

Interactive Correlation Clustering with Existential Cluster Constraints



Rico Angell



Nicholas Monath



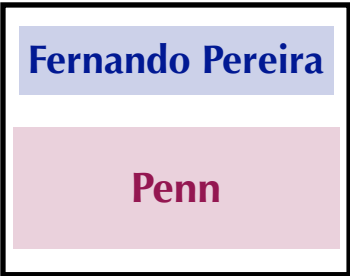
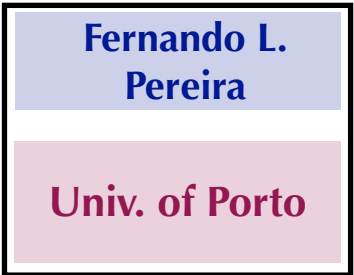
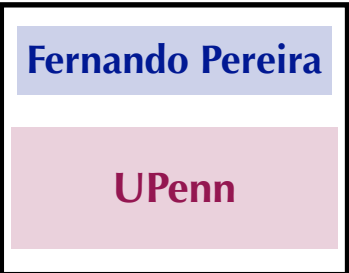
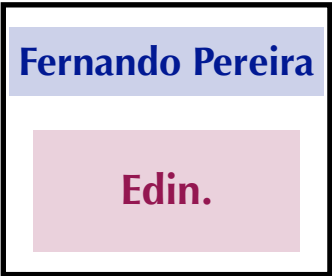
Nishant Yadav

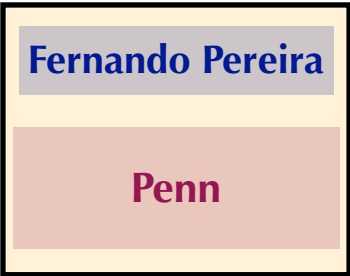
