# Embedding, Clustering and Coloring for Dynamic Maps

Sankar Veeramoni
University of Arizona

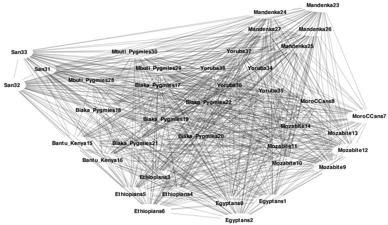
. . .

*Joint work with*

*Yifan Hu at AT&T Research and Stephen Kobourov at University of Arizona*

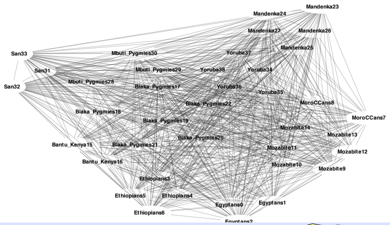
# GMAP

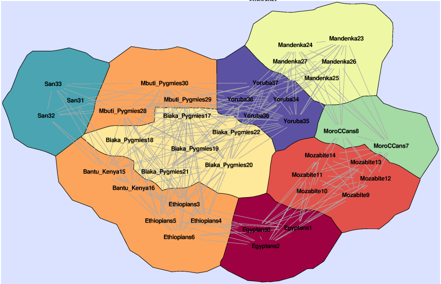
- Given high-dimensional data



# GMAP

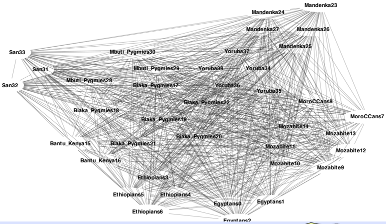
- Given high-dimensional data
- GMAP
  - clusters form countries
  - interconnected clusters share boundaries

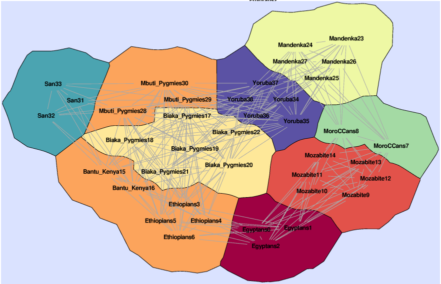


# GMAP

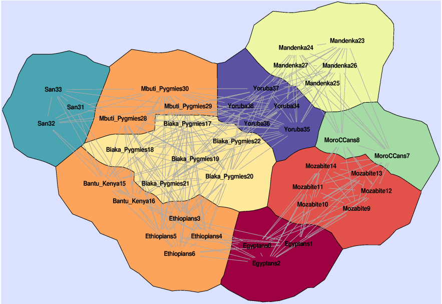
- Given high-dimensional data
- GMAP
  - clusters form countries
  - interconnected clusters share boundaries
  - visualize connectivity information (via edges)
  - visualize clustering information (via countries)

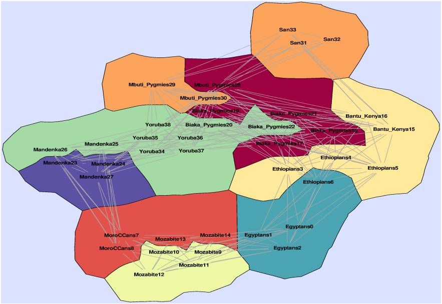


# Simulataneous map visualization

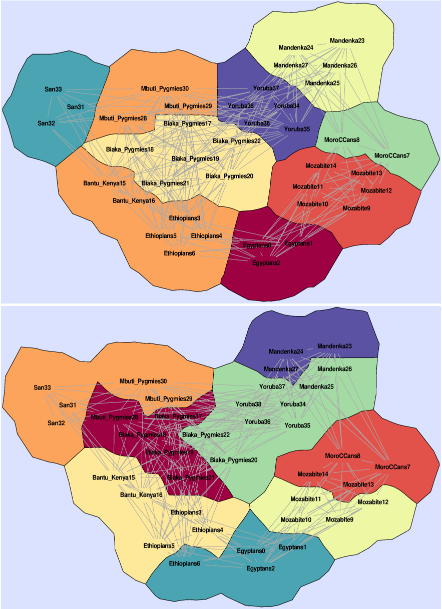
- Compare two or more maps obtained using different similarity metrics.



