Citation Analysis of Genetic Circuit Design Literature

Galen Michael Seilis
[redacted]

December 11, 2020

Contents

1 Introduction **3**
- 1.1 Genetic Circuit Design . . . 3
- 1.2 Network Analysis . . . 3
- 1.3 Literature Review . . . 4
- 1.4 Objective . . . 4

2 Methodology **5**
- 2.1 Obtaining Data . . . 5
- 2.2 Analysis . . . 6

3 Results and Discussion **6**
- 3.1 Exploratory Data Analysis . . . 6
- 3.2 Article Importance . . . 10
- 3.3 Author Importance . . . 11
- 3.4 Keyword Importance . . . 13
- 3.5 Authors → Keywords Associations . . . 16

4 Challenges **17**

5 Conclusion **18**

1 Introduction

This section will briefly introduce each of the main background topics of the project.

1.1 Genetic Circuit Design

Genetic circuit design is aims to design biological systems with programmed functionalities using molecular parts such as deoxyribonucleic acids, proteins, and ribonucleic acids. Inspired by the accomplishments of designing electrical circuits, the application of engineering principles have given us many examples of biological systems that have programmed functionalities. The earliest accomplishments of the field include the biological implementation of switches [1], represillators (i.e. repressor oscillators) [2], and logic gates [3]. The earliest biological implementations relied heavily on gene-gene interactions via promoter binding, but today's circuits also use miRNA [4], or even complex protein-DNA-RNA complexes such as CRISPR/Cas9 editing systems [5] Beyond basic circuits and switches, genetic circuit design promises advancements in drug production [6], biofuel production [7], and cancer detection. [8]

The most substantial limitations lay within biochemistry. Unlike with electrical components that can be sufficiently insulated by space and semi-conductive materials, most molecules within a cell are free to diffuse within their membrane compartment. As a result, components of genetic circuits sometimes have undesired cross-talk with other components due to chemical and physical interactions, which is often termed non-orthogonality. A second fundamental problem is metabolic burden, where producing and maintaining the desired circuit requires a substantial amount of energy and nutrients.

In the face of these limitations of single-strain circuits, an interest in using multi-strain colonies for distributed biological function has developed.[9] Multi-strain communities help avoid non-orthogonality of components by keeping them in separate cell lines, preventing opportunities for interaction.[9] They help with metabolic burden by distributing the components across strains so that no one bacterial strain has to synthesize all components of the biological circuit.[9]

Tools such as Gro, DiSCUS, BactoSIM, and BSim have allowed researchers to examine the emergent behaviour of bacterial communities to investigate their dynamics and long term stability.[10, 11, 12, 13] Other computer aided design tools such as Cello have empowered researchers to automatically design circuits for single-strain colonies with a relatively high success rate compared to hand-tuning approaches, however current automated methods do not convert these circuits into distributed multi-strain systems.[14]

1.2 Network Analysis

While people have considered network-like entities such as trading routes and spider webs since antiquity, the formal mathematics of graph theory that underpins network analysis has only been around since 1736.[15] Graphs are 2-tuples that contain a set of elements called the vertex set along with another set of 2-tuples called the edge set which is a subset of the Cartesian product of the vertex set. While pure mathematics often concerns itself with sets of numbers, graphs can be used to represent collections of either numerical or non-numerical things. Network analysis involves encoding observations into vertices and relations and further analyzing the properties of the resulting

'graph'.[1]

The methods of analysis vary, but include topics such as statistics, differential equations, linear albegra, and discrete mathematics.[16] The domains that have benefited from network analysis include epidemiology[16], dyamical systems[16], traffic and routing[16], distributed and concurrent computing[16], electrical engineering[17], cellular biology[18], social networks and public health[19], and citation analysis[20] among many others.

1.3 Literature Review

One of the challenges that researchers face is the ever-growing body of literature that they must assess for relevance to their own interests and work. One study found that from January 1st 2020 to June 30th 2020 there had been 23,634 studies indexed on Web of Science and Scopus related to Coronavirus disease 2019 (COVID-19)[21], illustrating that important and (relatively) narrow topics can have an overwhelming literature in terms of size and growth. While COVID-19 is given as an extreme example for the sake of illustration, there many topics that have been covered by thousands of studies.

The novice researcher getting into their first domain, as well as the experienced researcher expanding into additional domains, are confronted with databases containing a relatively large number of potentially relevant articles compared to the time they have available to perform reading. Therefore, entering into a new field in such a way that a researcher can quickly grasp the foundational knowledge needed to perform their own work requires prioritization of the literature. Often students and junior researchers are given key papers and concepts from their principal investigators that initiate an effective introduction to the topic.

Guidance is not always available, and therefore many researchers have to work out their own system of better understanding the literature. Sometimes it is possible to use search engines to query for landmark papers and review papers that are well-known. However, some subjects do not have well-known or easily-found landmark or review papers, or the current literature has moved on from those early papers. Therefore there is a use case for analyzing the literature itself to heuristically recommend where a researcher should begin their reading.

1.4 Objective

The objective of this study was to answer three basic questions about the genetic circuit design literature. The first is which papers should be read, the second is which authors should be followed, and the third is what keywords are prevalent to this subject through time. The former two questions pertain to the importance works and people of the discipline which should be considered first, while the third question pertains to what the general themes of the subject are.

Implicit in these questions is the vague query of *who and what is important* in a field of study. "Importance" in this study is not formally defined as it is both vague and varied in its meaning, however operational notions of importance will be used and discussed.

[1] While there isn't a complete consensus in the literature, the term "graph" often represents the pure mathematical object whereas a "network" is similar to a graph that itself represents real data in the form of attributes on the vertex and edge set in addition to encoding the relations between vertices. In this document I use these terms somewhat interchangeably.

2 Methodology

While much of this project's methodology is in the data analysis that will be implicitly available to the reader through the results and discussion, the general approach is outlined in this section.

2.1 Obtaining Data

This project did not perform any experiments because it did not include any controls, but data collected from the database can be considered valid for exploratory analysis. The main variables of interest were the article titles, authors, publication dates, and keywords. These variables can be found in the Web of Science database which has premium access available to UNBC students. However, the application programming interface for this database was not available to us, only the graphical user interface, because it is a separate subscription service that UNBC does not provide to their students. Manually searching and downloading the data for thousands of articles can take an exorbitant amount of time, and therefore we took a webscraping approach to obtain the data. In development, and iterative and test-driven development approach was taken to create a script that first extracted one article correctly before implementing loop structures for a broader search. While the implementation details are quite complicated due particular sequences of HTML tags being required to navigate and interact the web interface, an abstract procedure will be briefly outlined.[2]

First, we started with a collection of articles that were used to 'seed' the search of the webscraper. In general these seeds must be relevant to the topic of interest, but it was anticipated that better results would be obtained if the seeds were review papers or landmark papers. The papers chosen for this analysis are listed in Table 1.

