

# Where Are You Hiding? Identifying Diabetes Researchers



Mirong Luo<sup>1</sup>, PF Anderson<sup>1</sup>, Liang Gou<sup>2</sup>, Xiaolong (Luke) Zhang<sup>2</sup>, Kevin Coffman<sup>1</sup>, Mark MacEachern<sup>1</sup>  
 1) University of Michigan, 2) Pennsylvania State University

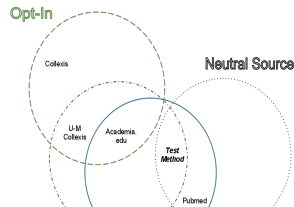


## Introduction

### Background

Diabetes research is studied in many disciplines. Large research universities provide a favorable environment for conducting translational research by hosting researchers from different disciplines. Researchers are challenged to discover appropriate collaborators. Existing tools for collaborator discovery have different abilities.

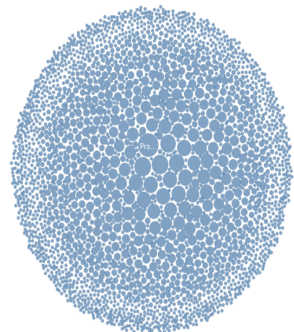
Figure 1: Tools Comparison



### Goals

- Define and validate a method to identify diabetes researchers working in different primary disciplines. **Campus-Wide**
- Depict a landscape of diabetes research at U-M:
  - Identify intra-institutional affiliations of diabetes researchers at U-M;
  - Identify both cross-departmental and same-unit collaborations through co-authorship analysis;
- Identify "collaboration gaps"
- Understand barriers to collaboration

Figure 2— MeSH term distribution for Strategy 2



## Methods

### Data Collection - MEDLINE Search Strategies

#### Strategy 1

Search primary diabetes terms from MeSH (Medical Subject Headings) limited by primary author address and last 5 years. **STRENGTH:** Provides data regarding campus researchers working in core diabetes research who may not be identified as such through campus resources. **WEAKNESS:** Fails to capture research not cataloged as primary diabetes MeSH terms. This included research by some known influential diabetes researchers on campus.

#### Strategy 2

Search name of researchers identified as affiliated with the University of Michigan diabetes research centers, limited by institutional address of the primary authors and last 5 years. **STRENGTH:** Provides data regarding diabetes-related research that is not identified as such through MeSH cataloging. **WEAKNESSES:** (1) Fails to capture research by campus researchers for whom their primary affiliation is with another campus unit. (2) Captures articles written by diabetes researchers on non-diabetes topics.

### Selected Strategy to Generate Data

Combine Strategies 1 and 2 above. Clean data to remove non-diabetes articles by applying Pareto's Principle. This was done by combining all MeSH terms harvested from Strategy 2, weighting by frequency distribution, retaining most-used terms to the level required for 80% of the citations retrieved in Strategy 2. This was 2846 unique terms of which 66 were used. [See Figure 2]

### Data Description

Final search strategy retrieved 829 articles. Data was cleaned by removing duplicates, articles from wrong authors with the same names, articles from other institutions and incorrect addresses for a final data set of 614 citations representing 847 number of authors and 2500 author relationships.

### Data Analysis - Visualization Search Strategies

**ScaleNetVis** was selected as our visualization tool. **ScaleNetVis** provides multiscale and cross-scale visualization of attribute-rich social networks. This new tool helps answer those questions concerning the social relationships that span different levels, such as how individuals in a specific group are related to other groups, in what ways social relationships in two groups may differ, who are those actors linking two groups, etc.

## Results

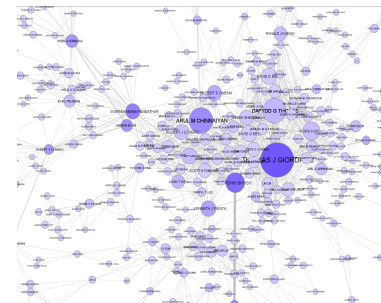


Figure 3 – Close view of the largest component in the whole co-author network

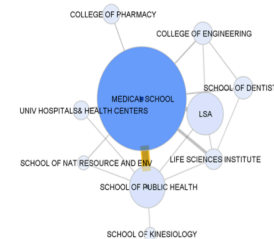


Figure 4 – Collaboration network among different schools

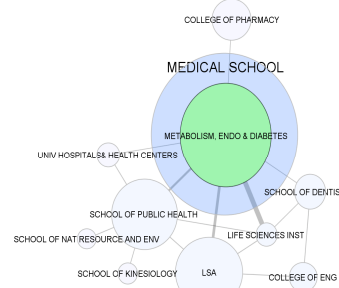


Figure 5 – Collaborations between "Metabolism, Endocrinology & Diabetes (MED)" with other schools, filtering out other departments in the Medical School.

## More Results

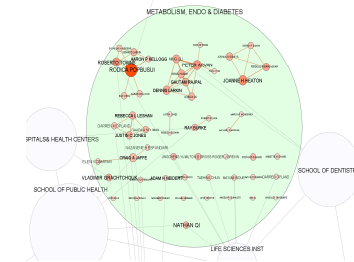


Figure 6 – Example: Collaborations between "Metabolism, Endocrinology & Diabetes (MED)" with other schools.

## Discussion

### How to define the boundary of diabetes research

Diabetes is an intricate and diverse topic, encompassing many disciplines, research techniques & methodologies, with etiologic, diagnostic and prognostic factors encompassing much of the human body, with core issues such as inflammation that cross discipline boundaries. These factors, which make it such a valuable area of study, also make it very difficult to clearly define the boundaries of the field. We used an application of the Pareto Principle to attempt to address this issue methodologically, while recognizing the inherent weakness of this approach. The ideal approach would have been to solicit a minimum of 2 diabetes researchers to apply a systematic review approach to determining which of the articles retrieved by Strategy 2 were actually diabetes-related.

### Applications of the results

This baseline study will allow us to later assess the impact of changes and interventions in the local environment as well as changes over time. Primarily, it is desirable to assess the impact on collaboration behaviors from the construction of a new physical space for campus diabetes researchers.

Identifying well connected hubs may allow more efficient information dissemination among the community and provide the opportunity to identify models of practice for collaboration within the community.

### Next Steps

- Refine search methodology with lessons learned and diabetes terms and concepts identified here.
- Interview selected researchers regarding collaboration behaviors and barriers.
- Extend search scope to broader international diabetes research community.