Data-Driven Assessment Modeling of Partnership Data

M. Seto & K. Medlin

Executive Summary

What does it mean to be in a strong partnership? Using Collaboratory's national dataset of community engagement data, we explored partnerships between higher education institutions and the community organizations with which they are partnered. Our goals were to (1) understand what quantitative characteristics from Collaboratory denote 'strong' community- university partnerships, (2) use those characteristics to create an algorithmic assessment model to identify the strongest partnerships in the Collaboratory dataset, and (3) reveal common themes that practitioners can leverage to cultivate stronger and more resilient partnerships.

With input from Collaboratory administrators, community engagement professionals, and institutional research team members, we identified four quantitative data points in Collaboratory data that we combined into a *partnership strength model*. The model identified 99 out of 2,083 community- university partnerships that might be classified as high-strength. The model's results represent an initial jumping-off point for future research, including qualitative assessment of the 99 strongest partnerships to validate the model. Additionally, we argue that quantitative assessment of qualitative partnerships is by no means a silver bullet, but instead represents a pragmatic method of high-level assessment and quick filtering of large datasets of qualitative partnership data that would otherwise be prohibitively time-consuming.

Bigger Data on Engagement through Collaboratory

Greater numbers of higher education institutions are integrating comprehensive community engagement as part of their strategy to achieve their mission and vision. Collaboratory collects data that help institutions articulate and examine their own engagement practices. For the purposes of this paper, activity refers to any undertaking where a university (or other institution) and a community partner are engaged in mutually beneficial work. Activities are Collaboratory's unit of analysis, consisting of various qualitative and

quantitative data fields reflecting details about who, what, where, when, and why the work is being undertaken. Activities vary in size and scope. Examples of Collaboratory activities are: a multi-year academic research partnership between a university and a local hospital to assess the efficacy of new cancer treatments; hosting a one-time community roundtable for neighborhood associations to discuss collaborative efforts to make safer public spaces; or a local company accepting student-interns for job-based training. Collaboratory is a software product specifically for higher-education institutions to standardize their data

collection for community engagement activities. Collecting this community engagement data allows institutions to begin to speak succinctly about and aggregate information on engagement activity into key metrics that help them better tell their story to those they value most - whether that be funders, collaborators, or the average citizen. But taken a step further, this data also provides a rich repository of information about work that is happening between various stakeholder groups - universities and schools, businesses, nonprofits, and the government sector. Promoting the intentional examination of our own data by scholars and others who seek to unpack processes and outcomes can lead to new insights about what makes partnerships thrive. We continue to increase our capacity and infrastructure for collecting and sharing de-identified interinstitutional data, in support of the democratization of knowledge around partnership characteristics. This paper represents the first in a series of research briefs intended to share the outputs and outcomes of Collaboratory's Research Fellowship Program, which supports exemplary and emerging scholar-practitioners who are committed to advancing the field of community engagement through data.

What do Strong Partnerships have in Common?

Our research set out to identify strong partnerships in the Collaboratory dataset, initially leading to the question 'what is a strong partnership?' Partnerships between universities and community organizations take various forms. Stemming from Clayton, Bringle, et al. (2010), we use the term partnership to refer to interactions that take place between a university and a community partner. Thus, a partnership consists of a series of interactions taking place in one or more activities between a university and their partner. Herein we assume that 'strong' partnerships broadly applies to those which exhibit properties of reciprocity - i.e., partnerships that are situated in the co-production of knowledge where both the institution and community organization are simultaneously

investigators, thought-leaders, and disseminators of new information (Saltmarsh, Hartley, and Clayton, 2009; Dostilio et al, 2012, Carnegie Foundation, n.d.). The concept of reciprocity is difficult to evaluate, as it relies on process, context, and relationships, rather than explicit characteristics that can be easily measured. In parallel to this research on partnership strength, a separate line of inquiry is underway to operationalize characteristics of reciprocity (see Janke et al., 2020).