Table 1: List of articles that were used as webscraper search seeds

Type	First Author	Publication Date	Citation
Article	Al-Radhawi	2020	[22]
Article	Menon	2019	[23]
Article	Misirli	2014	[24]
Article	Xiang	2018	[25]
Review	Brophy	2014	[26]
Review	Karkaria	2020	[9]
Review	Xia	2019	[27]

Each seed in Table 1 was queried against the Web of Science database for (1) articles that the seed cited, (2) information about the seed article itself, and (3) articles that cite the seed article. The results of these queries were downloaded, and the content of the results setup the next step of the scraping procedure. For each article among those collected from the first iteration, they were similarly searched against the database. Without stop conditions, such a search would likely continue indefinitely, so limitations are the search were included. A collection of keywords related to the discipline were used to bound the search under the rule that an article's citation relations would be only be explored if it had at least one keyword in common with the bounding keywords. This operationally defined the scope of the search, and creating a stop condition where all articles

[2] I will refer to it as a 'procedure' rather than an algorithm because many of the details pertaining to obtaining HTML tags and button-clicking that would complete the description are left out for simplicity.

with bounding keywords had been explored. The bounding terms chosen in our study were *genetic circuit*, *genetic circuits*, *genetic circuit design*, *gene circuits*, *genetic logic gate*, *synthetic biological circuit*, and *biological circuit*.

2.2 Analysis

With the data downloaded from the Web of Science database, it was then collated into graph representations for exploratory data analysis. Because there were no controls, no experimental interpretation should be taken from the results in this study. The rest of the methods of analysis are left to the results and discussion section.

3 Results and Discussion

This section summarizes and discusses the results from analyzing the network data described in the methodology.

3.1 Exploratory Data Analysis

This subsection examines various plots that were generated during the exploration of the data. They do not all pertain directly to the main goals of (1) most important articles, (2) most important authors, or (3) most important keywords, but they provide context to the dataset.

Having obtained the bibliographic data from the Web of Science database and collated it into a graph representation, it became possible to visualize structure in the data. Figure 1 is a plot of the citation graph which illustrates the pattern of citations in the genetic circuit design literature.

Visual inspection of the graph suggests that it is weakly connected, but not strongly connected, which is confirmed by utility functions in the NetworkX package.[28] This is what one would expect given that the search performed on the literature was conceptually an expansion of a few articles into this larger graph by expanding on citations (edges). Another observation about the graph is that there seems to be relatively node-dense center region with sparser paths leading outward. This suggests that there may be a power-law distribution of centrality. The nodes themselves are coloured from purple-to-yellow to represent low-to-high degree centrality, and thus Figure 1 seems to show a nearly-uniform distribution of degree centrality.[3]

[3]Note: Other measures of centrality including betweeness centrality and eigencentrality were also relatively uniform.

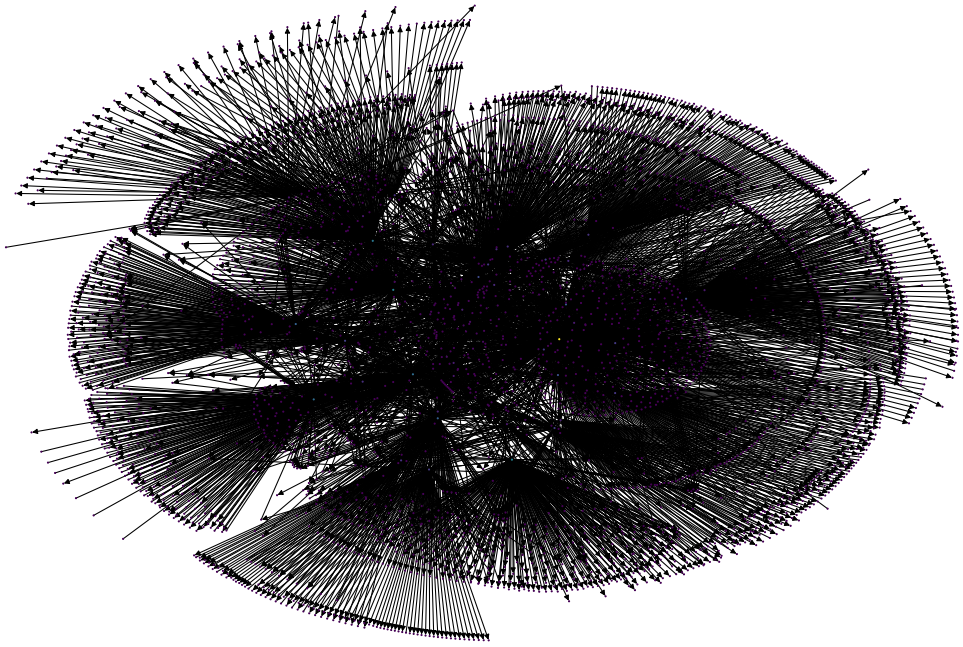

Figure 1: Directed network where the nodes are articles coloured by degree centrality and the directed edges represent the relation that one article points to the articles it cites.

From the attribute data of the citation graph, a second graph was made such that the nodes are the authors, and the undirected edges represent the relation between authors that they have coauthored at least once. This coauthorship network is given in Figure 2. Unlike the citation network, the coauthorship network is broken into many connected components that represent different groups of authors that have authored papers together.

Visual inspection of Figure 2 suggests that most such social groups range from one person up to a couple dozen people, with some of them being fully connected. The social groups that are not completely connected are possibly due to a combination of two phenonmenon. The first is that many graduate students will publish some papers with their supervisor's social group before leaving academia altogether, resulting a history of students within a professor's group that are separated by time. The second possibility is due to migration of academics among communities. As the social network of a researcher changes, so too will the members of the academic community they have opportunities to coauthor papers with.

Among the social groups there is a notable group that is larger than any of the others. This social group appears to have a many peripheral members, and the degree-centrality colouring suggests that there are subgroups that are more well-connected than others in that group. This social group is the largest, most internally connected group of researchers compared to the othres. It is within this group that one should expect to find the most influential researchers, as well as those who are the most likely to collaborate and be collaborated with.