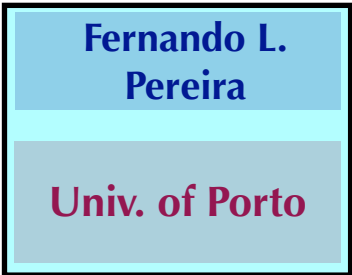
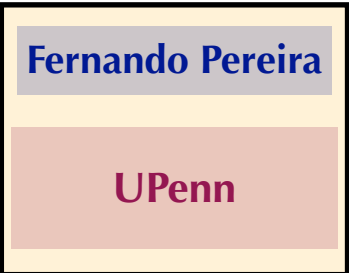
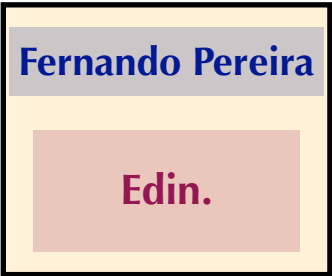


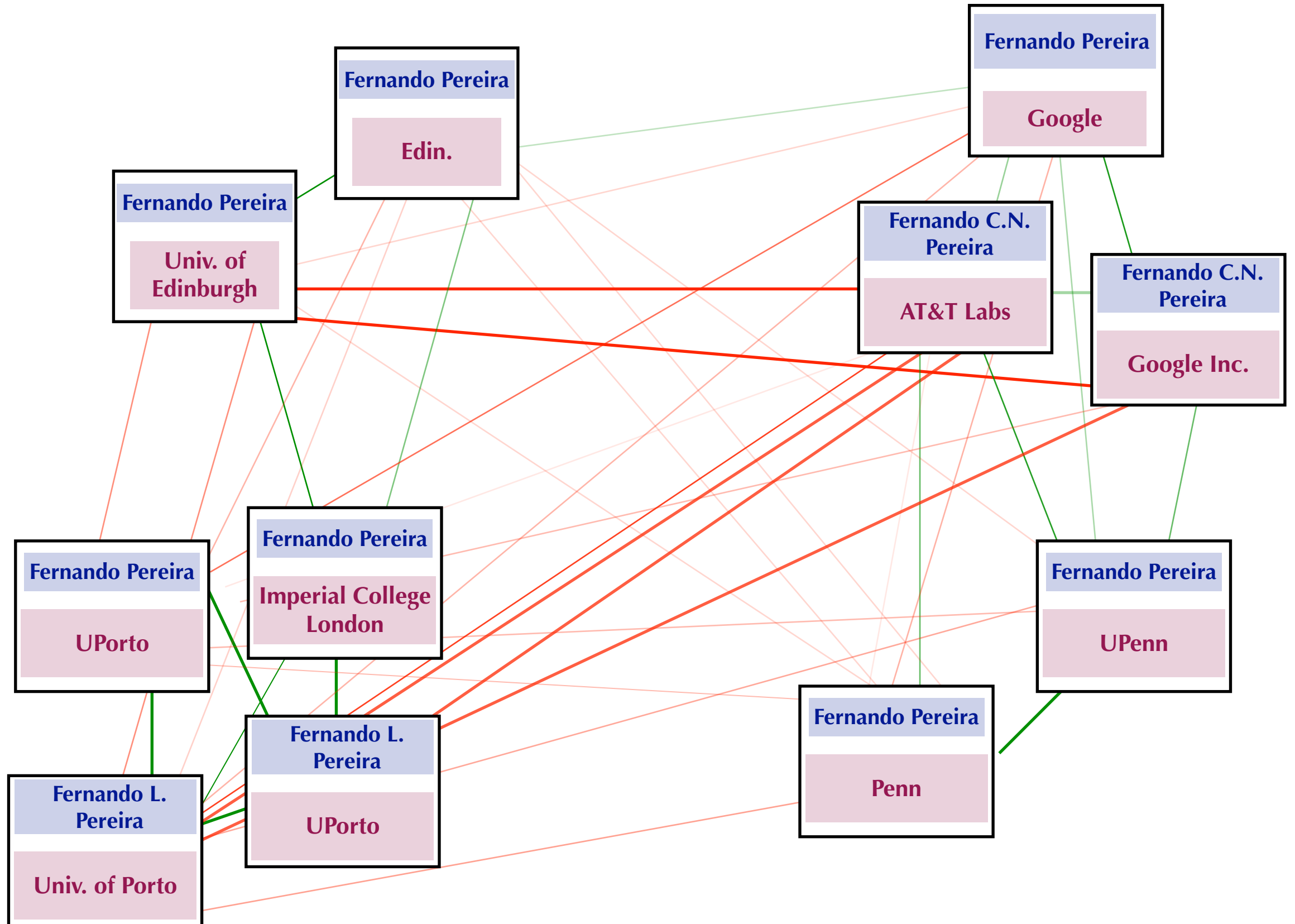
Andrew McCallum

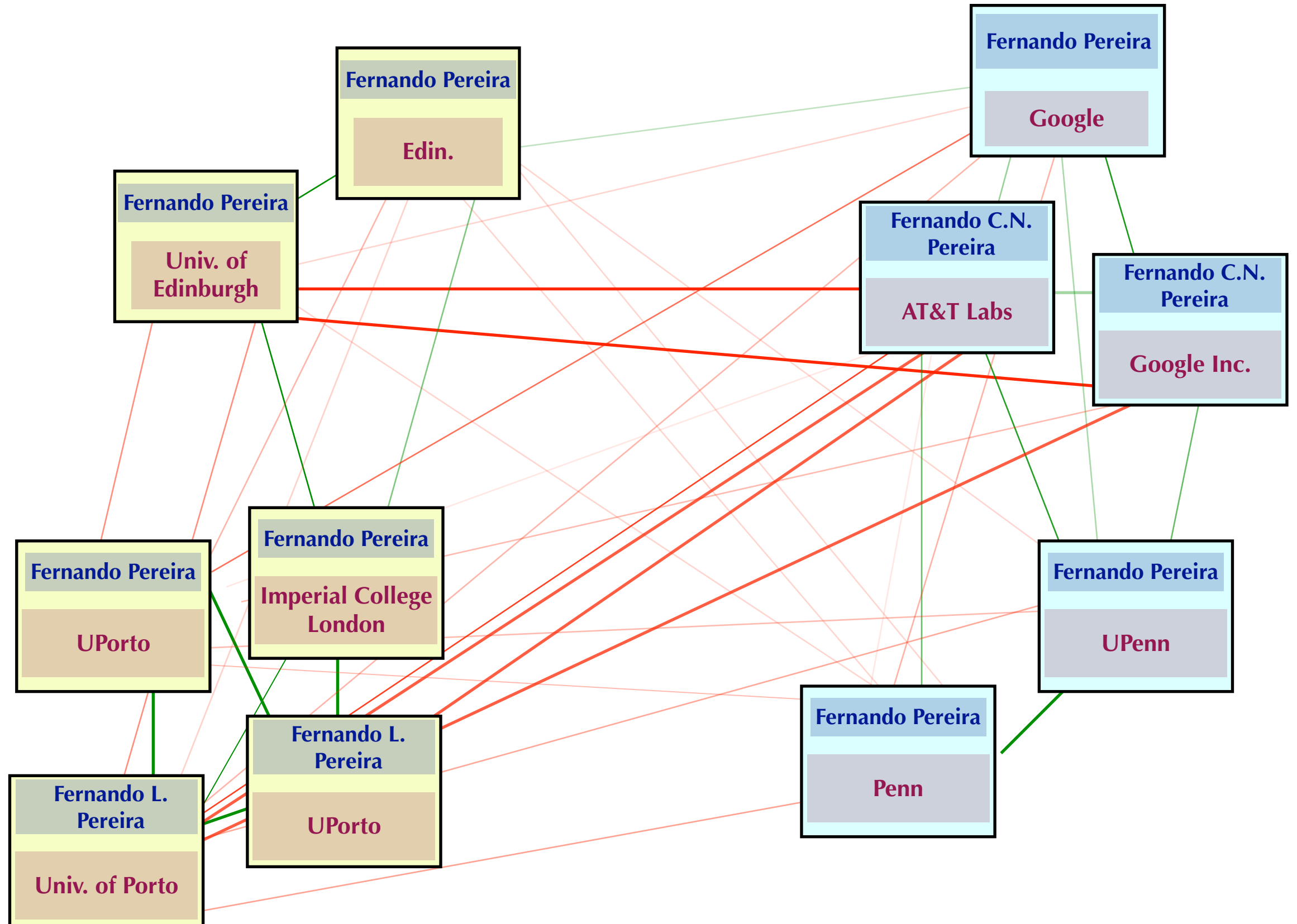