Anecdotally, 'strong partnerships' are those in which both parties might independently verify "yes, we are close partners". While these partnerships are contemporarily regarded as the pinnacle of engaged scholarship, there remains a lack of understanding into how to successfully identify these types of partnerships. Strong partnerships are often not immediately recognized, but rather identification of a strong partnership is done on a case-by-case basis to find partnerships that satisfy a rubric of, often qualitative, characteristics. For example, the Transformational Relationship Evaluation Scale (TRES) characterizes partnerships according to the degree of closeness, equity, and integrity reported among partners (Clayton et al., 2010).

Much of the research to date on partnership strength and characteristics of partnerships more broadly reflects small sample, qualitative case studies that represent individual and contextualized use cases. In a recent scoping review on community engagement literature, Janke, Florez, and Edwards (2021) identified that 81% of the articles they reviewed were based on an author's examination of specific partnerships and 19% addressed partnerships without referencing any singular partnership. Many studies attempt to identify the origins of partnerships in an attempt to identify root-causes for cultivating these partnerships (Dostilio, 2014; Boyle et al., 2011; Janke, 2008; Sandy & Holland, 2006). In these examples, a single partnership was often identified for deep, contextual analysis in an attempt to uncover the properties of that partnership unit that contributed to its success- they define a

partnership and through contextual analysis attempt to piece together what makes them successful. Our research takes an inverse perspective on such assessment.

We aim to identify ‘strong’ partnerships within a national dataset of community engagement data provided through Collaboratory. We initially began with a large dataset of community partnerships and attempted to model the strongest of those partnerships. Once identified, those partnerships should be qualitatively evaluated to determine if they do indeed meet our criteria to be considered “strong”. Secondly, our research centers around the concept of assessing large datasets of qualitative data using contemporary data analysis techniques. The Collaboratory dataset represents thousands of partnerships and it is impractical to identify strong partnerships by assessing each against a rubric of qualitative criteria. Instead we propose that an algorithmic method can be devised to provide selected partnerships for further investigation, essentially providing a short-list of the data to significantly reduce the time required to find strong partnerships in large datasets.

Algorithmic Approaches to Assessing Engagement

Initially coined in the mid 1990s, the term ‘big data’ has only recently become a buzzword, generally used in reference to the handling and analysis of big datasets (Kitchin, 2014, pg 67). Big data refers to large collections of data that are collected and aggregated quickly, and contain a variety of information to be assessed and analyzed (ibid, pg 68). As technological improvements have accelerated over the last decade, so too has the capacity to record and store vast amounts of data into a singular collection for use in government programming and assessment, private enterprise performance analysis, and virtually all parts of modern, technocratic life. As big data continues to permeate all spaces of contemporary academia, it is imperative that methods be developed to utilize

large-scale data in ways that most effectively identify partnerships suitable for further research.

With a nod to D’Ignazio and Klein (Data Feminism, 2020), algorithmic assessment does not represent a singular solution to big data parsing. Algorithmic bias, variable data ingestion methods, and intersectional power dynamics influence automated data analysis in unseen, and often unintentional ways, in some cases causing unintentional harm to those very populations most at-risk (Automating Inequality, 2017). Furthermore, we ground our assessment against the ‘fourth-paradigm of science’ (Kitchin, 2014, pg. 130), where we acknowledge that data analysis must address the traditional scientific method. We explicitly note that our algorithmic modeling does not allow ‘the data to speak for itself’, but rather that it represents empirical evidence for modeling purposes. Any results developed here must be tested and validated according to conventional theories, so we ground our assessment on the understanding that data analysis must continue to follow the human-to-machine paradigm in which we imagine possibilities in the dataset and then allow the data to confirm or deny it (Prensky, 2001). Therefore, we acknowledge that algorithmic selection of strong partnerships is, in some ways, problematic.

