Mapping and visualization of publication networks of public university faculty in computer science and electrical engineering

Conor Carroll, Nupur Garg, Theresa Migler, Barbara Walker, and Zoë Wood

[1] California Polytechnic State University San Luis Obispo, CA ccarro07, zwood@calpoly.edu
[2] UC Santa Barbara, CA barbara.walker@ucsb.edu

Abstract

We present our process and development of a web-based system to explore the publication networks of faculty in California public universities in the fields of computer science and electrical engineering. Our project explores collaboration networks in the fields of computer science and electrical engineering with a focus on publication networks and an analysis of these collaborations with a focus on geospatial organization (which institutions are collaborating with which other institutions). We present our data gathering process, which relies on the Scopus[9] database (Scopus represents a "comprehensive overview of the world's scientific research output across all disciplines"), and we present the development of a web-based tool using python and the Google Maps API [12] in order to allow visualizations and explorations of geospatial structures of the publications networks. These visualizations drove further network analysis, which we present here as well.

1 Introduction

To support scientific innovation, team sizes can reach into the thousands (consider CERN for example, where some high energy physics research teams are as large as 2000 researchers [1]). Figuring out how to support successful large scientific and engineering teams is in all of our interest. The importance of this is emphasized by work by the National Research Council on "Enhancing the Effectiveness of Team Science" [6]. When considering how to move forward as a community it is useful to examine past collaboration communities and networks in general to understand various communities and network measures [16, 3, 5, 15]. Collaboration networks can reveal under laying structures and reveal inequities or system trends [10, 2] and potentially point us towards important work for supporting inclusive and diverse teams.

This project specifically explores collaboration networks of faculty from the University of California system in the fields of computer science and electrical engineering as shown via their publication networks. We present our initial observations and analysis of these collaborations with a focus on geospatial organization of the network across the globe. We present our data gathering process, which relies on the SCOPUS [9] database (Scopus represents a "comprehensive overview of the world's scientific research output across all disciplines"), and we present the development of a web-based tool using python and the Google Maps API [12] in order to

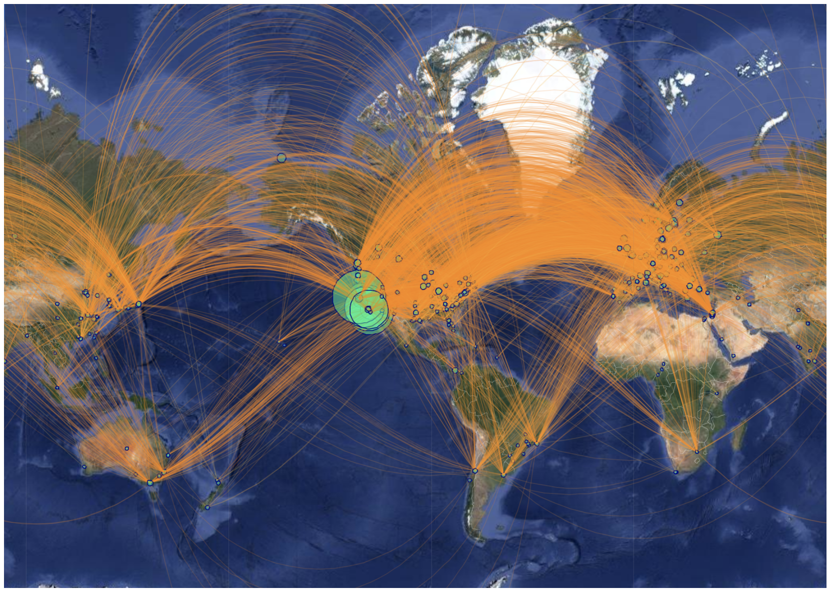

Figure 1: An overview image of the entire publication network visualized using Google Maps API.

allow visualizations of the publications networks. These visualizations drove further network analysis, which we present here as well. Ideally, this work will help develop a tool we can use to identify and classify discrepancies in collaboration networks, with future work aiming to discover social trends based off recurring patterns in collaborative communities.