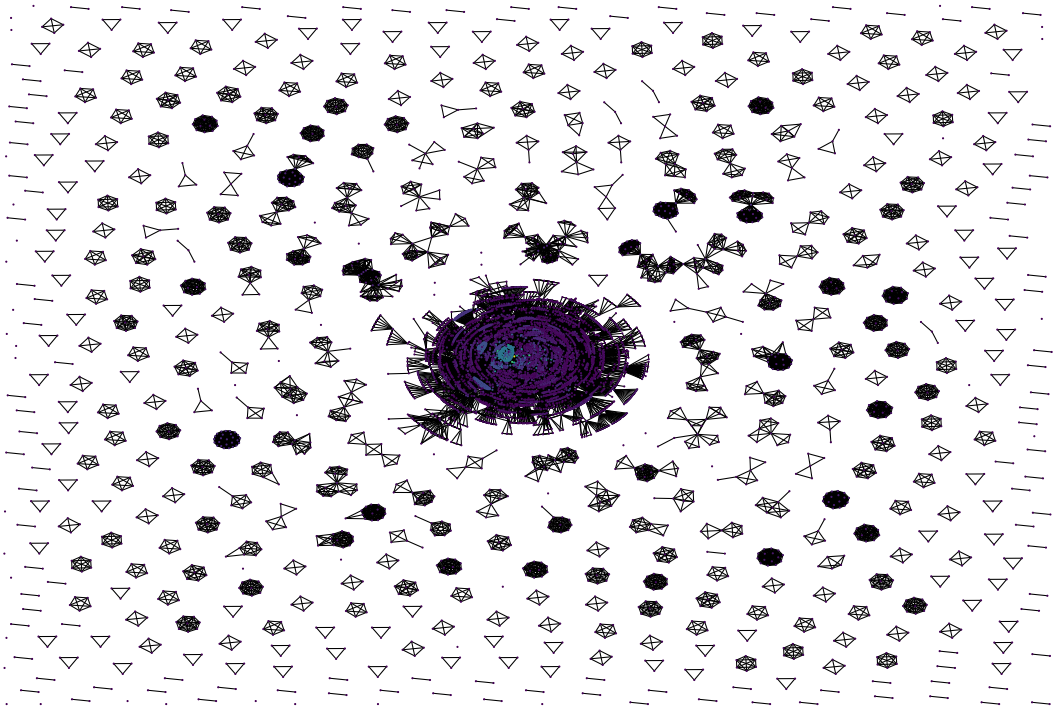

Figure 2: Network where the nodes are researchers coloured by degree centrality and the edges represent the relation at least one article had both researchers among its authors.

From the attribute data of the citation network, a third network was constructed where the nodes represent the keywords annotated to the papers, and the edges represent whether two keywords are used together in at least one article. While this keyword network is similar to the coauthoriship network in have multiple connected components, it has only 8 components. The notably-largest of the eight connected components is likely due to the fact that the search procedure used a keyword relation to expand the search. The remaining keywords must be from papers that were discovered by the search which did not themselves have any of the closure keywords, resulting in the citations of those articles not being explored.

The degree centrality shows that most of the nodes are of similar centrality, with a small number in drawn in the center of the largest connected component being particularly central. This higher-centrality terms are not necessarily all closure keywords, however they are more likely due to their role in the search.

Another notable property of the largest connected component is that while it has peripheral nodes like the previously-discussed networks, there is a higher amount of connectivity among the peripheral nodes leading to fewer of those fan-like structures in the plotted representation of the graph.

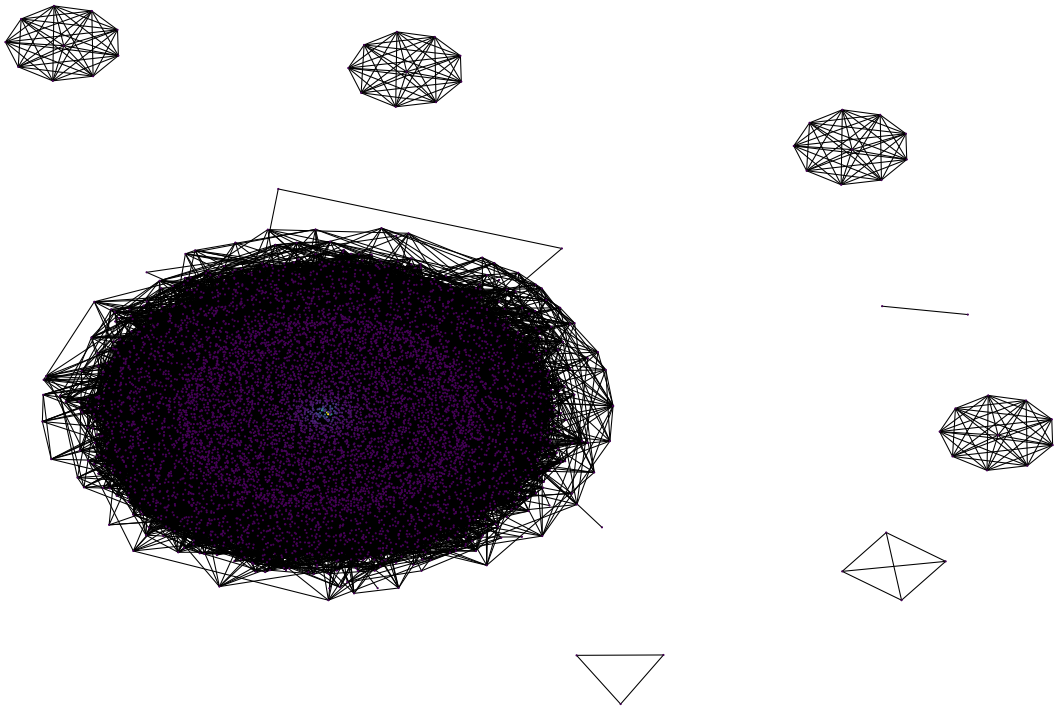

Figure 3: Network where the nodes are keywords coloured by degree centrality and the edges represent the relation at least one article had both keywords among its keywords.

Visual inspection of all three networks suggested that their degree centrality distributions may be power-lawed (i.e. scale-free). This is further evidenced from visually inspecting Figure 4 which illustrated that very few of the nodes have most of the connections and most nodes have very few connections. A network is scale-free if it satisfies the relation

$$P(k) \sim k^{-\gamma} \tag{1}$$

where $P(k)$ is the fraction of nodes in the network with k connections and γ is a scale parameter.[16] This defintion considers only the proportionality to be sufficient as similar exponentials with other terms or factors are also sometimes considered scale-free.[16] Evidence of these networks being scale-free further emphasizes that there is a minority of articles, authors, and keywords are are exponentially more important than others in terms of how they relate to the rest of the members of the network.

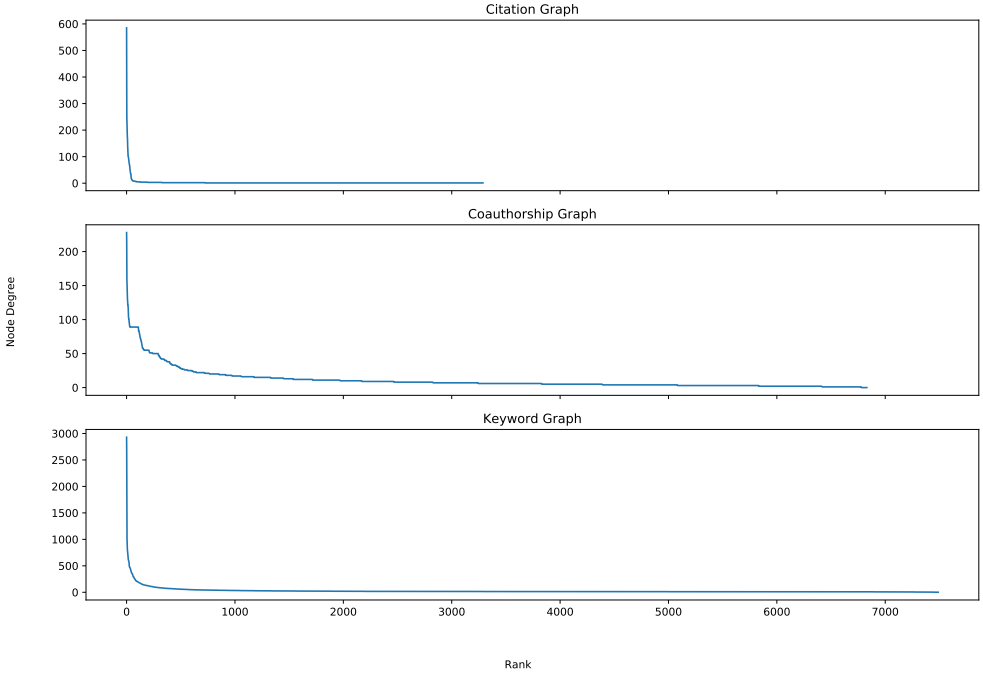

Figure 4: The degree distributions of the (top) citation network, (middle) coauthorship network, and (bottom) keyword network. The total degree is shown for the citation network.