Most importantly we note that the end goal of this research is not to fully replace qualitative assessment of partnerships, but rather to provide an initial pass at algorithmic analysis of large partnership datasets to produce a short-list of partnerships of interest to be further evaluated. Our goal with this research was to draw more objective and generalizable conclusions about key characteristics of partnership strength using a large dataset on the understanding that qualitative data such as this can not be fully assessed with a quantitative model as we develop here. Should the model be proven to identify strong partnerships, the applications of such a modeling process may be widely felt, including potentially automated high-level assessment of community partnership data and initial parsing of large-scale qualitative datasets through an

initial quantitative process.

Building the Model

This research is an initial exploration of community-university partnerships using quantitative methods to analyze the Collaboratory dataset. At the time of this exploration (November 2020), there were 2,439 unique activities in the dataset representing 31 institutions, 3,857 unique community partners, and 2,081 community-university partnerships. Our goal with this research is to diverge from the traditional small-scale, qualitative assessment of a small sample of community partnerships to instead begin parsing large datasets in an effort to bring selected characteristics to the forefront in an attempt to parse big data.

Our initial research question of 'what do strong partnerships have in common' led us to discover that the unique characteristics of community-university partnerships means there are varying ideas on what quantitative metrics might be indicators of success. While there is general consensus that reciprocal partnerships are the hallmark, the methods used to identify those partnerships are almost entirely qualitative. For our purposes we must uncover quantitative metrics for each partnership that can be measured and compared to other partnerships. A model refers to the process of evaluating data via a computer algorithm. In our case we developed a descriptive model (Provost & Fawcett, 2013, pg 46), wherein we allow the computer to read all of the partnerships in the data set, and follow a set of rules to identify only partnerships that satisfy all of the rules in the algorithm. A common example of an algorithmic model is in weather forecasting. Barometers, thermometers and other sensing equipment collect data which is processed through a computer program that predicts incoming weather events.

To avoid development of biased modelling algorithms, we hosted a focus group and conducted a small survey of Collaboratory administrators, community engagement professionals, and institutional IR teams

to discover what proxy data might be most influential in identifying the strongest partnerships in Collaboratory data. Proxy data refers to specific data that is selected because it is similar to, but not the exact same as, a dataset that is missing (Provost & Fawcett, 2013, pg 212). Consider that in many data science applications there is not an immediate data point for the question at hand, but there are other data points available that allow for an approximation to be formed. An example is estimating the height of a tree from only a remaining stump. The tree is unavailable, but the rings in the stump provide a data point from which we can estimate the tree height. The tree rings are considered to be 'proxy data' for estimating the height of a tree.

Our survey uncovered that the fields most commonly considered to be indicators of strong partnerships were (in descending order of popularity) (see Figure 1):

- Length of time co-engaged with a partner
- The total number of activities with a partner
- The number of students involved in activities with a partner
- The number of faculty or staff members involved with a partner
- The frequency of new activities originating with a partner
- The total funding amount allocated to activities with a partner
- The geographic distance between a partner and institution (negligible results, discounted from model)

With these proxy data in mind, we strategized methods of developing a descriptive model to assess the entire Collaboratory dataset and filter only the strongest partnerships. Each proxy variable was assessed independently to evaluate its suitability for use in the model. If the data was acceptable we assigned filter values such that only the upper tier of partnerships were selected. The model applied all filters at once to the dataset and only those partnerships that satisfied the criteria for each proxy value were included in the final dataset of 'strongest

Figure 1. Algorithmic Assessment Model of Partnership Strength

partnerships'. Our model is built on a jupyter notebook using the pandas library to parse and evaluate our quantitative data and the seaborn library to visualize our results, the following sections describe our process through developing the model.

Duration of time spent co-engaged

The consensus from our focus groups and survey indicated that the duration of time a university and partner were co-engaged is the leading quantitative indicator of partnership strength. With this in mind we aggregated the duration of all activities in a partnership. This allowed us to calculate the sum total of simultaneous and sequential activities.