Starting from a list of faculty members in University of California computer science and electrical engineering departments the Scopus database was used to aggregate all publication data up to the point of data collection (1967-2017). These fields are of interest in terms of both their challenges with respect to diversity [18, 14] and their newness (many UCs only formed computer science departments in the last 40 years). The exact data acquisition protocol is described in detail below, which resulted in over 400,000 records. Once this data was acquired, we chose to represent it as a graph with a geospatial embedding to explore community structure and trends. Each node in the graph represents the primary institution of any author or co-author. Edges between nodes represent a co-authored publication. This publication network graph was then visualized using the GoogleMaps API. See Figure 1 for an example map. We then computed various metrics for this network to begin to analyze its general structure.

2 Related Works

An overview of Social Network graphs analysis is described by Campbell [3] and an overview of community metrics is discussed in Chakraborty [5]. Likewise there is a large body of work related to scholarly networks generated via publications [20, 8, 13]. This current project does not address citation networks but co-authorship networks, thus the work of Liu [13] is more relevant. We are overall, less interested in rank, then revealing structure and geospatial trends of interest.

We are also interested in exploring the community nature of the co-author network. Communities have been examined via networks in general [16, 3, 15]. Of particular interest is that collaboration networks can reveal under laying structures and reveal inequities or system trends [10, 2] and potentially point us towards important work for supporting inclusive and diverse teams. Our work also considers the importance of various institutions via a DeGroot analysis [11, 7, 17].

3 Data Acquisition

The first stage of this project was the compilation of publication data for the authors of interest. The technical side of the project was completed in Python due to the simplicity of the language and the multitude of tools available. SQLite was the chosen database management system due to its simplicity and efficiency. Being a file-based database made it easily portable. Additionally, the project had no need for user management or high performance. Lastly, the object relational model (ORM), Peewee was used to interface between Python and SQLite. As with the previous technologies it was lightweight and easy to use.

Collecting Data: The first approach to collecting data was data mining Google Scholar. A sample set of faculty from 3 institutions was used, aggregated from the department websites. This approach presented problems, mainly that out of the 360 scholars queried, only around 25% of them had Google Scholar pages - leading to a limited size of the data. Lastly, after further investigation we discovered Google Scholar allows self-citations when aggregating their citations.

After our initial tests, we decided to change approaches to collecting data using the Scopus API. "Scopus indexes content from 24,600 active titles and 5,000 publishers which is rigorously vetted and selected by an independent review board" [9]. A collaborator provided a list of 921 scholars who were affiliated with schools within the UC system, and basic information about the scholars.

To build the database to represent scholars networks, we accessed four pages within the Scopus API: search, author, affiliation id, and scopus id. We searched for matches on the author's name, found the associated author id,where the scholar's campus matched the campus provided in the original scholars list. Given the author id, publications were collected from the author page. The affiliation id page searched for details for a given affiliation id. The scopus id page searched for metadata regarding a publication. Using the Scopus API we were able to collect information for 626 out of 914 scholars (68%).

3.1 Databases

The aggregated information was stored in five databases: Affiliation, Scholar, ScholarCitation, Publication, and ScholarXPublication.

- Affiliation - The Affiliation database contains information collected from the affiliation id page in the Scopus API. This includes name, country, country code, city, and postal code.
- Scholar - Scholar represents a scholar who has published papers.
- ScholarCitation - ScholarCitation contains citations for a given scholar aggregated by year.
- Publication - Publication contains citation information for a given publication. This includes title, publisher, number of citations, year, and number of authors.

- ScholarXPublication - ScholarXPublication contains the relationship between scholars and citations. This database enabled the least amount of information to be duplicated information.

Database Assumptions In our first pass through the data, many scholars had co-author papers with scholars who were not part of the provided list of scholars. In other words, many UC faculty collaborate with scholars outside the UC system. Outside collaborators also appeared as co-authors on multiple publications. These scholars needed to be added to the database. To prevent duplicates of the same outside-UC scholar, we made some assumptions regarding what identified a unique scholar. For scholars recorded from Scopus, we stored the scopus id and used that to identify a unique scholar. If source is not Scopus and there is a common paper we matched the first letter of first name and last full name. We determined publications to be the same if the title of the publication was the same. For publications that were published more than once, we took the publication that was produced earlier.