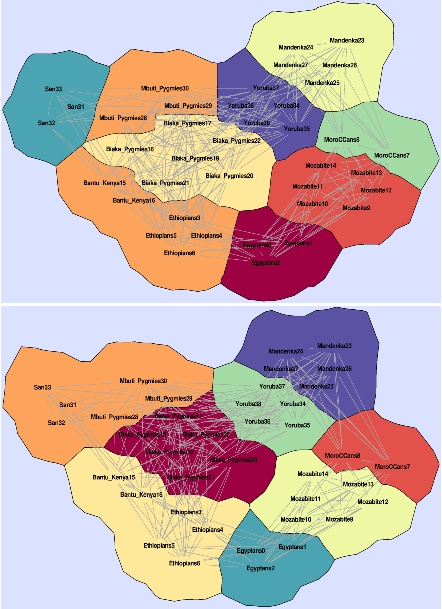
# Simulataneous map visualization

- Compare two or more maps obtained using different similarity metrics.
- Dynamic Map Layout



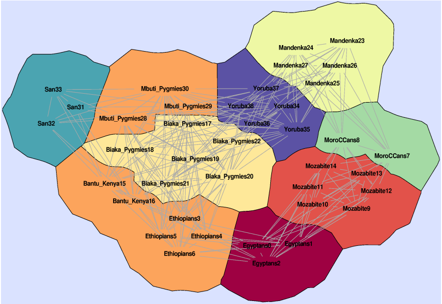
# Simulataneous map visualization

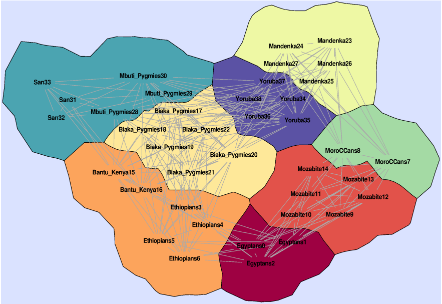
- Compare two or more maps obtained using different similarity metrics.
- Dynamic Map Layout
- Dynamic clustering



# Simulataneous map visualization

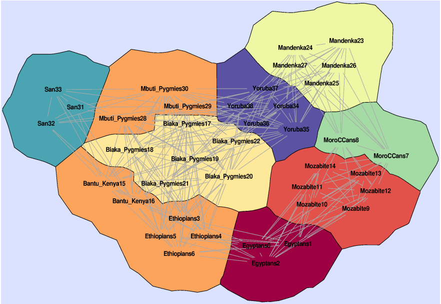
- Compare two or more maps obtained using different similarity metrics.
- Dynamic Map Layout
- Dynamic clustering
- Dynamic Map Coloring

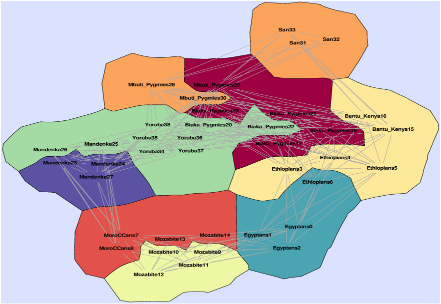


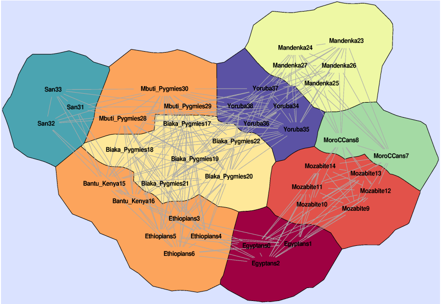
# Dynamic Map Layout

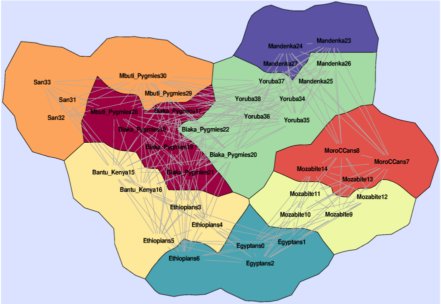
- two graphs with very similar topology can have very different drawings