UMassAmherst

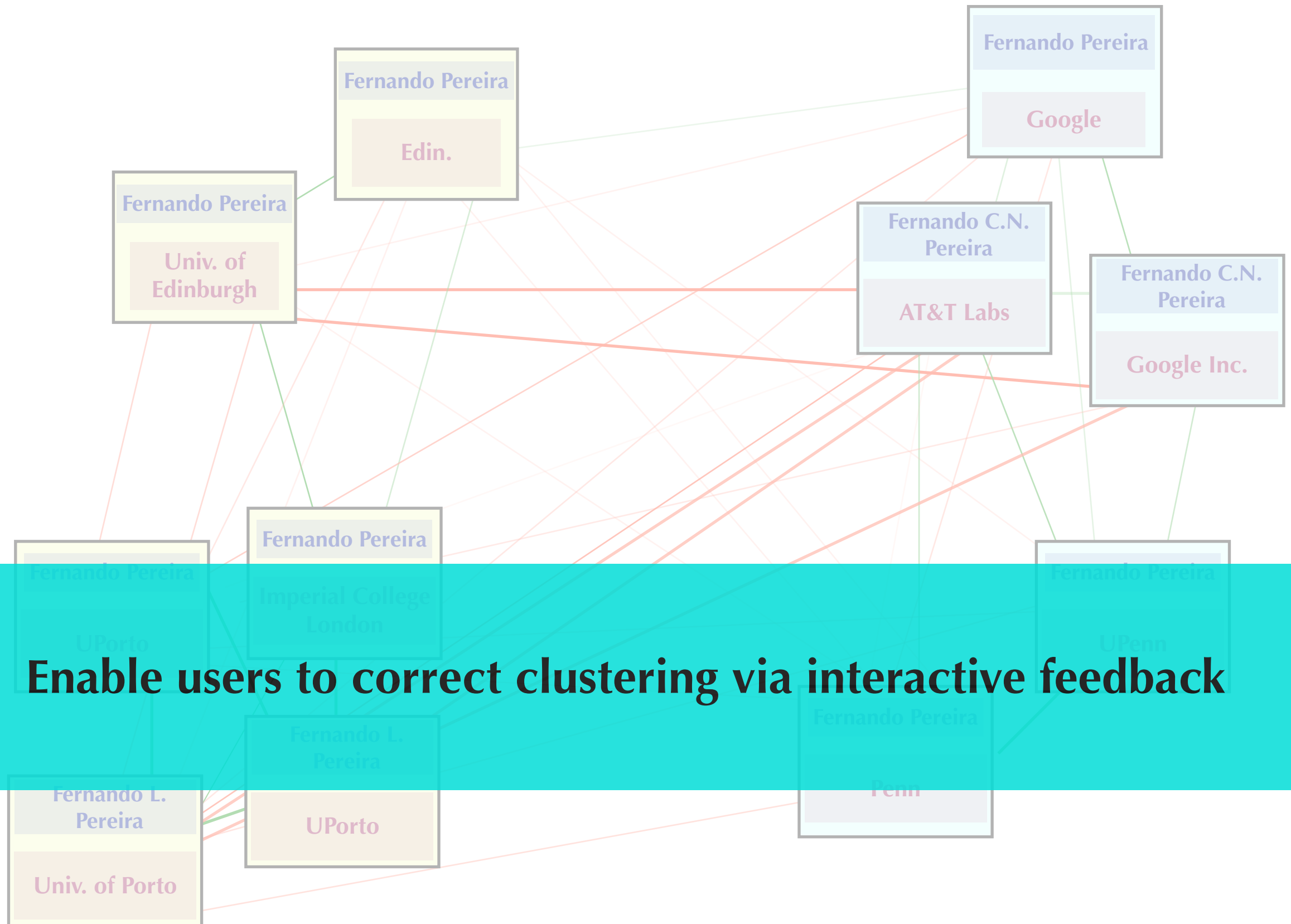
Manning College of Information
& Computer Sciences