After the SQLite database was completed, it was converted into a CSV file for ease of use for statistical analysis.

4 Visualizing the network

Once all the information had been acquired, we parsed the CSV file and built a graph to represent the data. This data was explored by various statistical measures, including validating Freemans work [10] that teams of more diverse co-authors have higher citation counts [19, 4]. This current project focuses on the development of a web-based geospatial embedding to explore community structure and trends. Each vertex in the graph represents the primary institution of any author or co-author. Edges between vertices represent a co-authored publication. The graph is weighted and allows self-loops, where the weights represent the number of published papers between institutions. This publication network graph was then visualized using the GoogleMaps API. Weights and self-loops are not drawn (to reduce drawing complexity, however, we do consider this information when conducting all analysis.

All construction and analysis of the graph was done in Python, using NumPy for mathematical computations. To draw the graph with the GoogleMaps API, we wrote a Python package that wrapped the API and generated the appropriate html files. The GoogleMapsAPI does not directly support drawing networks, so we constructed the graph by individually drawing circles and connecting them with polylines. Latitudes and longitudes were precalculated to relieve strain on the browser.

The vertices are drawn with respect to their institute's geographic location, and track three attributes: the name of the institution, the number of publications, and the number of institutions they have collaborated with, labeled degree. The vertex sizes are initialized linearly with respect to their publication count, and are resized as the user zooms in and out of the map. Additionally, the edges decrease in opacity as the map zoom increases to allow users see geographic details.

We note that many institutions tend to faithfully publish with authors of the same institution. To account for this, we size institution nodes based on their publications field, but also include the field degree to represent the number of edges leaving an institution. To account for the sheer size of the data, we do not draw multiple edges, so if multiple publications have been authored between the same universities, this is represented with only one edge, however, each unique publication is accounted for in the degree field listed per institution. While these maps are interactive, the size of the data greatly taxes the browser, so various methods to reveal the general structure with fewer edges were explored. All maps can be accessed here: Maps: http://users.csc.calpoly.edu/∼ccarro07/Maps.html

5 Analysis and Observations

The final network consists of 602 vertices (institutions), 19,223 publications and 1,571,966 edges (co-authorship between two different institutions).

We quickly see imbalances in the network when looking at the degree distribution. Figure 2 emphasizes how a sample of the population lies several standard deviations above the mean. Observing the distribution as a cumulative density function further highlights this; the first 80% of vertices have degree below $\sim$ 40000, which corresponds to an average of 28723.887, but we have a high standard deviation of 74768.137. This indicates that certain institutions collaborate with many institutions. These could represent particular projects (such as those like CERN) which result in very large teams.

We also see a disparity when looking at authors and their publication count, figure 3. We would expect the trend to be decreasing as the number of publications increases, but we see a spike at 85 publications. This occurrence seems to contradict intuition, and could be indicative of a community in our graph, where many authors are frequently collaborating.

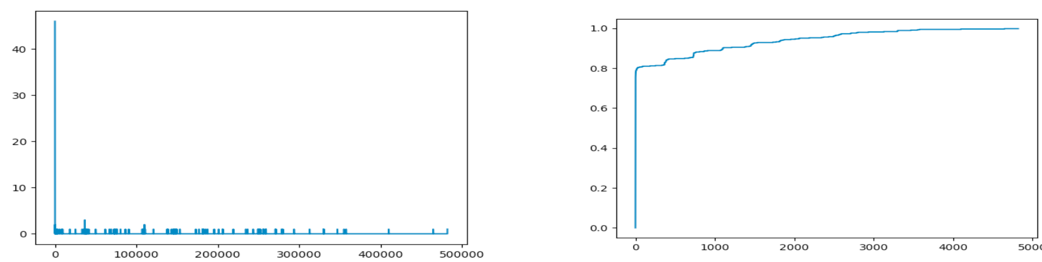

Figure 2: Left: Degree distribution of our data, plotting degree by frequency at which that degree appears in the graph. Right: Cumulative probability function representing the distribution. Plots degree by percentage of vertices that have that degree or less.

