic biology products: Contaminant bioremediation and invasive carp control as cases. *Environment Systems and Decisions, 38*(4), 517–527.

Trump, B. D., Cegan, J., Wells, E., Poinsatte-Jones, K., Rycroft, T., Martin, D., Warner, C., Perkins, E., Warner, C., Wood, M., & Linkov, I. (2019). Co-evolution of physical and social sciences in synthetic biology. *Critical Reviews in Biotechnology, 39*, 351.

Tucker, J. B., & Zilinskas, R. A. (2006). The promise and perils of synthetic biology. *The New Atlantis, 12*, 25–45.

Weisstein, E. W. (2017) Outdegree. From MathWorld – A Wolfram Web Resource.

Xie, M., & Fussenegger, M. (2015). Mammalian designer cells: Engineering principles and biomedical applications. *Biotechnology Journal, 10*(7), 1005–1018.