The results indicate that a significant number of partnerships are short-term in nature, 30 days or less. The Collaboratory system collects dates in a YYYYMM

Figure 2. Cumulative duration of partnerships (in days) in Collaboratory dataset

format, therefore, the shortest interval for an activity is 1 month (30 days). The mean duration of partnerships is 1067 days. For our strength modeling we assumed that the model filter value should be set at 1825 days (we calculated this using the mean + 1 standard deviation, a convention that allowed us to roughly slice the upper crust of the dataset). The model filter will select only partnerships where the cumulative duration of the partnership is longer than 1825 days. This selects 748 partnerships, or approximately the top 36% of activities to be processed through the remaining model filters.

Total number of activities per partner

Related to, but not the same as, duration of partnership, our survey indicated that the number of activities per partner was an indicator of a strong partnership. Our focus group revealed that continually initiating new partnership activities points to mutual trust and willingness to engage, indicators of strong partnerships. This proxy value was combined with another field identified by our survey, 'frequency of new activity initiation', as the count of activities and the frequency with which they are initiated are intrinsically linked. Analysis of activity counts per partnership revealed that the majority of partnerships exist as single-activity partnerships. That is to say that most partnerships (61%) present as a single activity. Therefore our model was set to select 'strong partnerships' as those where more than one activity is listed in the data. Note that this does not imply that all single-activity partnerships are not strong, but rather for our modelling methodology, we must ascribe a cut-off value.

Figure 3. Count of faculty, staff, and students involved in partnership activities

Figure 4. Partnership Strength Model

While the number of participants in a partnership was generally regarded as related to the strength of a given partnership, for modeling purposes we were required to omit this proxy field from the algorithmic model. While investigating the counts of faculty, staff, and students we found there to be wide variation in the volunteered data in the dataset. We found wide discrepancies in both faculty/staff summary statistics and student summaries. In each case we noted that the counts were sparse, and were dramatically different - ranging from zero into the thousands. Numerical data with the wide distributions are often unreliable without sufficient context into why the data looks as it does. Due to the uneven distribution of the data and lack of contextual reasoning for the data presented in this way, we leave these counts out of our model.

Funding allocated to a partnership

Finally, we assessed funding details to determine their adequacy for inclusion in the strength model. We found that there were 460 funded partnerships in the dataset, with a mean funding value of \$599,356. Again, there is a disparity in the values of funding that presents with the standard deviation of funding dollars being higher than the mean, however with such large dollar values (the max funding was found to be over 12 million dollars), we are able to include this metric in our assessment model by assuming that any partnership with such high funding dollars allocated to it must be supported in ways that would denote a strong partnership. While the dataset supports more refined data analysis of funding, including

public/private sources or annual values, we decide here to simply qualify partnerships as either funded or unfunded. For modelling purposes we assign a filter to identify only partnerships with a listed funder for any amount, which left us with 460 partnerships.

Building the model

We used the four acceptable proxy values to develop an algorithmic model that ingested all partnerships, evaluated their characteristics against the filter values we assigned above, and produced a list of only partnerships that satisfy all given conditions. We are left with a list of 99 partnerships that the model suggests to be 'strong partnerships'.

When we query the entire dataset for these values we end up with a model that looks like this:

```

SELECT partnerships
WHERE duration > 1825 days
AND number_of_activities > 1 AND
partnership_is_funded
  
```

Putting it All Together

The resultant 99 partnerships simultaneously satisfy all filtering criteria. They have endured for longer than 1825 days, they have more than one activity together, and they are funded. Of the 2083 partnerships in the dataset, the model selected only 4.75% of them as 'high strength'.

This quantitative exercise allowed us to quickly evaluate a complex, interconnected dataset of over 2000 activities to identify the top 5% given metrics

supplied by industry professionals. Recall that our intention is not to now assume that these 99 partnerships are the strongest in the dataset, but rather to devise a method to quickly parse qualitative data based on the limited quantitative elements therein. The logical next step to this research is to qualitatively assess a sample of these 99 partnerships to ground-truth the model's assessment that they are indeed strong. We propose that a modeling method like this is a suitable means of quickly slicing a large dataset of community partnerships to quickly identify a subset of partnerships that are meaningful in given ways. In the case of this model we assess based on perceived indicators of partnership strength, though the method will hold with varied quantitative inputs to describe other types of relationships.