3.2 Article Importance

In this subsection we try to identify particular articles in the discipline genetic circuit design as recommended for a first reading.

One notion of how important an article is pertains to how often it is cited. If an article is cited, that implies that a researcher thinks that article is worth the attention of their readers. While attention is a limited measure of importance, and possibly orthogonal to other notions of importance, it indicates what articles are worth paying attention to by the researchers in a given field. Figure 5 shows the distribution of articles in order of how cited they are for the top such articles. At the top of the ranked list is Brophy 2014, which was one of the articles that was used to seed the search of articles. Of the other seed articles, none of them made this top-ten list. While the last names of the first authors give some hint at the most important authors, that ignores tha the last author is often a principal investigator or supervisor that may have influenced a large quantity of the literature less directly.

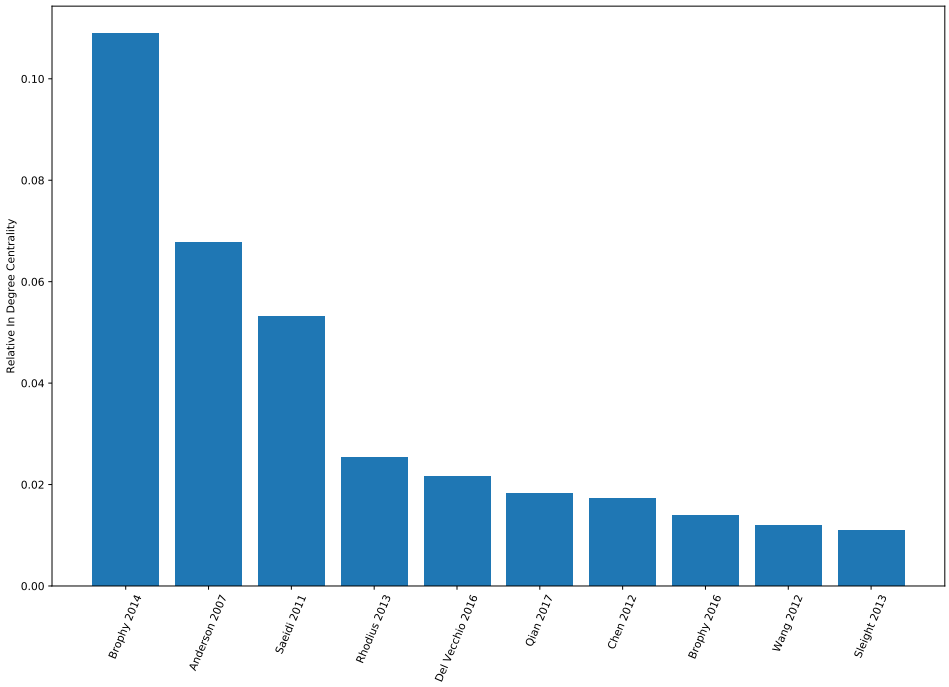

Figure 5: The rank-ordered distribution of top-cited articles.

Given that Brophy 2014 had already been read as background material for this project, the next three papers to read according to Figure 5 are Anderson 2007, Saeidi 2011, and Rhodius 2013. None of the recommended papers in the top ten are more recent than 2016, but it is unlikely that all studies since then are unimportant. Rather, it takes time for studies to accumulate citations and so older studies are more likely to be on this list.

3.3 Author Importance

In this subsection we try to identify particular authors whose work should be followed within the genetic circuit design literature.

While not all articles are high-quality, there is something to be said for looking up authors that have published a lot of papers. It is possible that such authors have exchanged quality for quantity, but it is also possible that they have an excetionally high productivity. Writing articles requires thinking about what to write, and thinking about what to write requires thinking about the subject matter. With the expectation that more experiences in writing about a subject lead to improved understanding, one expects the content to be more refined in the later works of such authors. At least, that is what is supposed to have been the case when articles are submitted for publication, but it is certainly not a certainty. There is an ongoing discussion of quality assurance in publication because many poor-quality articles have been published[29], and that writing alone is not sufficient to really understand a subject and produce good research articles.[30] Because tenureship and other career benefits and advancements in part depend on metrics such as the H-index, there can be a high incentive to publish a high quantity of poor-quality articles.[31] With these cautions in perspective that no assessment of number of publications will be a completely accurate measure

of author importance, it may still be worth tentatively examining the 'fastest-publishing' authors.

Figure 6 gives the rank-ordered distribution of the authorship of papers, which indicates that Silver, Stirling, O'Keefe, Way, Bitzan, and Oliver have published substantially more articles than other authors in genetic circuit design. Reading their most-cited articles would give some indication of whether their high publication rate indicates if their work should be followed or avoided.

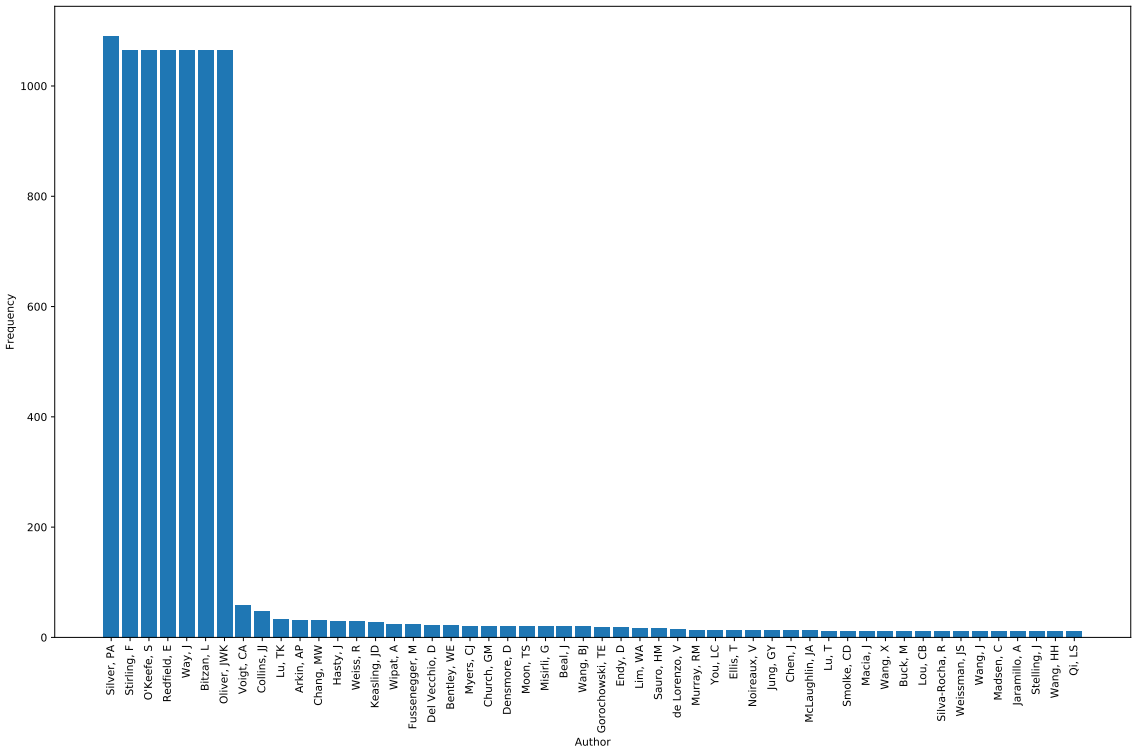

Figure 6: The rank-ordered distribution of authorship of articles.

