Generating an Optimal Editorial Board for a Research Journal

Bobby Ullman, with guidance from David Walker

Motivation

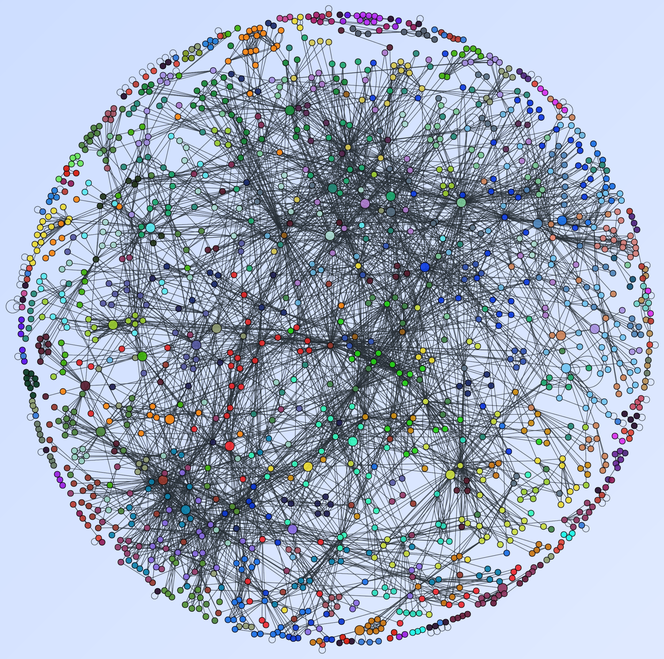
- We seek to find a representative group for a large community such as:
  - an editorial board for a journal;
  - top scientists for the President's Council of Advisors;
  - a program committee for a conference; or
  - membership of a university department.
- A well-chosen representative group:
  - covers all specialized fields in the community;
  - is comprised of leaders in those fields; and
  - represents the fields proportionally to the prevalence of the field in the community.
- Currently, most representative groups are selected via social mechanisms. A small well-connected group of people use their personal knowledge of the community and suggestions from colleagues to select an editorial board.

Our Approach

- We generate an editorial board by algorithmic analysis of data by:
  - collecting clerical data for past published papers;
  - representing the important relationship data of a scientific community in a graph;
  - applying graph clustering algorithms and vertex centrality measures to break the community into topic clusters; and
  - selecting representative members from each cluster.
- These representative members represent the community topically by:
  - clustering to ensure topical coverage;
  - selecting the most frequently publishing contributors to produce leaders in each field; and
  - assigning membership relative to the size of the field in the community to ensure proportional coverage.
- We produce a refined list of candidates to serve as a starting point for the traditional editorial board selection process.

The Community Graph

- Each vertex is a contributing author.
- Edges represent co-authorship and citations.
- Weight between co-authors is inversely proportional to the number of authors on a paper.
- High edge weight between authors indicates a close academic relationship.



Stochastic Blockmodeling

- Partition the vertex set into B blocks so as to minimize some carefully chosen objective function.
- Measure the number Ω of ways to assign edges that satisfy the current block arrangement:

$$\Omega = \prod_{a \geq b} \Omega_{ab} \quad \Omega_{ab} = \binom{n_a n_b}{e_{ab}} \quad S = \ln \Omega$$

- Measure the quantity of information L in the graph:

$$P = \binom{B}{2} \qquad L = \ln \left[\binom{P}{E} B^N \right]$$

- Minimize the description length $D = S + L$.
- Use randomization (Monte Carlo Interference) to efficiently minimize D.

Quality Heuristic

- Take a graph G with vertex set V. We judge the quality of a representative set S by the average representation distance of the set:

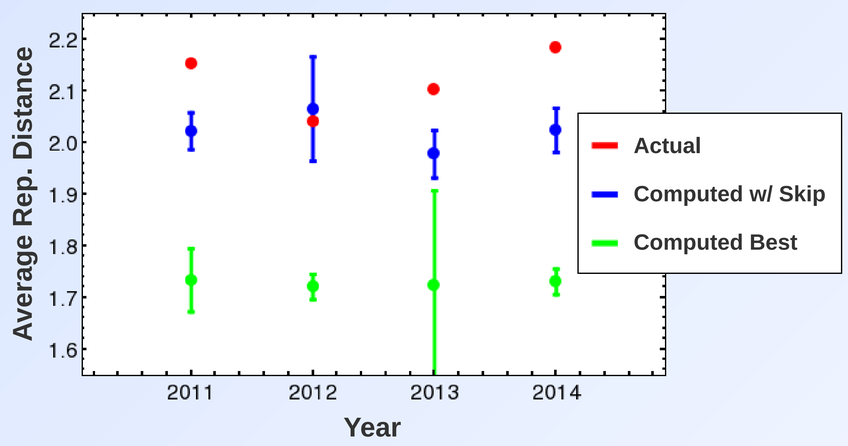
$$\alpha_G(S) = \frac{1}{|V|} \sum_{v \in V} d_S(v),$$

where $d_S(v)$ is the distance from v to S.

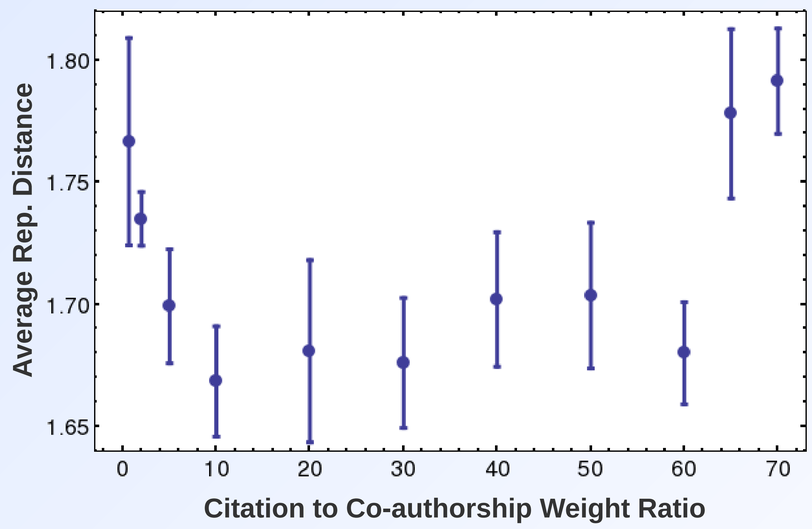
POPL Conference

- Principles of Programming Languages (POPL)

Our approach performs better



Balancing citations and co-authorship is best



Future Work

- Apply this technique to larger communities.
- Change edge weighting assignments.
- Use relationship data to determine centrality.
- Test with other quality heuristics.