Limitations

We have suggested that an automated system for initially processing qualitative community engagement data is a pragmatic approach to meeting the needs of big data analysis. Within the context of our research, we do note some limitations to this approach. First and foremost is the assumption that the model developed here is accurate in identifying strong partnerships. While the methods we used were a consensus of industry professionals, we do not assert that this is the only means of identifying strong partnerships in this dataset. We firmly believe that there may be strong partnerships that are bypassed by this model, or that the model selects partnerships that are not strong by traditional scoring metrics. A programmatic model such as this is limited by the selection criteria that we assign. As time goes on we will continue to refine the model.

Additionally, this research is limited by the size of the focus group we consulted in the early phases of this research. Our focus group consisted of Collaboratory administrators, community engagement professionals and some higher education leaders (vice-chancellors) that sat in on a roundtable discussion about how to

deploy an algorithmic model such as this. That same focus group was presented with the survey described in this paper. With roughly 10 responses to the survey to rely on, there is opportunity for bias in the survey results. Again, we note that this research represents only a first attempt at deploying such a process for analysis of community engagement activities. Lastly, each institution that uses Collaboratory chooses a different rollout strategy that affects Collaboratory's dataset. While we are working to approach generalizability, the dataset currently reflects the structural and cultural choices and limitations of the institutions with which we work, and therefore is likely not generalizable to all partnerships.

Conclusions & Next steps

This initial foray into exploring the Collaboratory national dataset has led us to develop an exploratory model for gauging partnership strength between an institution and community organizations. The model was built using input from administrators, community engagement specialists, and institutional leadership teams in both a focus group and a survey. Using their input we identified 4 proxy data fields that can be included to identify the strongest partnerships according to those metrics. There are numerous next steps for this research, most importantly is developing a method of qualitative assessment of community partnerships so that we can evaluate the accuracy of our model. Should we qualitatively discover that the 99 resultant partnerships are strong, we can have higher confidence in the validity of the model.

It is important to note that the method is not a silver-bullet. We do not proclaim that these 99 activities are the strongest in the dataset, but rather that these 99 activities satisfy a consensus about what might indicate a partnership is strong. As previously noted, assessment of qualitative datasets must not be left solely to algorithmic models, to do so repudiates the context and wisdom of the analyst (Prensky, 2001), however the analysis tools presented by modern

big-data computing tools are too powerful to ignore in the case of parsing large datasets such as this. This initial foray into exploring the Collaboratory national dataset allowed us to develop an exploratory model to gauge strength using contemporary tools to investigate a large, relational dataset of activities, work that previously was either conducted by hand or not performed at all. Previously, identifying strong partnerships was largely dependent on first-hand knowledge of a partnership and word-of-mouth details about a relationship that is known to be transformational and democratic. Through the modeling technique developed here, we are able to quickly parse a complex dataset to identify a subset of relationships that should be of interest, thereby significantly reducing the amount of time it takes to find relationships of note. An added benefit is shedding the burdensome, and often biased method of finding strong relationships through anecdotal evidence or collective memory.

The modeling method of using quantitative indicators for parsing generally qualitative data is feasible, given the appropriate inputs. When presented with the problem of parsing large scale qualitative data such as Collaboratory data, this method is a viable solution to quickly identifying only those data points that are most likely to be usable in future research. Additionally, this modeling and exploration has resulted in an initial model of evaluating partnership strength, giving us insight into what institutions believe to be important indicators of strength. As research in this area matures, we can use this modeling technique to continue to parse national-level data while applying lessons and metrics identified here.