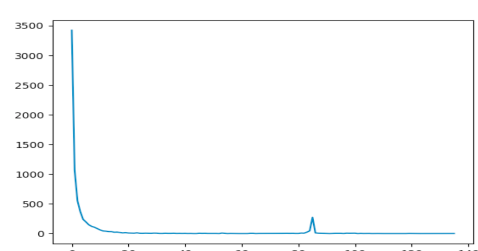

Figure 3: Histogram monitoring authors' publications. We plot the number of publications (*x-axis*) by the number of authors who have published that amount (*y-axis*). *Average publication number:* 21.617. *Standard deviation:* 9.131.

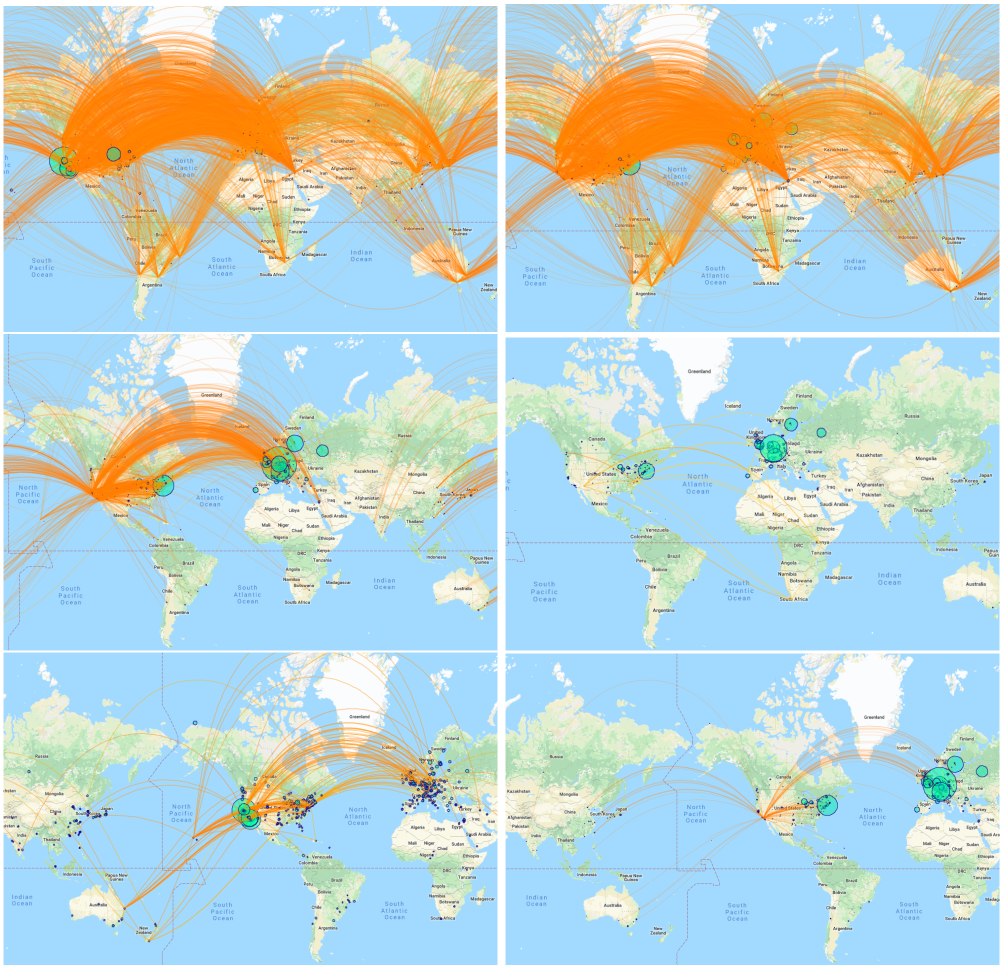

Figure 4: A two-step neighborhood centered at various institutions in the order they appear in the Perron vector. In order of appearance from top right to lower left:University of Bonn (Rank 1), Tokyo Metropolitan University (Rank 100), University of Hawaii at Manoa (Rank 200), Center of Disease Control and Prevention (Rank 300), University of Otago (Rank 302), and Sookmyung Women's University (Rank 400).