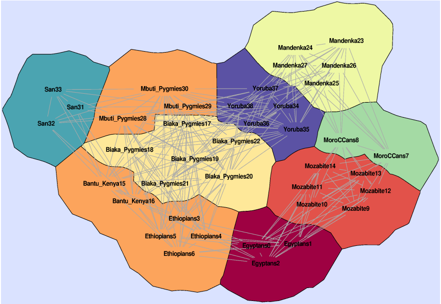
- two graphs with very similar topology can have very different drawings
- compute node positions to preserve mental map

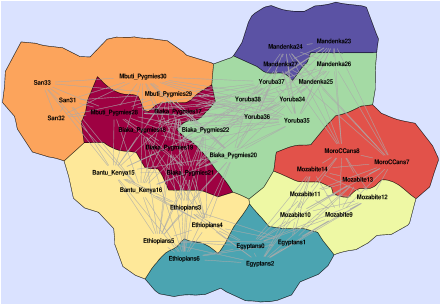


# Dynamic Map Layout

- two graphs with very similar topology can have very different drawings
- compute node positions to preserve mental map
- Approach :
  - initialize the node positions
  - apply multidimensional scaling (MDS) layout
  - Procrustes transform the coordinates of the nodes



# Procrustes transformation

- find the best alignment of two input layouts

# Procrustes transformation

- find the best alignment of two input layouts
- Apply rotation, translation, and scaling

# Procrustes transformation

- find the best alignment of two input layouts
- Apply rotation, translation, and scaling
- Mathematically
  - Node positions from first layout, $y_i, \ i = 1, 2, \ldots, |V|$
  - Node positions from second layout $x_i, \ i = 1, 2, \ldots, |V|$

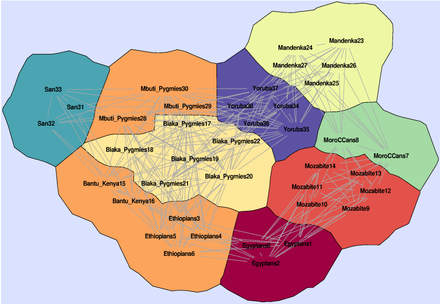
# Procrustes transformation

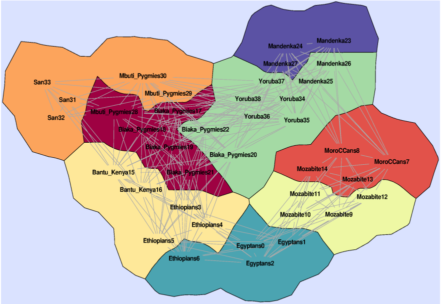
- find the best alignment of two input layouts
- Apply rotation, translation, and scaling
- Mathematically
  - Node positions from first layout, $y_i, \; i = 1, 2, \ldots, |V|$
  - Node positions from second layout $x_i, \; i = 1, 2, \ldots, |V|$
  - find a translation vector $b$, scaling value $\rho$ and rotation matrix $T$ that minimize

$$\sum_{i=1}^{|V|} \|y_i - (\rho T x_i + b)\|^2. \tag{1}$$

# Dynamic clustering

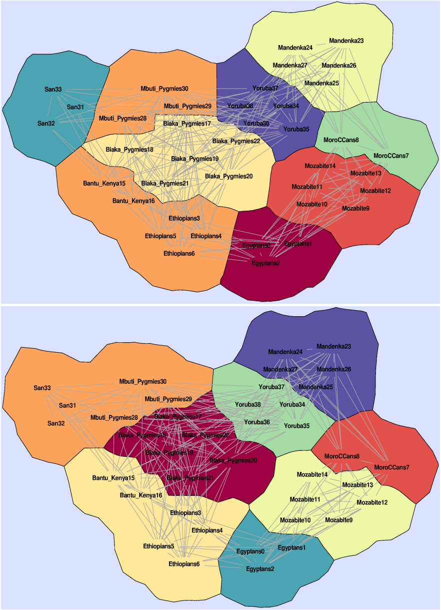
- cluster stability $\Rightarrow$ preserve the clustering structure
- Modularity $\Rightarrow$ cluster quality measure
    - Modularity based clustering by Blondel et al.