References

- Boyle, M. E., Ross, L., & Stephens, J. C. (2011). Who has a stake? How stakeholder processes influence partnership sustainability. *Gateways*, 4, 100-118.
- Public Purpose Institute. (2021). *Community engagement classification (U.S.)*. <https://public-purpose.org/initiatives/carnegie-elective-classifications/community-engagement-classification-u-s/>
- Clayton, P. H., Bringle, R. G., Senor, B., Huq, J., & Morrison, M. (2010). Differentiating and assessing relationships in service-learning and civic engagement: Exploitative, transactional, or transformational. *Michigan Journal of Community Service Learning*, 16(2).
- Medlin, K. (2020). *Collaboratory dataset* [data file and codebook]. Collaboratory [producer].
- Dostilio, L. (2014). Democratically Engaged Community- University Partnerships: Reciprocal Determinants of Democratically Oriented Roles and Processes. *Journal of Higher Education Outreach and Engagement*, 18(4), 235. <https://openjournals.libs.uga.edu/jheoe/article/download/1159/1158/1562>
- Dostilio, Lina D.; Brackmann, Sarah M.; Edwards, Kathleen E.; Harrison, Barbara; Kliever, Brandon W.; Clayton, Patti H. (2012). Reciprocity: Saying What We Mean and Meaning What We Say. *Michigan Journal of Community Service-Learning* 19(1), 17-32.
- D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. The MIT Press.
- Eubanks, V. (2017). *Automating Inequality: how high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- Janke, E.M. (2008). Shared partnership identity between faculty and community partners.
- Doctoral dissertation, the Pennsylvania State University.
- Janke, E. M., Medlin, K. D., Norris, K. E., & Shelton, T. L. (2020, November 19) Examining mutual benefit & reciprocity as core distinctions between community engagement and public service: Implications for scholars, practitioners, & administrators [Webinar]. Campus Compact. <https://www.youtube.com/playlist?list=PLZpEMvWFeizGZ7zIFRxcYVoiAn8B>

hxKxo

- Janke, Emily, Flores, Santos, and Edwards, Kathleen. Dataset for "Community-Academic Partnerships in the Community Engagement Literature: A Scoping Review." Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2021-08-04. <https://doi.org/10.3886/E146781V2>
- Kitchin, R. (2017). The Data Revolution : Big data, open data, data infrastructures & their consequences. Sage.
- Prensky, M. (2001). H. Sapiens Digital: From Digital Immigrants and Digital Natives to Digital Wisdom. Innovate, 5(3). <https://www.learntechlib.org/d/104264>
- Provost, F., & Fawcett, T. (2013). Data Science for Business: What you need to know about data mining and data-analytic thinking. O'Reilly.
- Saltmarsh, J., Hartley, M., & Clayton, P. (2009). Democratic Engagement White Paper. New England Resource Center for Higher Education. https://repository.upenn.edu/gse_pubs/274
- Sandy, M., & Holland, B (2006). Different Worlds and Common Ground: Community Partner Perspectives on Campus-Community Partnerships. *Michigan Journal of Community Service Learning*, 13(1), 30-43.

About the Authors

Matthew Seto

Matt is the data, analytics, and BI reporting specialist at Collaboratory. Matt was one of the original members of University of Washington Tacoma's Office of Community Partnerships, where his work included managing technical projects with the Center for Applied Urban Research. His background is as a software developer, working at several Seattle-area startups before moving back to Academia. Matt's graduate work centered on applying cloud-computing solutions and automating processes to collect voluntarily generated spatial data and aggregate them through a pre-defined logic model. In his spare time, Matt can be found hiking, mountain biking, and playing with his two young children.

Kristin Medlin

Kristin oversees Collaboratory's research and development efforts that support, promote, and enhance the field of study related to community engagement and socially effective impact. She works to raise awareness about the value of community engagement data, the imperative to develop a data culture, and helps to build systems and structures for data access, analysis, and use. Prior to jumping into the corporate sector, Kristin served as the communications and partnerships manager in the Institute for Community & Economic Engagement at UNC Greensboro, where she co-founded Collaboratory alongside Drs. Emily Janke and Barbara Holland. Kristin's scholarly work explores cross-sector partnerships and tracking and assessment of community engagement. Kristin holds masters degrees from UNC Greensboro both in Public Affairs and in Educational Research, Assessment, and Evaluation.