GRAND Multidisciplinarity: The Way We Are

Emily Maemura
University of Toronto
e.maemura@mail.utoronto.ca

David Turner
University of Alberta
dwt@ualberta.ca

Jessica Perrie
University of Toronto
perrie@cs.toronto.edu

Kelly Lyons
University of Toronto
kelly.lyons@utoronto.ca

Eleni Stroulia
University of Alberta
stroulia@ualberta.ca

ABSTRACT
As part of the Media Enabled Organizational Workflow (MEOW) project, we are exploring if being part of the GRAND NCE is correlated with multidisciplinarity. Using subject fields from Scopus and the count of a researcher’s publications in each field, we developed a measure which can be used to evaluate an individual’s multidisciplinarity over time. Based on recent publication data, researchers who are part of GRAND were found to be more multidisciplinary, compared with a similar set of researchers not in GRAND. We also found that GRAND members publish in a significantly higher number of subject fields.

Keywords
Multidisciplinarity, collaborative research, research networks, bibliometrics.

1. INTRODUCTION
Funding agencies have established large-scale programs that bring researchers from multiple disciplines together. Previous work has studied how knowledge is exchanged between researchers and how multidisciplinary research can foster learning [3]. Multidisciplinary collaboration can also address problems that transcend a single distinct academic field. But how can a researcher’s multidisciplinarity be measured, or compared to a specialized research focus in a single field? Further, how are the products of multidisciplinary research published and disseminated across different fields? With its focus on multidisciplinary projects and teams, the GRAND network is an ideal site for studying this new breed of interdisciplinary large-scale research collaboration in an attempt to answer these questions.

Recent studies have explored multidisciplinary research by: measuring the outcomes (e.g., number of publications) of multidisciplinary projects; identifying the need for collaboration tools [2]; and by performing social network analysis of existing interdisciplinary teams to explore how knowledge exchange occurs [3]. Building on this research, we focus on the work of individual researchers (as opposed to groups) and we propose a metric for estimating their multidisciplinarity; we then use this metric to compare GRAND researchers with their Canadian peers outside GRAND. More specifically, the research question motivating this work was to assess whether or not individual researchers within the GRAND network are measurably and significantly more multidisciplinary in their publication output than other similar researchers who are not part of GRAND.

This work makes two contributions. First, the new bibliometric measure for assessing an individual’s publication breadth and disciplinary diversity enables us to examine a new set of research questions about collaboration, productivity and contribution. Second, our study of the GRAND community reveals that the research output of individual GRAND researchers is significantly distinct from that of their peers, which may imply that indeed the type of research conducted in such large-scale researcher networks is different from that in more traditional research environments. Just as the h-index [4] metric has come to provide important information about a researcher and can potentially impact career development, the development of a metric for multidisciplinarity can provide an additional dimension by which to measure a researcher’s work and research products

In the rest of this paper, we outline the existing literature on multidisciplinary research and citation analysis tools; we describe the process for gathering publication data for GRAND Principal Network Investigators (PNIs) and Collaborating Network Investigators (CNIs); we present our findings, and development of a ‘Multidisciplinary Measure’ to compare the GRAND researchers against a control population; and we discuss our findings and present some conclusions.

2. RELATED WORK
The term ‘multidisciplinary research’ can be used to refer to many different types of work, and Huutoniemi, et al. [5] have presented a useful conceptual framework for interdisciplinary research, distinguishing interdisciplinary work by three criteria: taking place between closely related or diverse fields; addressing research problems ‘ported’ from one field to another or alternately conceived across fields; and, finally epistemologically or instrumentally oriented. While these different types of interdisciplinary work have been variously described as research collaboration, cross-disciplinary research, interdisciplinary research, or multidisciplinary research, we will use ‘multidisciplinary’ herein to broadly encompass all of the above.

Exploring the benefits of multidisciplinary work for individual researchers, Sonnenwald [10] notes the differences across the lifecycle phases of a project. From a broad survey of different disciplines, it is concluded that the metrics for success vary significantly across disciplines, specifically in the importance of publication and citation metrics, suggesting that collaborating researchers should be aware of these different goals. However, it is generally understood that publication metrics, including the h-index, describe a researcher’s impact and influence in their field.