To explore the geospatial nature of the network, we considered two-step neighborhoods about an institution. In doing this, we target a particular institution, and draw an edge between institutions if: (a) one of the institutions is the target one, or (b) if either institution is a neighbor of our target institution. The two step neighborhoods are effective in decreasing the overall amount of data drawn, and highlighting the relative connectivity that a vertex has to the graph as a whole. With a graph as large and connected as our network, it becomes difficult to observe details other than general trends. Focusing on a vertex (or set of vertices) and it's local connections can help discern these details; in our two-step map of the University of Otago (Figure 4) we are able to see the prevalence of connection to Europe, and lack of connection to continental Asia. In the larger graph, there is no easy way to make out these details. Furthermore, by observing the connections that Australia has to East Asia, and the United States' east coast, and that the University of Otago does not have these connections, we are able to deduce that the appearance of Australian institutions in our two-step neighborhoods comes from mutual collaborators, not direct collaboration.

We next explored viewing the map using random walks. This process involves specifying a start location and a number of steps, and allowing the "walker" to wander until it uses up it's step count. We determine how the walker moves by randomly picking an edge to traverse based at whichever institution the walker is currently located. While random walks provide a effective way to shrink the graph for visualization purposes, it is very dependent on initial position and number of allowable steps, making it difficult to know how representative a walk is of the whole graph.

Examining the asymptotics of random walks allows us to rank the vertices. The Perron vector [17] of a normalized transition matrix encodes the limiting behavior of the system, and tells us the likelihood that a random walker will be at a given vertex. This allows us to rank the vertices by increasing probability. When applied to our simplified graph (no weights, no loops), the Perron vector ranks the institutions based off of their degree. However, when we do include weights, i.e. an edge has a stronger weight if multiple papers have been published between the institutions, we observe some very interesting behavior. See table 1. For example, even though the network was generated starting from University of California scholars, the first UC institution in the Perron vector does not appear until rank 28. Likewise, from the visualizations, we noted that there is strong trend for scholars to work with European collaborators, this visual effect is reflected likewise in the Perron vector as 9 of the top 10 institutions are European.

Figure 5: Left: Vertices sized based off publication count. Right: Vertices sized based off asymptotic ranking.

Knowing we have a skewed degree distribution, and that the data was collected from the UC system, we can hypothesize that our model will favor California schools, and we see this when we size vertices based off their publication number. It seems surprising then, that we do not see a better representation of California when ranking the universities based off willingness to collaborate, table 1. We see European institutions dominating the top 10 spots, and do not see a UC until rank 28: UC Berkeley. In fact, there are a couple of institutions that dominate the visualization in our publication-based model, such as UCLA and UCSD, however, they do not appear in the top 100 of the ranking, table 2.