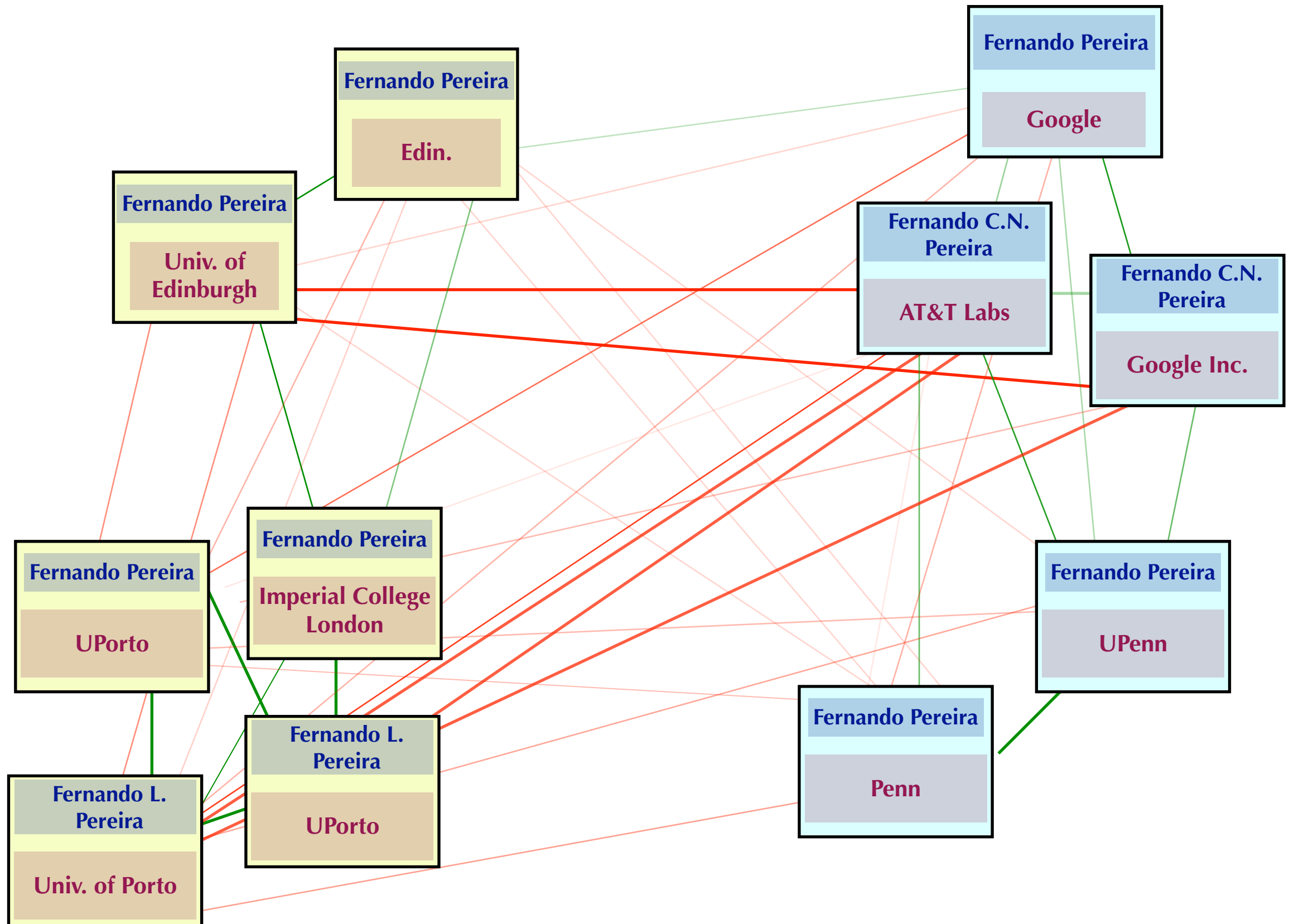


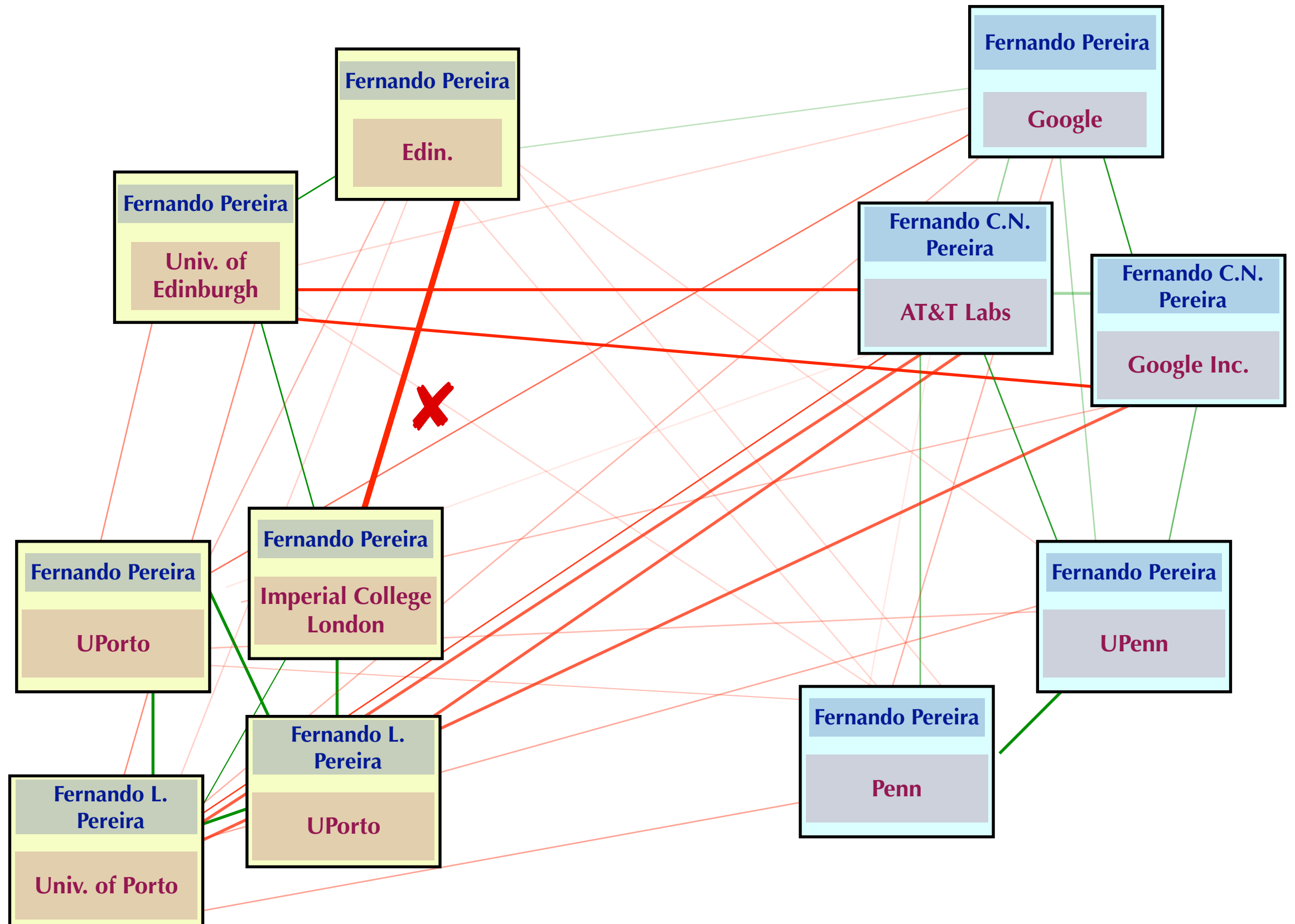


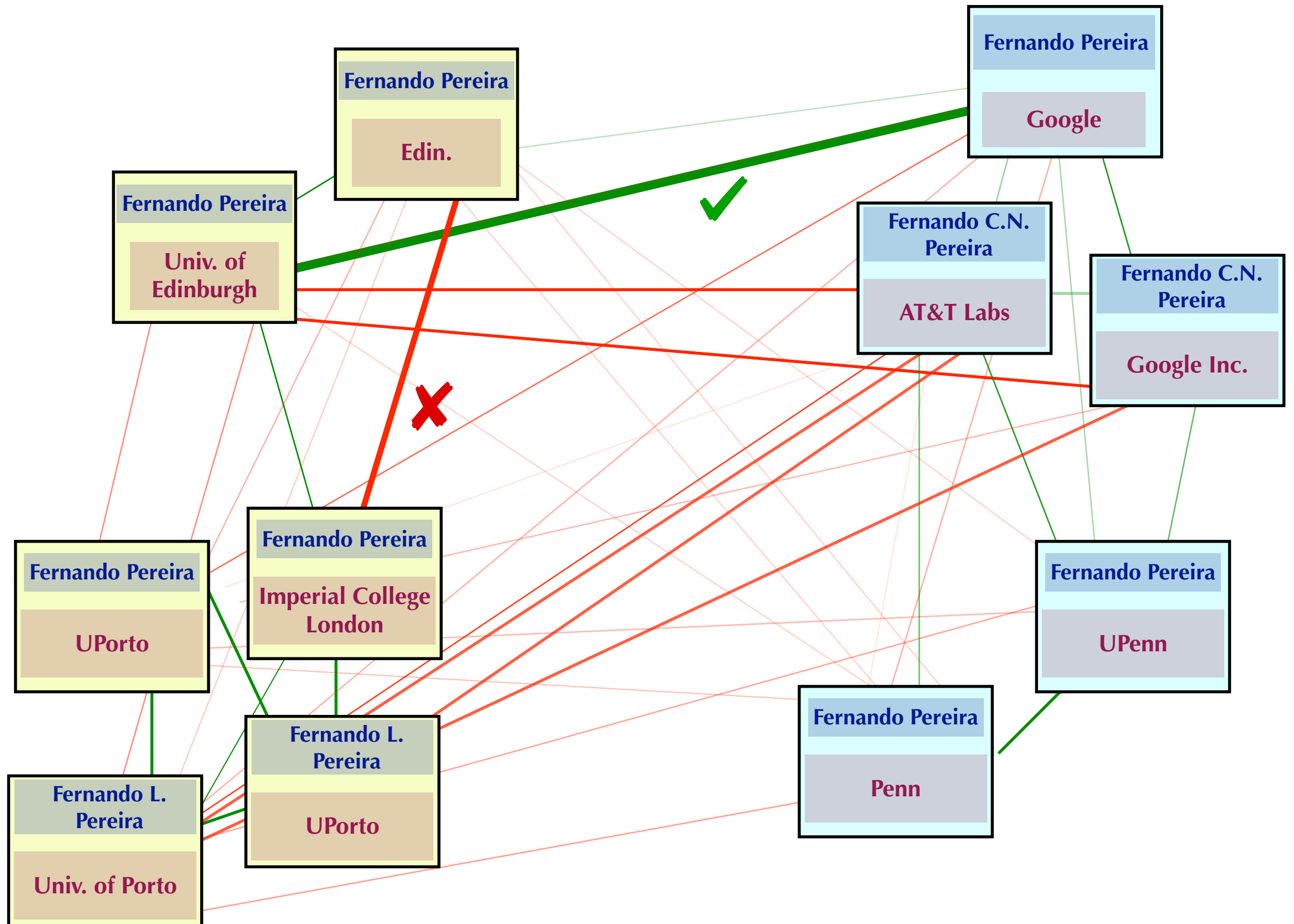