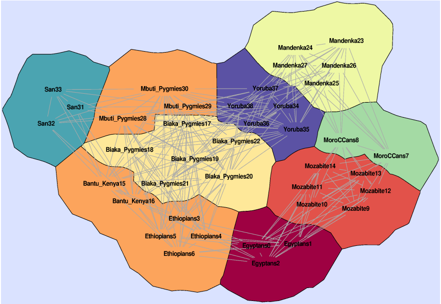
# Dynamic clustering

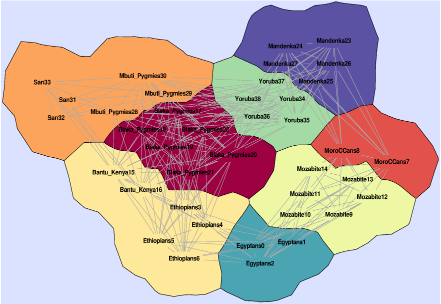
- cluster stability ⇒ preserve the clustering structure
- Modularity ⇒ cluster quality measure
  - Modularity based clustering by Blondel et al.
- Heuristic to combine the two objectives - modularity and cluster stability



# Dynamic clustering

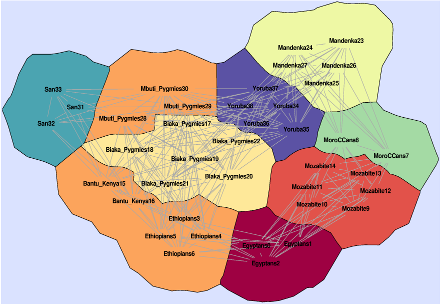
- cluster stability ⇒ preserve the clustering structure
- Modularity ⇒ cluster quality measure
  - Modularity based clustering by Blondel et al.
- Heuristic to combine the two objectives - modularity and cluster stability
- Dynamic variation of Blondel's clustering algorithm
- Clustering of one map is used as pre-clustering.

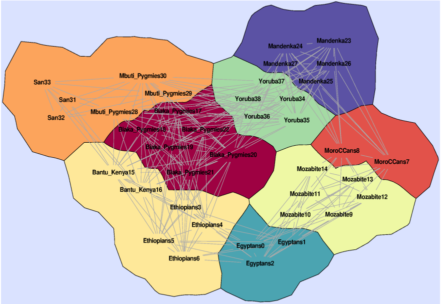


# Dynamic Map Coloring

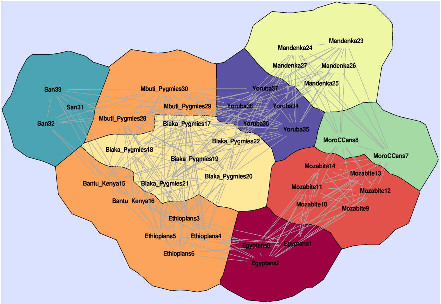
- same color for similar countries

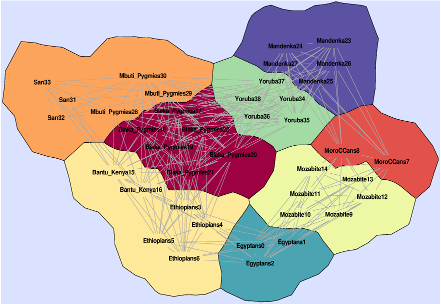


# Dynamic Map Coloring

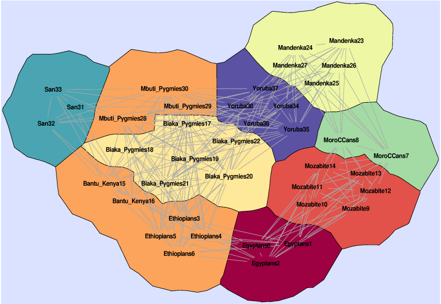
- same color for similar countries
- similar countries ⇒ more nodes in common.