Ranking	Institution	Probability
1	Universitat Bonn	0.0306923431
2	Brookhaven National Laboratory	0.0295787876
3	Universitat Freiburg im Breisgau	0.0261025543
4	Bergische Universitat Wuppertal	0.0227725802
5	Stockholms universitet	0.0226096457
6	University of Birmingham	0.0220857731
7	Joint Institute for Nuclear Research: Dubna	0.0210338908
8	University of Oxford	0.0210199522
9	Universite de Geneve	0.0210011129
10	University of Amsterdam	0.0199134614
11	University of Wisconsin Madison	0.0186969262
12	LIP - Lisboa	0.0178291726
13	Johannes Gutenberg Universitat Mainz	0.0177259383
14	University of Chicago	0.0172897702
15	Centre de Physique des Particules de Marseille	0.0172389805
16	Universitat Gottingen	0.0165046932
17	University of Liverpool	0.0164694968
18	University of Manchester	0.0163378763
19	Universite Blaise Pascal	0.0162814857
20	Horia Hulubei National Institute of Physics and Nuclear Engineering	0.016064961
21	Michigan State University	0.0160088249
22	University of Glasgow	0.0159015173
23	Lancaster University	0.0158749131
24	University of Pennsylvania	0.015502455
25	University of Toronto	0.0150664779
26	Yale University	0.014924674
27	University of Cambridge	0.0139503765
28	UC Berkeley	0.013138759
29	Tel Aviv University	0.0131120912
30	Ludwig-Maximilians-Universitat Munchen	0.0128072254
31	Rutherford Appleton Laboratory	0.012454625
32	University Michigan Ann Arbor	0.0119350166
33	McGill University	0.0117913033
34	Humboldt-Universitat zu Berlin	0.0116460625
35	Ceske vysoke uceni technicke v Praze	0.0116407798
36	Universitat Heidelberg	0.0115602036
37	Columbia University in the City of New York	0.0115393913
38	University of Tokyo	0.0112685763
39	University of Illinois	0.0110215014
40	Duke University	0.0097722519
41	Niels Bohr Institute	0.0095266409
42	Harvard University	0.0094456191
43	Technische Universitat Dresden	0.009421179
44	University of Pittsburgh	0.0093343654
45	National Technical University of Athens	0.0093338563
46	University of Texas at Arlington	0.0092809025
47	Universitetet i Oslo	0.0092544257
48	UCL	0.0092530891
49	UC Irvine	0.0092483156
50	University of Ljubljana	0.0091216595

Table 1: First 50 entries of the Perron vector ranking the institutions based off likelihood of collaboration.

Ranking	Institution	Probability
28	UC Berkeley	0.01313876
49	UC Irvine	0.00924832
108	University of California: San Diego	0.00020551
109	University of California: Los Angeles	0.00016688
110	University of California: Santa Barbara	0.0001476
115	University of California: Riverside	6.95e-05
123	University of California: Santa Cruz	2.094e-05
399	UC MERCED	3.2e-07

Table 2: University of California system ranked.

The initial ranking achieved with the Perron vector seems to correspond to universities that were historically very active in the 1900s. As computer science and electrical engineering departments are young fields, we decided to split the graph based off publications before and after 1995, with hopes to highlight any differences in ranking. Interestingly, we see no changes in ordering or probabilities for the graph only including publications 1995-2017. However, the ranking based on 1967-1995 prioritizes California institutions, and the University of Bonn, ranked 1 in the full data-set, does not appear in the top 50.

Experimentation with more date thresholds yielded similar results, with the original ranking seeing only minor changes in ordering and probability. This indicates the presence of high impact publications that stabilize the topology of the graph.

An initial histogram does not provide much insight as the graph scales are dominated by the high number of papers that have few authors. However, ignoring all "small" papers (all papers with 50 authors or less), we see a subset of publications each have 400 authors or more. With our model, each one of these publications contributes over 80000 edges to the graph, and would explain the invariant behavior of the ranking. Pinpointing these papers show that they tend to be physics-based and share many of the same collaborators. Choosing to exclude these papers and only analyze publications from teams of 200 people or less showed drastic changes in the ranking.

Ranking	Institution	Probability
1	UC Berkeley	0.08713082988106795
2	University of California: San Diego	0.07191857098310861
3	Lawrence Berkeley National Laboratory	0.07062675397567741
4	University of California: Los Angeles	0.0583990378190574
5	University of California: Santa Barbara	0.051650407590539765
6	UC Irvine	0.045770412935992974
7	Broad Institute	0.041494053187222886
8	UC Davis	0.029600427635976895
9	University of California: Riverside	0.024321796071094904
10	Stanford University	0.02011225444340302
11	University of Wisconsin Madison	0.01812998351819812
12	Uniformed Services University of the Health Sciences	0.013452715042986486
13	Universita degli Studi dell'Aquila	0.012962715488440404
14	Universita degli Studi di Milano	0.012962715488440392
15	Universita degli Studi di Padova	0.010401354180586984
16	University Michigan Ann Arbor	0.009198628001248326
17	University of California System	0.009198628001248144
18	Massachusetts Institute of Technology	0.007995901821907883
19	University of California: Santa Cruz	0.007327720611163956
20	Fermi National Accelerator Laboratory	0.007149538954964621
21	Harvard Medical School	0.005969085482649674