Yet, the calculation of these metrics can vary depending on the source of publication and citation data. Existing studies have compared both open-access and subscription-based online publication databases available for calculating these metrics, in particular Web of Science, Scopus, and Google Scholar. With a focus on the work of Israeli scientists, Bar-Ilan [6] found that Web
The study of collaboration specifically through bibliometric analysis is largely influenced by the work of Katz & Martin [7]. Their influence has led to work such as the extensive study by Cummings & Kiesler [2] which highlights gaps in existing collaboration methods and practices in order to increase the success of future projects. Porter, et al. [9] have furthered work in creating a metric for multidisciplinarity, working with the National Science Foundation and citation data from Web of Science. However, we found that their methods were not extensible to our project, particularly in the intensive work and coordination with Web of Science to create an initial mapping of subject categories. Additionally, where they use citations as a basis for measurement, we have instead chosen to use only the categorization of a publication’s venue, and do not cover citations in our multidisciplinary metric.

Reviewing these existing studies, bibliometrics have focused on influence and exposure by citation count, which is often narrowly focused in one specific field of study. Few existing measures account for a researcher’s reach or coverage in a wide range of disciplines through publication venues, and the intention of our work is to address this gap by contributing a new multidisciplinarity metric to existing bibliometrics analysis tools.

### 3. METHODOLOGY

#### 3.1 Choosing a source for publication data

The previous studies of citation databases noted above ([6], [8], [1]) have found that Scopus and Web of Science are the most accurate and complete databases from which to pull publication data. Our main criteria for selecting a database for our data collection and analysis were: the source is a recognized venue for providing publication information; data is available for a large set of researchers and their publications for each year; that existing categorization associates fields or disciplines with the publications; and, that the database be accessible and convenient for extracting data.

We also considered additional sources including Google Scholar, and Microsoft Academic Search, but based on the conclusions from the surveyed literature, as well as our own review of the different available tools, Scopus was chosen as our source for data collection, since it provides a balance of subject coverage, available data for our chosen time frame (within the past 5 years), and fewest errors.

#### 3.2 Defining the population of researchers

The scope of our initial analysis included the 199 active network investigators (NIs) in GRAND from 2007-2013. To determine if there is a significant impact of involvement in GRAND on the measurement of multidisciplinarity, we designated a ‘control’ group of similar researchers: for each GRAND NI, we noted their academic position, department, and institution, and manually paired them with a person of the same academic position, within the same department and institution who is not a part of the GRAND network. For each GRAND researcher, R, we visited the website of the academic unit of R and found the list of faculty members in the same academic unit. Each academic unit uses slightly different information architectures for their website so, given the specific order or presentation of researchers on the site, we manually found the first person in the given list at the same rank as R who was not a member of GRAND and had an author profile in Scopus. If the person did not have an author profile in Scopus (which was the case for approximately 20 people), we selected the next person on the given list at the same rank as R. We discovered that in some cases (approximately 15 individuals) there was no person in that unit at the same rank who was not also in GRAND. These GRAND researchers were excluded from our data set for analysis.

Once the lists of GRAND members and control researchers were finalized, we determined the unique author ID for each person in Scopus using their first and last names. Since there can be multiple people with the same name, the data from this gathering process was then further inspected and cleaned to ensure we had the correct set of author IDs for both the GRAND and control researchers. During this process we eliminated those GRAND researchers who do not have an author ID in Scopus (and removed their corresponding researcher in the control group). This resulted in 143 GRAND researchers, each paired with a researcher in the control group for the final analysis. We then gathered publication data for each of the 286 researchers for the years 2007 to 2013. This timeframe was chosen to provide us with data for three years prior to joining GRAND (2007-2009) and all the years since the start of GRAND (2010-2013). The publication data gathered included the number of publications per year and the number of publications in each Scopus subject area category (see Table 1).

### 3.3 Defining the taxonomy of disciplines

There is no widely accepted standard taxonomy of subjects to apply a consistent method for cataloging journal articles across academic disciplines. Our original goal was to implement a hierarchy in order to compare relative distance between academic fields of study so that collaboration between more closely related fields could be distinguished from collaboration across fields at a higher level of the hierarchy. We reviewed the Library of Congress subject headings, the Classification of Instructional Programs used for census data, the US National Academy of Sciences fields of study, as well as the categories used within each of the Google Scholar, Microsoft Academic Search, and Scopus sites themselves. None was found to provide significant benefits for our analysis, so we proceeded to use the hierarchy from Scopus.