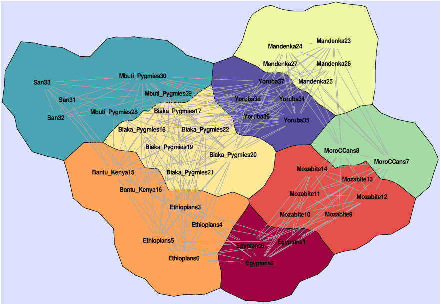


# Dynamic Map Coloring

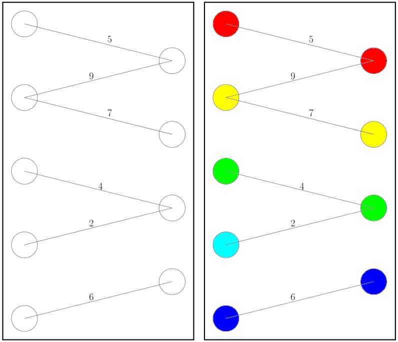
- same color for similar countries
- similar countries ⇒ more nodes in common.
- can be modeled with maximum weighted matching (MWM) of a bipartite graph.



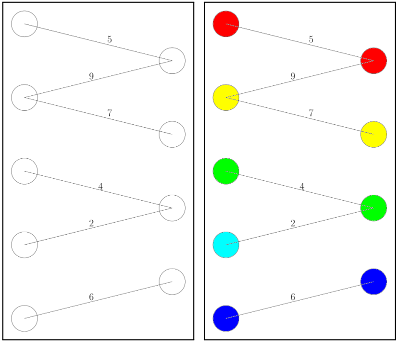
# maximum weighted matching

- Nodes - clusters(countries) in the two maps
- Edge weight - number of objects that are common between clusters



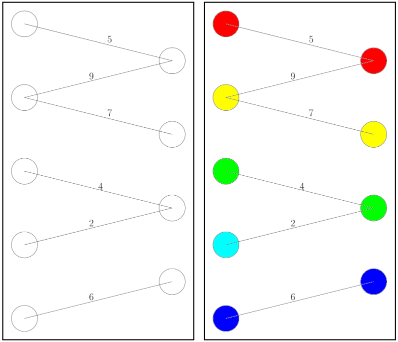
# maximum weighted matching

- Nodes - clusters(countries) in the two maps
- Edge weight - number of objects that are common between clusters
- Hungarian algorithm
  - Running time - $O(mn + n^2 \log n)$
  - m - edges, n - nodes



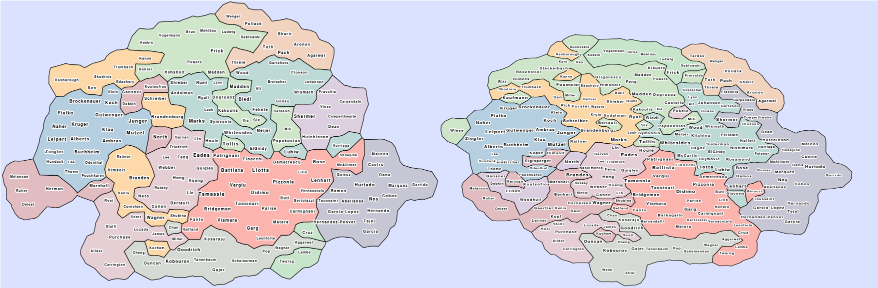
# maximum weighted matching

- Nodes - clusters(countries) in the two maps
- Edge weight - number of objects that are common between clusters
- Hungarian algorithm
  - Running time - $O(mn + n^2 \log n)$
  - m - edges, n - nodes
  - $m = O(n)$
  - $O(n^2 \log n)$



# Dynamic map visualization

- dynamic relational data
    - Research collaboration
    - last fm
    - social network



# Dynamic map visualization

- how to combine different similarity metrics.