22	University of Oxford	0.005946812775624283
23	University of Maryland	0.005879994654550843
24	Tsinghua University	0.005479085928105664
25	McGill University	0.005434540514053123
26	Purdue University	0.005412267807028206
27	IBM Thomas J. Watson Research Center	0.005323176978929842
28	University of Southern California	0.005144995322732104
29	Uniwersytet Wroclawski	0.005055904494632281
30	Lawrence Livermore National Laboratory	0.004677268475211254
31	Baylor College of Medicine	0.004654995768185594
32	Scripps Institution of Oceanography	0.004632723061160425
33	Pennsylvania State University	0.004476814111987638
34	Universitat Dortmund	0.004276359748764512
35	Ben-Gurion University of the Negev	0.004031359971490283
36	Carnegie Mellon University	0.003942269143390591
37	National Nano Device Laboratories Taiwan	0.003764087487193693
38	University of Arizona	0.00374181478016928
39	Stockholms universitet	0.0036749966590940222
40	Ecole Polytechnique Federale de Lausanne	0.0036749966590938214
41	Iowa State University	0.0034745422958708974
42	Universita degli Studi di Napoli Federico II	0.0034077241747961133
43	Brown University	0.0033631787607461567
44	Johannes Gutenberg Universitat Mainz	0.0033409060537218822
45	Uniwersytet Warszawski	0.0032963606396721754
46	University of Pennsylvania	0.003184997104547125
47	University of California: San Francisco	0.0031181789834732712
48	Delft University of Technology	0.0030959062764495953
49	Universiteit Gent	0.0030959062764490427
50	IEEE	0.003073633569423864

Table 3: Top 50 universities ranked after omitting high-collab publications.

We see exactly the bias towards California that we would expect from our data set; 7 of the top 9 institutions are associated with the University of California. By plotting the graphs and only drawing edges that appear between the top percentiles of institutions in the ranking, we can get a sense of the asymptotic community structure of the graph. Figure 6 shows two graphs that use two different ranking; one ranking includes the set of high-impact papers, and the other does not. By excluding the high-impact papers, we immediately see the graph refocus its attention onto California. However, there is still a strong European influence drawing connections to Italy.

Additionally, looking at the two-step neighborhoods of institutions of increasing rank further shows the weight that the ranking (high-impact papers included) puts on North America and Europe. Figure 4 depicts exactly this, as the rank of the institution gets worse, we see the edges within the two-step neighborhood shift from connecting North America and Europe, to connecting regions with Eastern Asia and Africa.

6 Conclusion and Future Work

We have presented our exploration of a co-author network, including data acquisition, visualization using the Google Maps API and graph analysis. This data set has allowed us to explore collaboration trends in the newly arising disciplines of computing and electrical engineering. Overall, starting with UC schools, our visualizations show a strong tendency for collaboration with European institutions. Further analysis shows that certain large projects with thousands of collaborators skews the network.

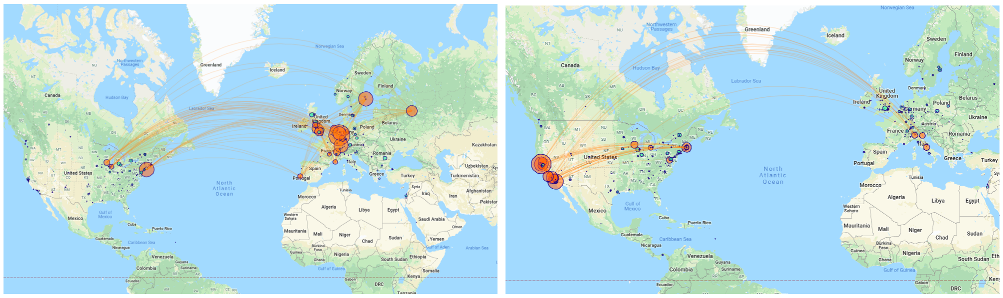

Figure 6: On the left: Top 20 institutions based off ranking with high-impact papers. On the right: Top 20 institutions based off ranking that omits high-impact papers.