The existing categories provided in the Scopus database are divided at a high level into four broad areas (life sciences, health sciences, physical sciences, and social sciences & humanities) which we will call fields below. Scopus further divides each of these into more detailed subject areas (27 subject areas in total). Unfortunately, the methods used by Scopus to categorize each are not openly available for review. While any one classification scheme may not be ideal in all cases, the Scopus Fields were deemed adequate for our purposes since they serve as a consistent

---

1 See Keywords and Index Terms and Journal Classification

classification for comparison. Table 1 shows the fields and subcategories.

Table 1. Scopus Subject Areas

<table>
<thead>
<tr>
<th>Subject Area</th>
<th>Subject Area Classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health Sciences</td>
<td>Medicine, Nursing, Veterinary, Dentistry, Health Professions</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>Arts and Humanities, Business, Management and Accounting, Decision Sciences, Economics, Econometrics and Finance, Psychology, Social Sciences</td>
</tr>
<tr>
<td>Life Sciences</td>
<td>Agricultural and Biological Sciences, Biochemistry, Genetics and Molecular Biology, Immunology and Microbiology, Neuroscience, Pharmacology, Toxicology and Pharmaceutics</td>
</tr>
</tbody>
</table>

3.4 Developing a Multidisciplinarity Measure

The data extracted from Scopus for each researcher includes the total number of papers published within our timeframe (2007-2013) and the number of publications for each of the 4 subject fields. In order to compare multidisciplinarity, we devised a measure based on the distribution of a researcher’s total publications across these four fields. The ideal, most multidisciplinary distribution is an equal percentage (25%) of publications in each of the four fields; conversely, the least multidisciplinary distribution is to have all publications categorized within a single field. The difference in the ratio of a researcher’s papers in a given field from the ideal (from 25%) is calculated for each field and squared to account for absolute value, and then the average difference of all four fields was calculated. The square root of this average value gives a measure of multidisciplinarity (MD1).

$$MD1 = \sqrt{\frac{\sum_{i=1}^{4} \left( \frac{\text{Publications in Field}_i}{\text{Total Publications}} - 0.25 \right)^2}{4}}$$

Since MD1 is based on percentages of a researcher’s total number of publications, more weight is given to a single paper when an author has a smaller total number of publications. This result is intentional, giving additional weight to papers from researchers who have published fewer papers distributed across a larger number of fields. However, this measure does not take into account the number of fields in which someone has published so it is possible for a researcher with coverage in a smaller number of fields (e.g., 2 of the 4 clusters) to rank higher (i.e., be more multidisciplinary) than a researcher with coverage in more fields. This might occur if a researcher has a smaller total number of publications, since each would contribute more, or a higher percentage, compared with a researcher who had published more total papers. Intuitively we want the multidisciplinary (MD) measure to provide more weight to publication coverage across a higher number of fields, so adjusting for this, our final multidisciplinary measure is: $MD = MD1/(\# \text{ of fields covered})$, where $\#$ of fields is an integer from 1 to 4. A value of $MD=0$ represents an even coverage of publications across all four fields. The highest possible value and, hence, the least multidisciplinary is $MD=0.433$.

4. FINDINGS

Result 1: The research output of GRAND researchers is significantly more multidisciplinary based on the MD metric

In a 1-tailed paired Wilcoxon Signed-Rank test, where the alternate hypothesis is that GRAND researchers’ MD values are lower than their corresponding control researchers’ values, it was found that GRAND researchers are significantly more multidisciplinary ($V = 3690.5, p < 0.05, r = 218.22$). Table 2 presents the medians, means and standard deviations for each groups’ MD values.

Table 2. Summary of MD values for GRAND and control

<table>
<thead>
<tr>
<th>Group</th>
<th>Median</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRAND</td>
<td>0.126</td>
<td>0.165</td>
<td>0.120</td>
</tr>
<tr>
<td>Control</td>
<td>0.155</td>
<td>0.196</td>
<td>0.131</td>
</tr>
</tbody>
</table>

Result 2: The research output of GRAND researchers covers more fields

As shown in Figure 1, when comparing the total number of fields covered, the frequency of GRAND researchers who publish in more fields than their control counterpart ($\Delta > 0$) is greater than the frequency of control researchers who publish in more fields than their counterpart GRAND researcher ($\Delta < 0$). To evaluate the hypothesis that GRAND researchers publish in more fields, a 1-sided Wilcoxon Signed-Rank test showed that those with GRAND membership published in significantly more fields ($V = 3278.5, p < 0.05, r = 193.86$). The median of GRAND researchers’ number of fields covered (median = 3) is also greater than the median of control researchers’ number of fields (median = 2).