Just like the number of publications that an author has, the number of times their work gets cited is not necessarily because their work is high value. However, looking to the authors whose work gets a lot of attention by others who work in the same field is an *indication* of their work being important. The rationale behind this is that if a researcher cites another, then that tacitly implies that the cited work is worth the attention of the reader. Figure 7 shows the distribution of weights where each weight as the sum of the number of times each article by the author was cited, divided by the number of authors of the article. This is given in Equation 2 as

$$W_i = \sum_j \frac{w_j}{n_j} \tag{2}$$

where W_i is the weight assigned to the ith author, w_i is number of times the jth article was cited, and n_j is the number of authors of the jth article, summed over all articles authored by the ith author. Comparing Figure 2 to Figure 7, we can see that the top seven authors in terms of publishing articles are all in the top eleven authors of the citation-weighted distribution. However, the citation-weighted distribution had four authors as the top four that were not in the aforementioned seven. The first was Voigt, which was the eighth rank in the authorship frequency distribution. Brophy,

Wong, and Nguyen were not on the authorship frequency bar plot at all, suggesting that they are cited a lot more than their number of publications would suggest. Of all these top articles, only Brophy was in the initial seed articles.

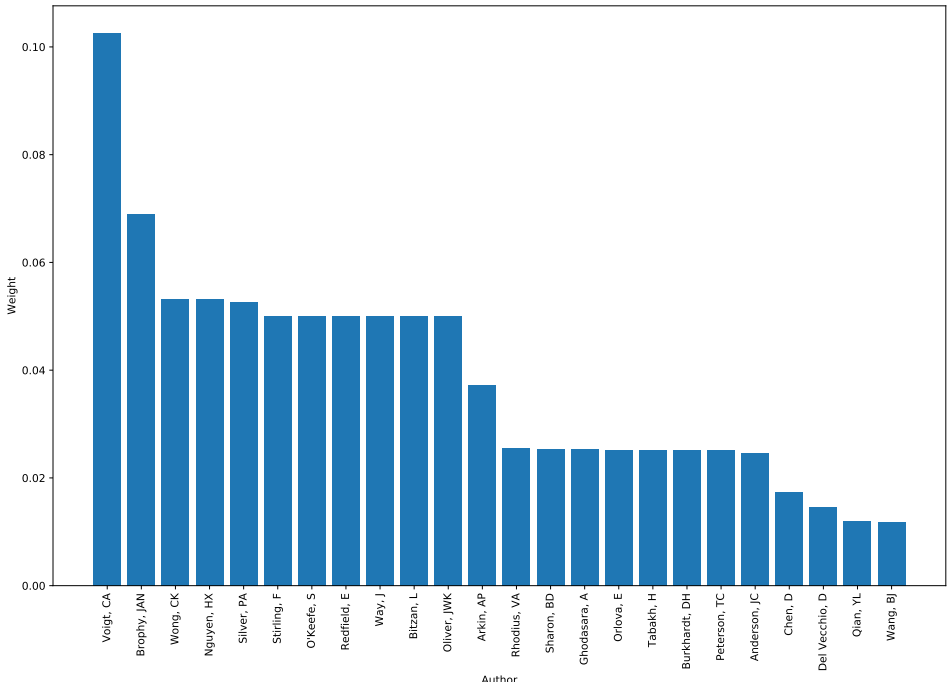

Figure 7: The rank-ordered distribution of author's citation weight. Only authors with a weight $>$ 0.01 are shown.

Based on the results given above the top researchers to follow are Voigt, Brophy, Wong, Nguyen, Silver, Stirling, O'Keefe, Redfield, Way, Bitzan, and Oliver.

3.4 Keyword Importance

In this subsection we try to identify particular keywords that are prevalent in the field of genetic circuit design in order to better understand the general themes of the subject.

While there is a lot of article-to-article variability in keywords, the overall literature should have an unequal distribution of keywords among there articles because not every article covers the same set of subtopics. Figure 8 shows the frequency distribution of the most common keywords in the literature, with only a four clearly surprising their neighbors. The first is the name of the bacterium, *Escherichia coli*, which is one of the most common model organisms in all of biology. *Escherichia coli* is a desirable model organism as it is easy to culture, to genetically modify, and poses relatively low risk to humans. The second keyword is *synthetic biology*, which is the study of how to design aritifial life organisms. This keyword is semantically similar to genetic circuit design, and one might even consider genetic circuit design to be a propr subtopic of synthetic biology. The next two keywords are getting at the same thing: *expression* and *gene-expression*. The former may have ranked higher because it is less particular to write about the expression of various traits of

organisms than to write specifically of gene expression. Both are related to genetic circuit design because genetic circuits are composed of genes that are expressed via transcription (and other processes), while one may also write about the expression of traits. These are two similar concepts oriented around the same underlying molecular biology and cell biology. The fourth term is *design*, which is naturally related to genetic circuit design. "Design" was not included as a closure keyword because it is too broad, and would expand the search far beyond genetic circuit design. None of the closure keywords made it into this top list of keywords, but all of them can be related to the various organisms of study, macromolecules, circuit theory, engineering, biology, and chemistry that underpin genetic circuit design.