Moreover, our ranking system provides a method to understand regions of high activity and connectivity while respecting the underlying graph topology. With such a high number of edges, deleting all edges associated with high-impact publications makes no noticeable changes to the graph, but is reflected in the ranking. This instability seems to reflect a disparity in collaboration; we would not expect a well-rounded, collaborative community to drastically change its statistical behavior by the deletion of a couple publications. This indicates that we are looking at a statistically significant event.

Future work includes additional exploration of collaboration networks. Of particular interest is to asses how scholars identity in terms of gender and ethnicity contribute to the geospatial nature of their collaborators. For example, do more diverse teams display more or less geospatial spread or distribution?

7 Acknowledgments

We would like to thank the Center for Research, Excellence, and Diversity in Team Science (CREDITS) for their input.

References

[1] Jeremy Birnholtz. When authorship isn't enough: Lessons from cern on the implications of formal and informal credit attribution mechanisms in collaborative research, 2008. the journal of electronic publishing.

[2] LG Campbell, S. Mehtani, ME Dozier, and J. Rinehart. Gender-heterogeneous working groups produce higher quality science. PLoS One, 2013.

[3] William M. Campbell, Charlie K. Dagli, and Clifford J. Weinstein. Social network analysis with content and graphs, 2013. Lincoln Laboratory Journal.

[4] Eduardo Cardenas-Torres and Michael Nava. Does author diversity affect uc journal publications?, 2018. "Student research presentation for UCSB McNair Scholars Program".

[5] Tanmoy Chakraborty, Ayushi Dalmia, Animesh Mukherjee, and Niloy Ganguly. Metrics for community analysis: A survey. *ACM Comput. Surv.*, 50(4):54:1–54:37, August 2017.

[6] National Research Council. Enhancing the effectiveness of team science, 2015. National Academies Press.

[7] Morris H. Degroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.

[8] Y. Ding. Applying weighted pagerank to author citation network. *Journal of the American Society for Information Science and Technology*, 2011.

[9] Elsevier. Scopus. "URL: https://www.elsevier.com/solutions/scopus".

[10] Richard B Freeman and Wei Huang. Collaborating with people like me: Ethnic co-authorship within the us. Working Paper 19905, National Bureau of Economic Research, February 2014.

[11] Benjamin Golub and Matthew O. Jackson. Naïve learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1):112–49, February 2010.

[12] Google. Google maps api. "URL: https://developers.google.com/maps/documentation/javascript/reference".

[13] X.M. Liu, J. Bollen, M.L. Nelson, and H. V. Sompel. Co-authorship networks in the digital library research community. *Information Processing and Management*, 2005.

[14] Joyce B Main and Corey Schimpf. The underrepresentation of women in computing fields: A synthesis of literature using a life course perspective. *IEEE Transactions on Education*, 60(4):296–304, 2017.

[15] M. E. J. Newman. Detecting community structure in networks. *The European Physical Journal B*, 38(2):321–330, Mar 2004.

[16] Arun Pandey, Roshni Chakraborty, Soumya Sarkar, and Joydeep Chandra. Analyzing link dynamics in scientific collaboration networks: A social yield based perspective. pages 1395–1402, 08 2015.

[17] Alberto Peretti. The importance of perron-frobenius theorem in ranking problems. Technical report, Department of Economics, University of Verona, 2014.

[18] C. Stephenson, A. Derbenwick Miller, C. Alvarado, L. Barker, V. Barr, T. Camp, C. Frieze, C. Lewis, E. Cannon Mindell, L. Limbird, D. Richardson, M. Sahami, E. Villa, H. Walker, and S. Sweben. Retention in computer science undergraduate programs in the u.s.: Data challenges and promising interventions. Technical report, ACM, 2018.

[19] Qi Wang and Michael Nava. Does diversity inputs by homophily index leads to greater contributions for uc authors?, 2018. "Student research presentation for UCSB McNair Scholars Program".

[20] E. Yan and Y. Ding. Discovering author impact: A pagerank perspective. *Information Processing and Management*, 2011.