Figure 1. A histogram of the difference in numbers of fields for GRAND researchers and their control counterparts

2 Adapted from (with ‘multidisciplinary’ classification removed) http://help.scopus.com/Content/h_subject_categories.htm
Result 3: The GRAND community includes a higher percentage of researchers who publish in more fields

The GRAND sample contains a higher percentage of researchers who have published in 3 or 4 fields, while the control sample contains a higher percentage of researchers who have published in 1 or 2 fields, as shown in Table 3. Applying a 2-sample z-test to compare sample proportions found that only marginally significantly more GRAND researchers publish in 3 and 4 fields than their control counterparts; for both 3 and 4 fields, z(143) = 1.6, p < 0.10. Similar results were found for 1 and 2 fields. In reviewing these results we must recognize how the assumptions for the test are not met: the two samples are not independent, and the GRAND population is not ten times as big as the sample.

Table 3. Number of individuals arranged by number of fields in which that individual has published. Within each number of fields, the larger values are bolded.

<table>
<thead>
<tr>
<th>#Fields</th>
<th>#GRAND</th>
<th>#Control</th>
<th>%GRAND</th>
<th>%Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21</td>
<td>30</td>
<td>14.69%</td>
<td>20.98%</td>
</tr>
<tr>
<td>2</td>
<td>47</td>
<td>60</td>
<td>32.87%</td>
<td>41.96%</td>
</tr>
<tr>
<td>3</td>
<td>46</td>
<td>34</td>
<td>32.17%</td>
<td>23.77%</td>
</tr>
<tr>
<td>4</td>
<td>29</td>
<td>19</td>
<td>20.28%</td>
<td>13.29%</td>
</tr>
</tbody>
</table>

5. DISCUSSION

Based on these three results, it can be concluded that GRAND researchers are significantly more multidisciplinary than their control counterparts, tend to publish in a greater number of fields, and, as a whole, contain a higher number of researchers who publish in more than two fields. However, this study does not distinguish between correlation and causation for multidisciplinarity and membership in GRAND. Researchers may not necessarily increase their multidisciplinarity as a result of joining GRAND, but could be more likely to join GRAND as a result of their multidisciplinarity. To understand these relationships, we will need to undertake further analysis by looking at a wider time frame that includes publication data for each researcher prior to their membership in GRAND.

5.1 Threats to Validity

There are a number of limitations in our methods and collection of data. Using publication data from Scopus has relied upon a categorization scheme which allows one publication to have more than one subject area. Since many of these subject areas fall within the same broader field, this results in a heavier weight within that field - i.e. a single paper will be classified in terms of subject area as both Computer Science, and Engineering; since these are both within the Physical Sciences field, the paper will essentially be counted twice in that field using our metric. Conversely, this limitation can also result in the same paper being categorized in subject areas from two separate fields, which will again alter the calculation. In any case, this weighting does not impact the measure of overall coverage of fields, confirming that GRAND members have greater coverage than the control group.

Our method of identifying researchers to include in the control group could also be improved. For example, it might be more representative to find all members of an academic unit at the same rank and compute a combined measure of their multidisciplinarity (using max., min., or mean values, for example).

While every effort was made to ensure that the author data retrieved from Scopus was a match for the researcher from GRAND or the control group, some authors had two separate entity IDs in the Scopus database, and their publications were split between them. At this time, we have not adjusted our methods to account for two separate sets of data for one author. Another limitation of the Scopus data is that there may be some missing papers from the later part of 2013, which have not yet been indexed and added to the system.

6. CONCLUSIONS

Our findings support our hypothesis that GRAND researchers are measurably more multidisciplinary than similar researchers who are not in GRAND. We hope to build on these findings in the future by completing further analysis to include other factors such as gender, age, stage of career, and how active each researcher is within GRAND. Future studies can also consider the measure of multidisciplinarity in relation to geographical location, institution, researcher productivity (through publications, presentations, and other artifacts created), and their impact (citation counts, h-index), as well as extend its application beyond the GRAND network.

7. ACKNOWLEDGMENTS

The authors wish to thank Kellogg Booth and Adrian Sheppard for their insights, guidance, and discussions about this work and Barry Wellman for motivating us in this direction.

8. REFERENCES