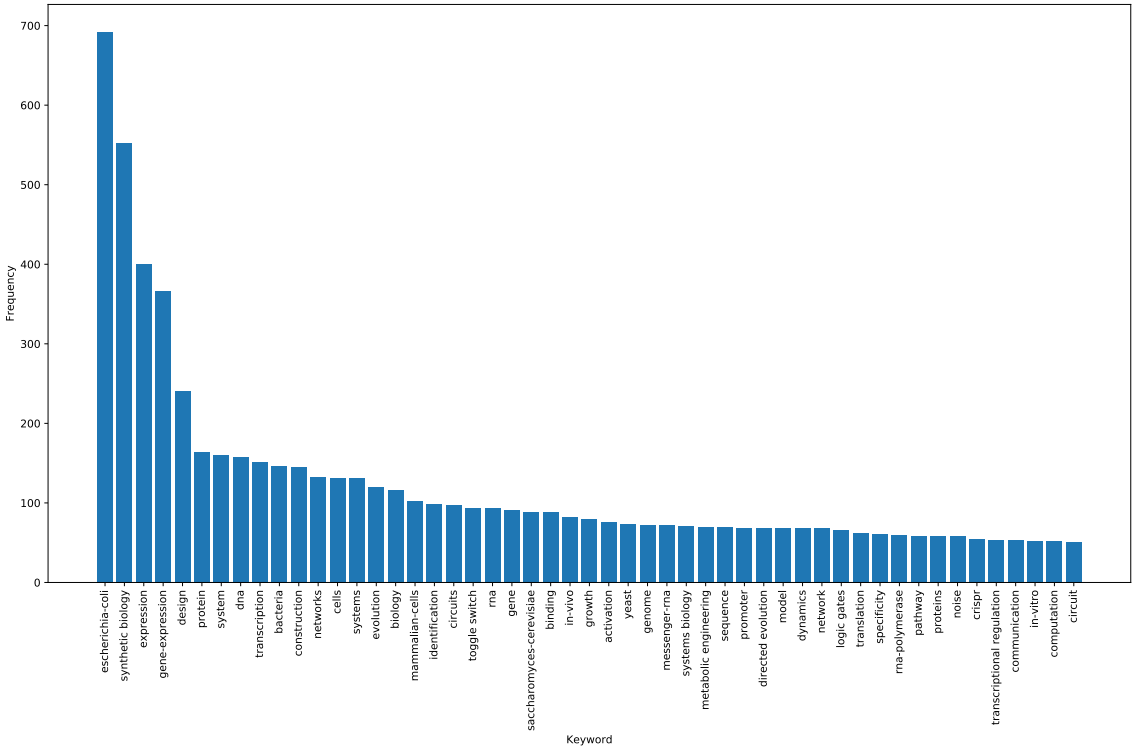

Figure 8: Frequency distribution of top 50 article keywords in rank order.

The keyword frequency is expected to change as a function of time because subjects grow in the number of concepts that have been invented or applied, and the number of articles published provided the field is either growing or shrinking. In the case of genetic circuit design, the field appears to be growing given the number of keywords used as a function of time (Figure 9). Figure 9 shows an exponential-like growth. Most of the terms particular to genetic circuit design start use in the 1990's or later, leaving earlier keywords to more general terms such as *Escherichia coli*. Notably the year 2020 has a smaller number of publications. This could be due somewhat to COVID-19 preventing researchers from publishing for ethical, technological, of logistical reasons, as well as researchers changing their focus towards researching COVID-19. This is plausible in part because many researchers in genetic circuit design have relevant backgrounds in molecular biology that could be repurposed toward studying coronaviruses. Another plausible alternative to this is that the number of papers indexed in Web of Science will lag behind the number of studies that have been published, and that one should expect the number of 2020 publications in genetic circuit design to increase in the coming year. A more peculular observation is that there are already some

papers indexed as published in 2021. While human error is always possible, it is more likely that papers that have been approved for publication in 2021 can be indexed in the Web of Science database, making it possible for researchers to get some early indication of papers coming up next year.

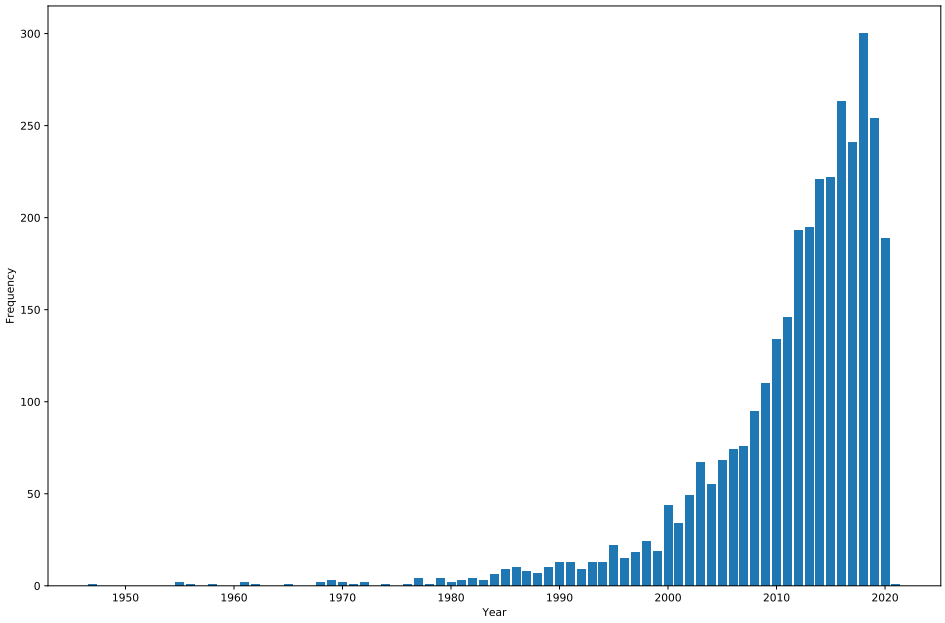

Figure 9: Frequency of keywords in articles published by year.

The information about top keywords and number of keywords over time can be combined into considering the top keywords over time. Figure 10 shows the same top keywords as they change over time. It is notable that "*Escherichia coli*", "synthetic biology", "expression", "gene-expression", and "design" have shown a trend in genetic circuit design papers that more closely resembles exponential growth than any of "protein", "DNA", "transcription", "bacteria", or "system".

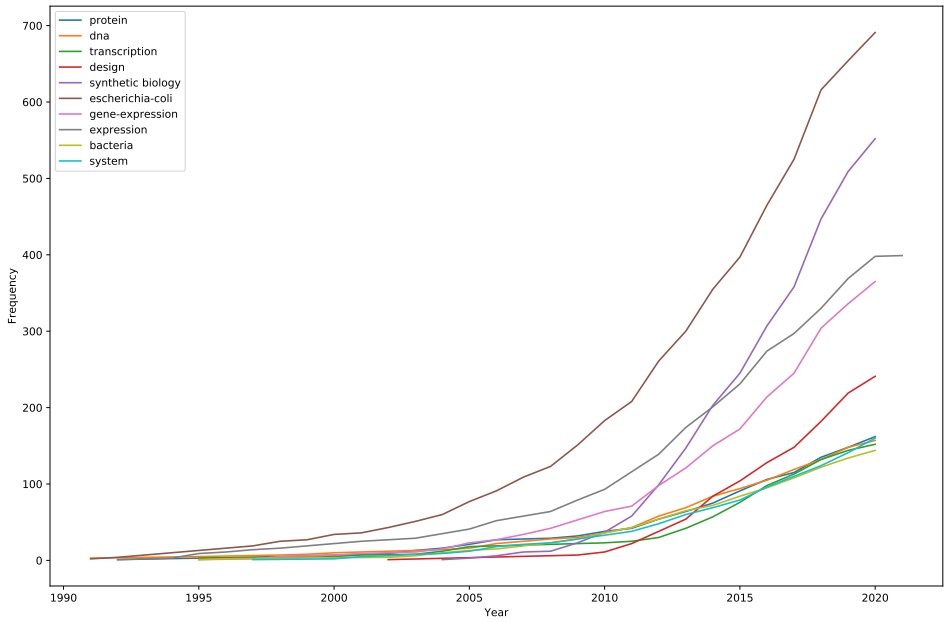

Figure 10: Frequency timeseries of top keywords of articles.