# Dynamic map visualization

- how to combine different similarity metrics.
- compute a good distance measure from multiple similarity metrics

# Dynamic map visualization

- how to combine different similarity metrics.
- compute a good distance measure from multiple similarity metrics
- assign weights to each metric and take a weighted sum

# Dynamic map visualization

- how to combine different similarity metrics.
- compute a good distance measure from multiple similarity metrics
- assign weights to each metric and take a weighted sum
- $w$ be a weight in the range from 0 to 100
- $M_1$ and $M_2$ be the two input matrices.
- compute the weighted matrix

$$M_w = \frac{(w \times M_1) + ((100 - w) \times M_2)}{100}.$$

# Dynamic map visualization

- how to combine different similarity metrics.
- compute a good distance measure from multiple similarity metrics
- assign weights to each metric and take a weighted sum
- $w$ be a weight in the range from 0 to 100
- $M_1$ and $M_2$ be the two input matrices.
- compute the weighted matrix

$$M_w = \frac{(w \times M_1) + ((100 - w) \times M_2)}{100}.$$

- Use slider to find the "optimum" weight.

# DEMO Setup

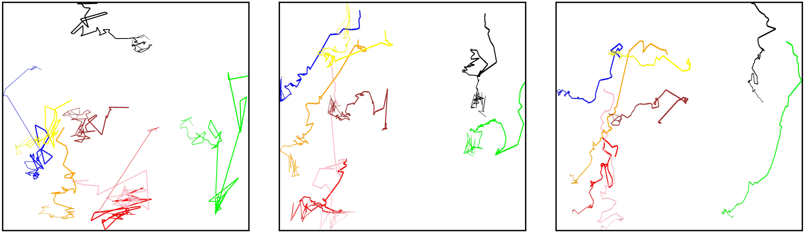
- Two input similarity matrices with distance measured using PCA analysis.
  - Y chromosome (NRY)
  - mitochondrial DNA (mtDNA)
- Number of nodes = 45

# DEMO Setup

- Two input similarity matrices with distance measured using PCA analysis.
  - Y chromosome (NRY)
  - mitochondrial DNA (mtDNA)
- Number of nodes = 45
- DEMO

# Layout Stability

- Trajectories of randomly selected nodes
- Compared the average distance travelled by one node in one iteration.
  - independent layout - 21.41 pixels.
  - layout initialized with positions from the previous frame - 13.19 pixels
  - Procrustes transformation - 8.43 pixels.



## Cluster Stability

- Rand $\rightarrow$ measure of cluster similarity
- C and $C^{'}$ $\rightarrow$ two clusterings of a graph $G$

# Cluster Stability

- Rand $\rightarrow$ measure of cluster similarity
- C and $C^{'}$ $\rightarrow$ two clusterings of a graph $G$
- $S_{11}$ $\rightarrow$ the set of pairs clustered together
- $S_{00}$ the set of pairs in different clusters

$$rand(C, C^{'}) = 1 - \frac{2(|S_{11}| + |S_{00}|)}{n(n-1)} \qquad (2)$$

# Cluster Stability

- Rand $\rightarrow$ measure of cluster similarity
- C and $C^{'}$ $\rightarrow$ two clusterings of a graph $G$
- $S_{11}$ $\rightarrow$ the set of pairs clustered together
- $S_{00}$ the set of pairs in different clusters

$$rand(C, C^{'}) = 1 - \frac{2(|S_{11}| + |S_{00}|)}{n(n-1)} \tag{2}$$

- 0 if the two clusterings are identical
- 1 if one clustering is a singleton clusters and the other one with all nodes in the same cluster.

# Cluster Stability

- Rand measure between each pair of successive iterations and averaged these values over all successive pairs.

# Cluster Stability

- Rand measure between each pair of successive iterations and averaged these values over all successive pairs.
- Without our heuristic, value = 0.0631
- With the heuristic, value = 0.0252.
- improvement of a little more than 60% with the heuristic.

# Conclusion

- a practical approach for visualizing
  - dynamic relational data
  - multiple relationships on the same data

# Conclusion

- a practical approach for visualizing
  - dynamic relational data
  - multiple relationships on the same data
- dynamic cluster stability - Modified modularity clustering algorithm
- color stability using MWM
- layout stability using affine transformations.

# Future Work

- formal evaluation of the effectiveness of our method

# Future Work

- formal evaluation of the effectiveness of our method
- combine more than 2 similarity metrics,

# Future Work

- formal evaluation of the effectiveness of our method
- combine more than 2 similarity metrics,
- deploy the system in a generic dynamic data setting.


Thank You!