Local Algorithms for Finding Densely Connected Clusters

Peter Macgregor He Sun

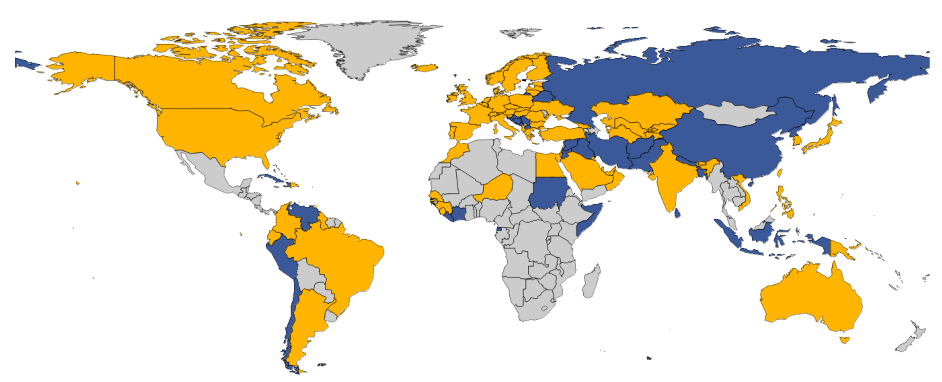
University of Edinburgh

Problem Statement

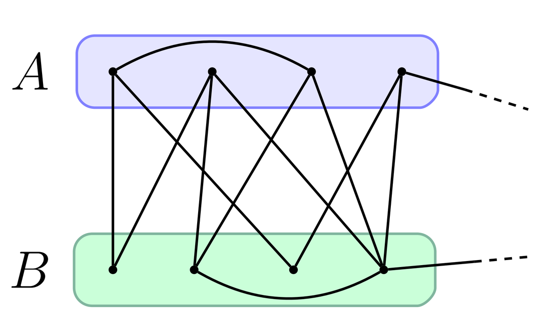
Given a graph, find two clusters A and B which are densely connected to each other and loosely connected to the rest of the graph.

Constraint in the local setting: the algorithm should run in time proportional to the size of returned clusters A and B.

Application: Interstate Dispute Graph



Bipartiteness



- Many edges between A and B
- Few edges inside A or B
- Few edges to the rest of the graph

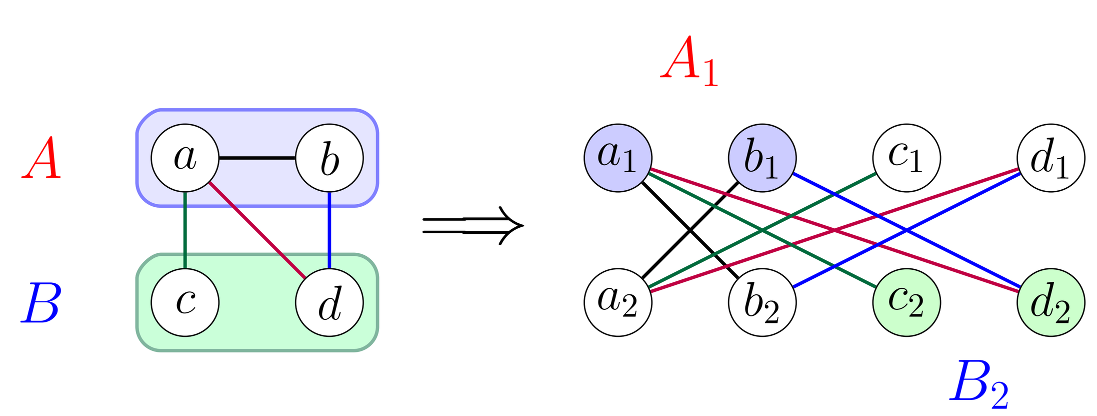
Bipartiteness Ratio [Trevisan 2012]

$$\beta(A,B) = 1 - \frac{2w(A,B)}{\mathrm{vol}(A \cup B)}$$

Reduction by Double Cover

Given a graph G, its double cover H is constructed as follows:

- every vertex v has two corresponding vertices v_1, v_2;
- for every edge $\{u,v\}$, there are edges $\{u_1, v_2\}$ and $\{u_2, v_1\}$ in H.



$$\beta(A,B) \quad = \quad \Phi(A_1 \cup B_2)$$

Lemma

For sets A and B in the graph G, it holds that $\beta_G(A,B) = \Phi_H(A_1 \cup B_2)$.