In this section we do not recommend following any particular keywords as the top-ranked keyswords tend to be more general terms from fields of study that genetic circuit design are influenced by, including terms from molecular biology. However, we do recommend examining such a list of terms to explore what terms are related to genetic circuit design and watch out for these terms in papers to understand what role these concepts play in the field.

3.5 Authors → Keywords Associations

Given that each article has a collection of authors and a collection of keywords, we used an association rule learning approach to look for associations between researchers and keywords. The interpretation of this would be to address what topics a given researcher tends to study. FP-Growth is an algorithm for frequent pattern mining[32], and those frequent patterns can be used to calculate association rules. We FP-Growth and other utility functions in the MLxtend package to extract frequent association rules that give the keywords associated with the works of a given author.[33] These rules are visualized in Figure 11 for the genetic circuit design literature.

Interestingly, none of the top recommended authors appear in Figure 11. One reason that this might be is that authors that have published many papers have greater variation in the keywords they've used, making their particular author-name-to-keyword pairings less supported than authors that are either specialized or have fewer publications.

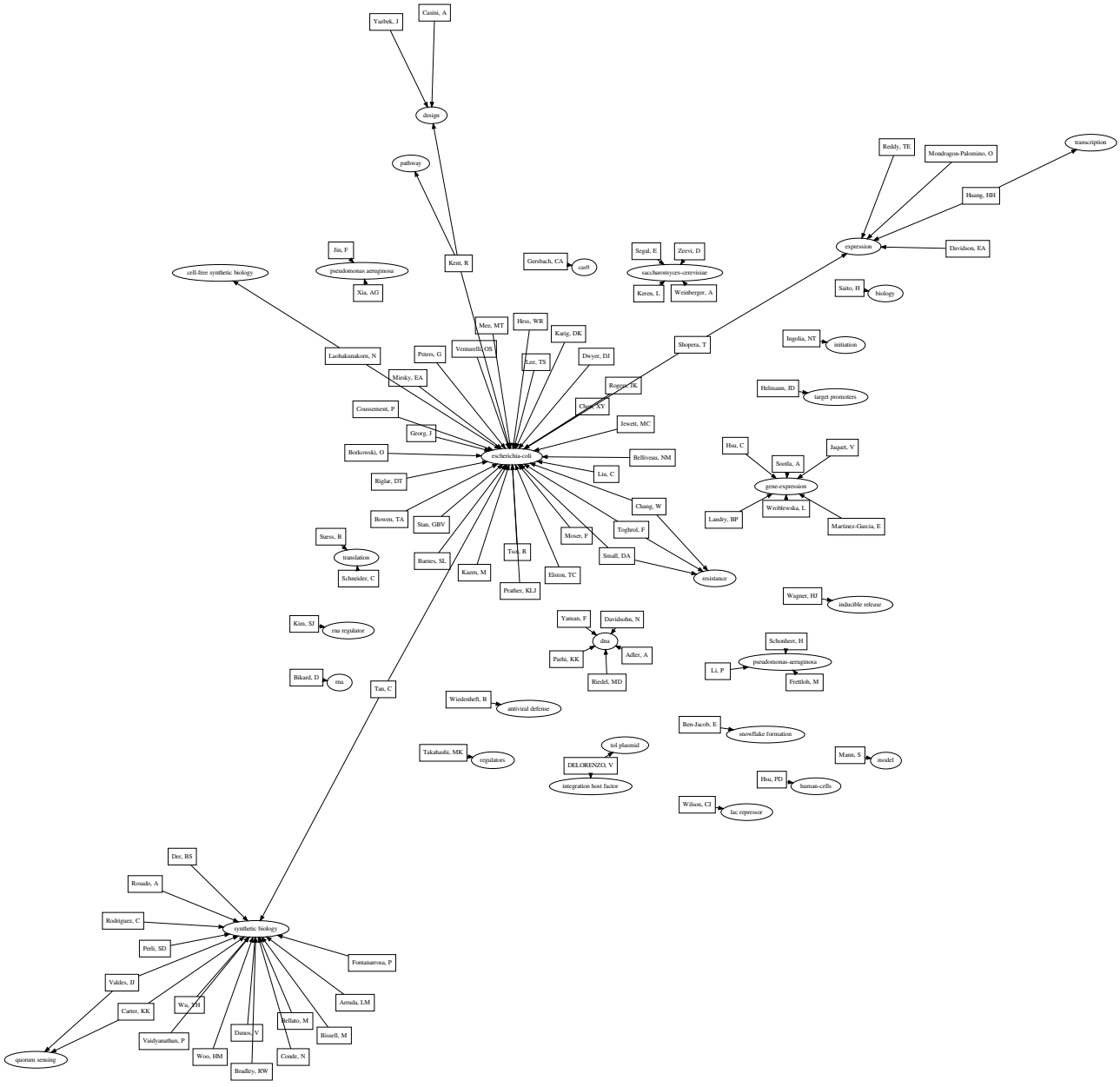

Figure 11: Directed graph illustrated association rules whose antecendent are authors and whose consequent are keywords.

4 Challenges

This section briefly summarizes some challenges when trying to complete this project.

The first challenge was getting the webscraper to work. There ended up being a variety of corner cases pertaining to the particular structure of the HTML. While I developed the script until it seemed to obtain all relevant entries from the database, I have doubts that it will remain stable as Web of Science changes its HTML structure over time.

The second challenge encountered was the performing the Fisher's Exact tests, which turned out to involve more combinations of authors and keywords than what my laptop would be able to process. I turned to association rule mining as a substitute to still obtain associations between authors and keywords. Performing a similar analysis between maximal cliques and keywords was similarly prohibitive as their Cartesian product was of a similar order of magnitude size as with the product between authors and keywords.

A third challenge was attempting an alternative of Fisher's Exact test that would give more rich information about the associations between authors and keywords. This approach was to one-hot encode the authors and keywords into sparse matrices, which would have the same number of rows corresponding to the number of articles, and perform canonical correlation analysis between the two matrices. Canonical correlation analysis attempts to find linear combinations of the original variables through multidimensional rotation such that the transformed variables are maximally correlated. This would give loading scores of how stronly associated each author was to each keyword. Unfortunately it was taking too long to complete, so I halted the calculations despite the preferable results that they might produce.

I had a similar idea to the maximal clique analysis, except that it looked at the maximal paths. I coded a brute force implementation of it, but since finding all the maximal path problems is a generalization of a problem that was already NP-complete, the progress was very slow. If I had obtained the maximal paths in the citation graph, I could have looked for whether which authors tended to occur in these paths as evidence of their importance to the field.

Despite these challenges, I made an effort to accomplish the more general goals of the project using other techniques.

5 Conclusion

In this study we produced a list of recommended studies and authors to examine more closely, and showed how keyword frequency gives hints at the common themes found within the literature of genetic circuit design. This approach has limitations, including validity of measurement, choice of initial seed articles for searching the database, maintainability of webscraper code, and choice of closure keywords. But, this project serves as a proof of concept that an automated approach can be used to create recommendations for starting a more traditional literature search or review.