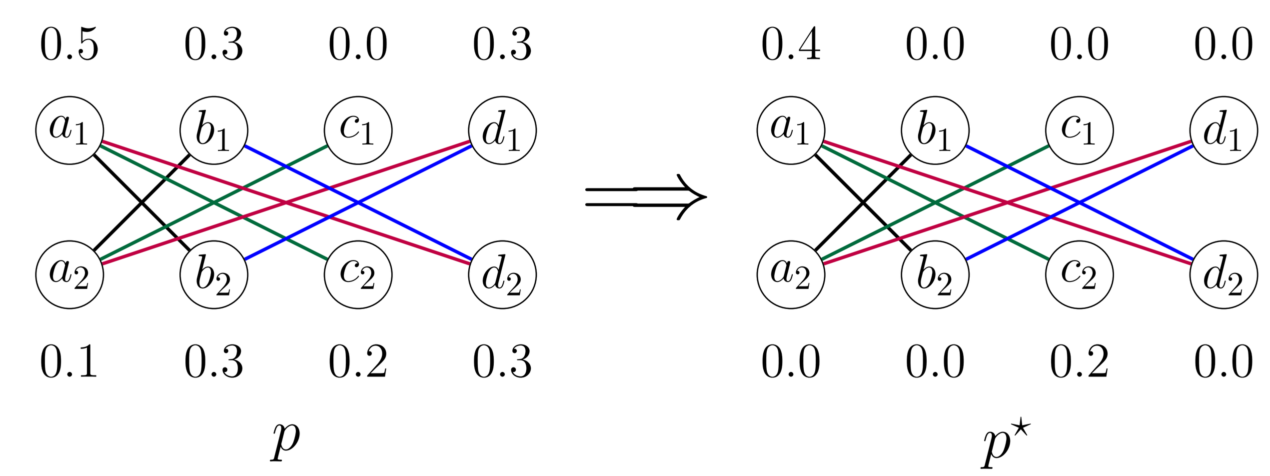
The Simplify Operator

Simplified Vector

For any $p \in \mathbb{R}^{2n}_{\geq 0}$, the simplified vector $p^\star$ is defined by

$$p^\star(u_1) \triangleq \max(0, p(u_1) - p(u_2)),$$
$$p^\star(u_2) \triangleq \max(0, p(u_2) - p(u_1))$$

for every vertex u.



The Algorithm for Undirected Graphs

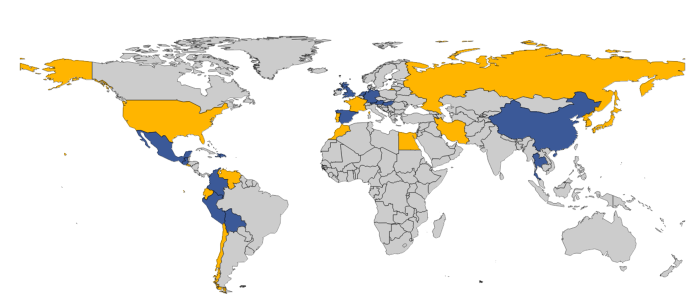
Algorithm (Informal)

1. Construct the double cover
2. Apply ACL-like method to compute vertex importances in the double cover
3. Compute the simplified importance vector
4. Return the vertices with high importance

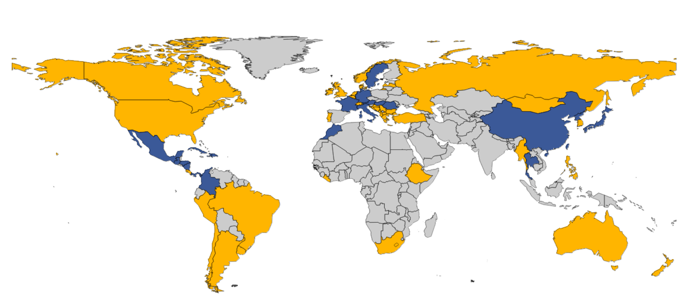
Theorem (Informal)

Given a graph G and a starting vertex u from the target set $A \cup B$, there is a local algorithm which returns (A', B') with $\beta(A', B') = O\left(\sqrt{\beta(A,B)}\right)$.

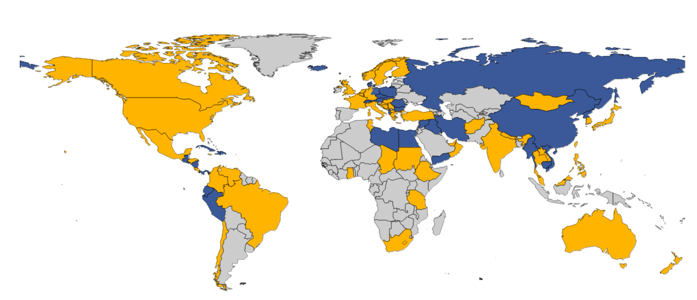
Experimental Result: Interstate Dispute Graph



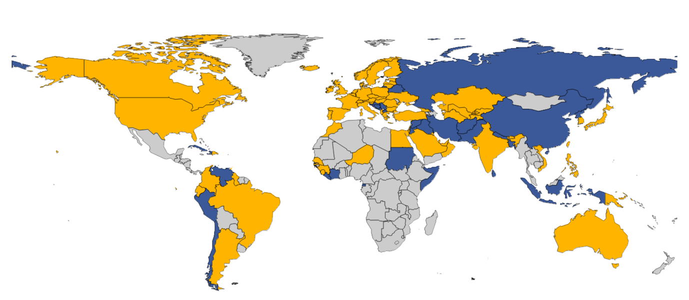
1816 -- 1900



1900--1950

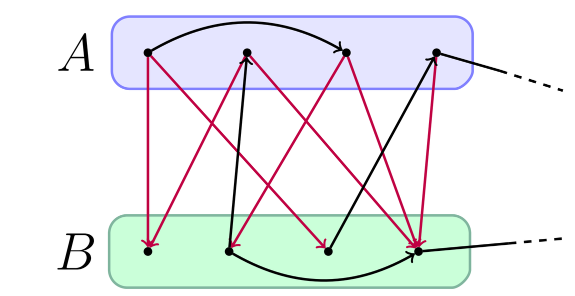


1950 -- 1990



1990 --2010

Flow Ratio



- Many edges from A to B
- Few edges inside A or B
- Few edges to the rest of the graph

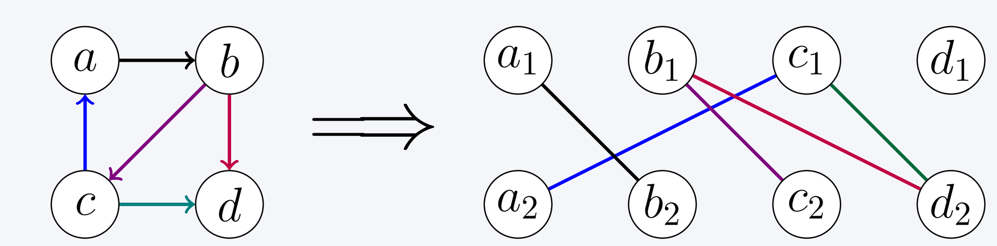
Flow Ratio

$$F(A,B) = 1 - \frac{2w(A,B)}{\mathrm{vol}_{\mathrm{out}}(A) + \mathrm{vol}_{\mathrm{in}}(B)}$$

The Algorithm for Directed Graphs

Given a graph G, its semi-double cover H is constructed as follows:

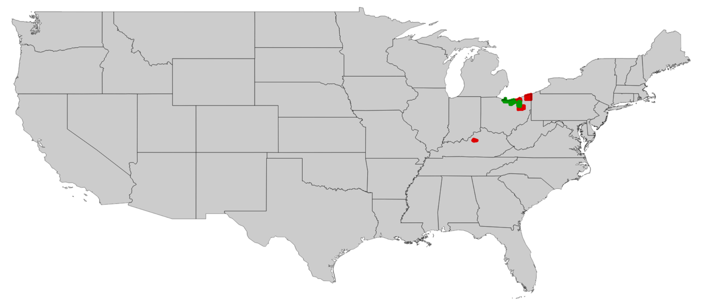
- every vertex v has two corresponding vertices v_1, v_2;
- for every edge (u,v), there is an edge $\{u_1, v_2\}$ in H.



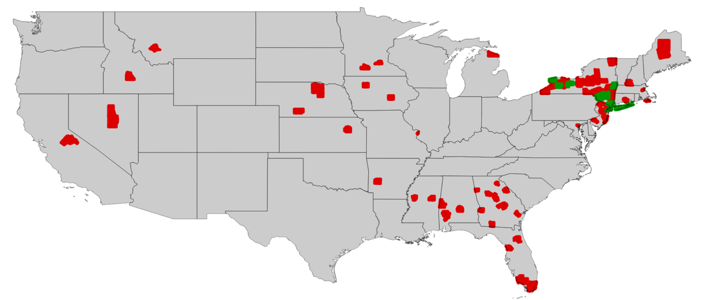
Theorem (Informal)

There is a local algorithm based on the semi-double cover which returns two clusters A and B with a bounded flow ratio $F(A,B)$.

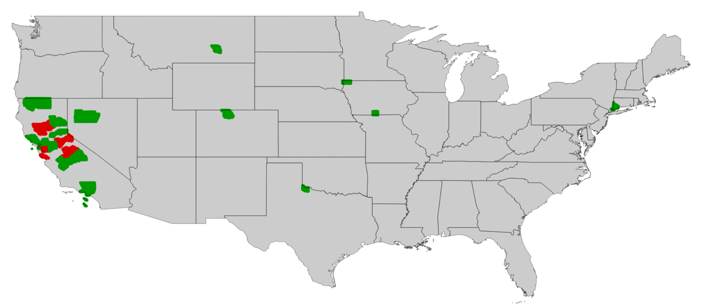
Experimental Results: USA Migration Graph



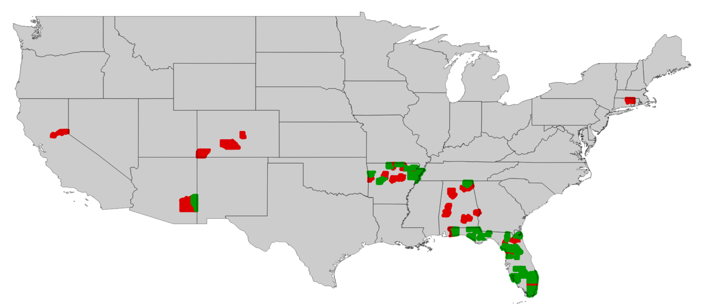
Migration to Cleveland, Ohio



Migration to NYC and Buffalo, NY



Migration in California, to LA and SF



Migration to Florida and Arkansas

General migration trend from rural to urban areas