References

[1] Timothy S. Gardner, Charles R. Cantor, and James J. Collins. Construction of a genetic toggle switch in escherichia coli. *Nature*, 403(6767):339–342, January 2000.

[2] Michael B. Elowitz and Stanislas Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, January 2000.

[3] Piro Siuti, John Yazbek, and Timothy K Lu. Synthetic circuits integrating logic and memory in living cells. *Nature Biotechnology*, 31(5):448–452, February 2013.

[4] Francisco J. Navarro and David C. Baulcombe. miRNA-mediated regulation of synthetic gene circuits in the green alga chlamydomonas reinhardtii. *ACS Synthetic Biology*, 8(2):358–370, January 2019.

[5] Lior Nissim, Samuel D. Perli, Alexandra Fridkin, Pablo Perez-Pinera, and Timothy K. Lu. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/cas toolkit in human cells. *Molecular Cell*, 54(4):698–710, May 2014.

[6] Dae-Kyun Ro, Eric M. Paradise, Mario Ouellet, Karl J. Fisher, Karyn L. Newman, John M. Ndungu, Kimberly A. Ho, Rachel A. Eachus, Timothy S. Ham, James Kirby, Michelle C. Y. Chang, Sydnor T. Withers, Yoichiro Shiba, Richmond Sarpong, and Jay D. Keasling. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, 440(7086):940–943, April 2006.

[7] J.L. Fortman, Swapnil Chhabra, Aindrila Mukhopadhyay, Howard Chou, Taek Soon Lee, Eric Steen, and Jay D. Keasling. Biofuel alternatives to ethanol: pumping the microbial well. *Trends in Biotechnology*, 26(7):375–381, July 2008.

[8] Pejman Mohammadi, Niko Beerenwinkel, and Yaakov Benenson. Automated design of synthetic cell classifier circuits using a two-step optimization strategy. *Cell Systems*, 4(2):207–218.e14, February 2017.

[9] Behzad D. Karkaria, Neythen J. Treloar, Chris P. Barnes, and Alex J. H. Fedorec. From microbial communities to distributed computing systems. *Frontiers in Bioengineering and Biotechnology*, 8, July 2020.

[10] Seunghee S. Jang, Kevin T. Oishi, Robert G. Egbert, and Eric Klavins. Specification and simulation of synthetic multicelled behaviors. *ACS Synthetic Biology*, 1(8):365–374, July 2012.

[11] Martín Gutiérrez, Paula Gregorio-Godoy, Guillermo Pérez del Pulgar, Luis E. Muñoz, Sandra Sáez, and Alfonso Rodríguez-Patón. A new improved and extended version of the multicell bacterial simulatorgro. *ACS Synthetic Biology*, 6(8):1496–1508, May 2017.

[12] Angel Goñi-Moreno and Martyn Amos. DiSCUS: A simulation platform for conjugation computing. In *Unconventional Computation and Natural Computation*, pages 181–191. Springer International Publishing, 2015.

[13] Antonio Prestes García and Alfonso Rodríguez-Patón. BactoSim – an individual-based simulation environment for bacterial conjugation. In *Advances in Practical Applications of Agents, Multi-Agent Systems, and Sustainability: The PAAMS Collection*, pages 275–279. Springer International Publishing, 2015.

[14] A. A. K. Nielsen, B. S. Der, J. Shin, P. Vaidyanathan, V. Paralanov, E. A. Strychalski, D. Ross, D. Densmore, and C. A. Voigt. Genetic circuit design automation. *Science*, 352(6281):aac7341–aac7341, March 2016.

[15] Leonard Euler. Solutio problematis ad geometriam situs pertinentis. pages 128–40, 1736.

[16] M. E. J. Newman. *Networks : an introduction*. Oxford University Press, Oxford New York, 2010.

[17] Mahmoud Saleh, Yusef Esa, and Ahmed Mohamed. Applications of complex network analysis in electric power systems. *Energies*, 11(6):1381, May 2018.

[18] Iman Habibi, Effat S Emamian, and Ali Abdi. Quantitative analysis of intracellular communication and signaling errors in signaling networks. *BMC Systems Biology*, 8(1), August 2014.

[19] Danielle M. Varda, Rich Forgette, David Banks, and Noshir Contractor. Social network methodology in the study of disasters: Issues and insights prompted by post-katrina research. *Population Research and Policy Review*, 28(1):11–29, November 2008.

[20] Eugene Garfield. *Citation indexing - its theory and application in science, technology, and humanities*. Wiley, New York, 1979.

[21] Jaime A. Teixeira da Silva, Panagiotis Tsigaris, and Mohammadamin Erfanmanesh. Publishing volumes in major databases related to covid-19. *Scientometrics*, August 2020.

[22] M. Ali Al-Radhawi, Anh Phong Tran, Elizabeth A. Ernst, Tianchi Chen, Christopher A. Voigt, and Eduardo D. Sontag. Distributed implementation of boolean functions by transcriptional synthetic circuits. *ACS Synthetic Biology*, 9(8):2172–2187, June 2020.

[23] Govind Menon and J. Krishnan. Design principles for compartmentalization and spatial organization of synthetic genetic circuits. *ACS Synthetic Biology*, 8(7):1601–1619, June 2019.

[24] Goksel Misirli, Jennifer Hallinan, and Anil Wipat. Composable modular models for synthetic biology. *ACM Journal on Emerging Technologies in Computing Systems*, 11(3):1–19, December 2014.

[25] Yiyu Xiang, Neil Dalchau, and Baojun Wang. Scaling up genetic circuit design for cellular computing: advances and prospects. *Natural Computing*, 17(4):833–853, October 2018.

[26] Jennifer A N Brophy and Christopher A Voigt. Principles of genetic circuit design. *Nature Methods*, 11(5):508–520, April 2014.

[27] Peng-Fei Xia, Hua Ling, Jee Loon Foo, and Matthew Wook Chang. Synthetic genetic circuits for programmable biological functionalities. *Biotechnology Advances*, 37(6):107393, November 2019.

[28] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. Exploring network structure, dynamics, and function using networkx. In Gaël Varoquaux, Travis Vaught, and Jarrod Millman, editors, *Proceedings of the 7th Python in Science Conference*, pages 11 – 15, Pasadena, CA USA, 2008.

[29] D. T. Carrell and M. Simoni. 'easier ways to get a publication': the problem of low quality scientific publications. *Andrology*, 6(1):1–2, December 2017.

[30] Bhushan Patwardhan. Good publications need good research. *Journal of Ayurveda and Integrative Medicine*, 6(2):73, 2015.

[31] Ulf Sandström and Peter van den Besselaar. Quantity and/or quality? the importance of publishing many papers. *PLOS ONE*, 11(11):e0166149, November 2016.

[32] Jiawei Han, Jian Pei, Yiwen Yin, and Runying Mao. Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data Mining and Knowledge Discovery*, 8(1):53–87, January 2004.

[33] Sebastian Raschka. Mlxtend: Providing machine learning and data science utilities and extensions to python's scientific computing stack. *The Journal of Open Source Software*, 3(24), April 2018.