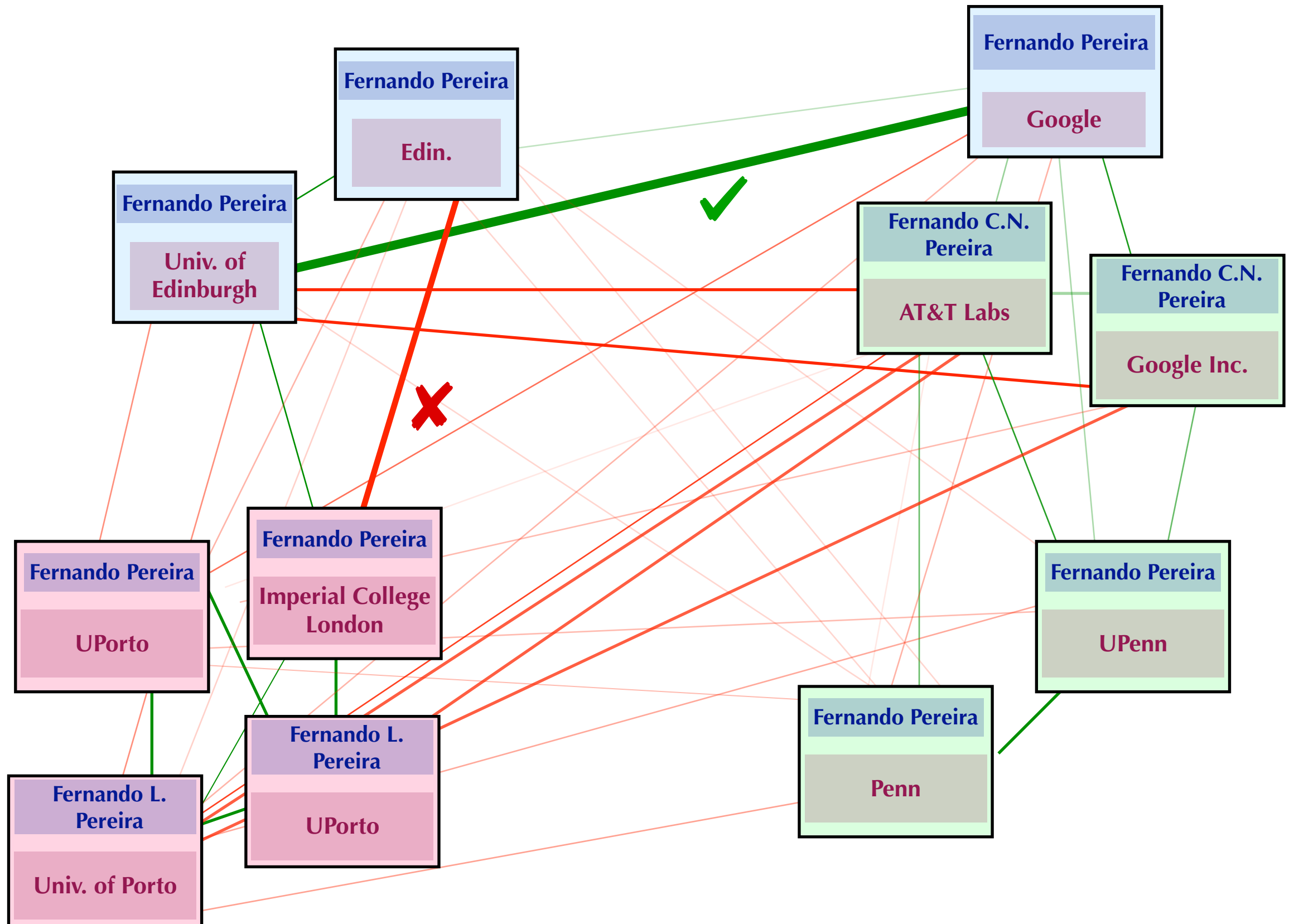


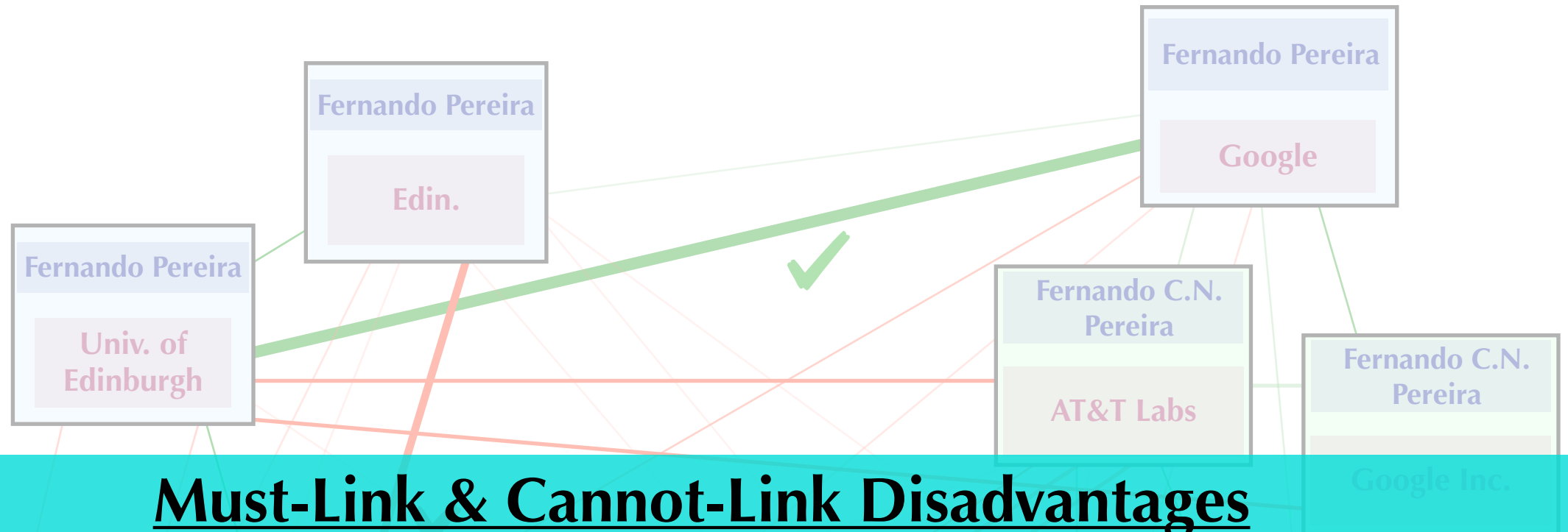






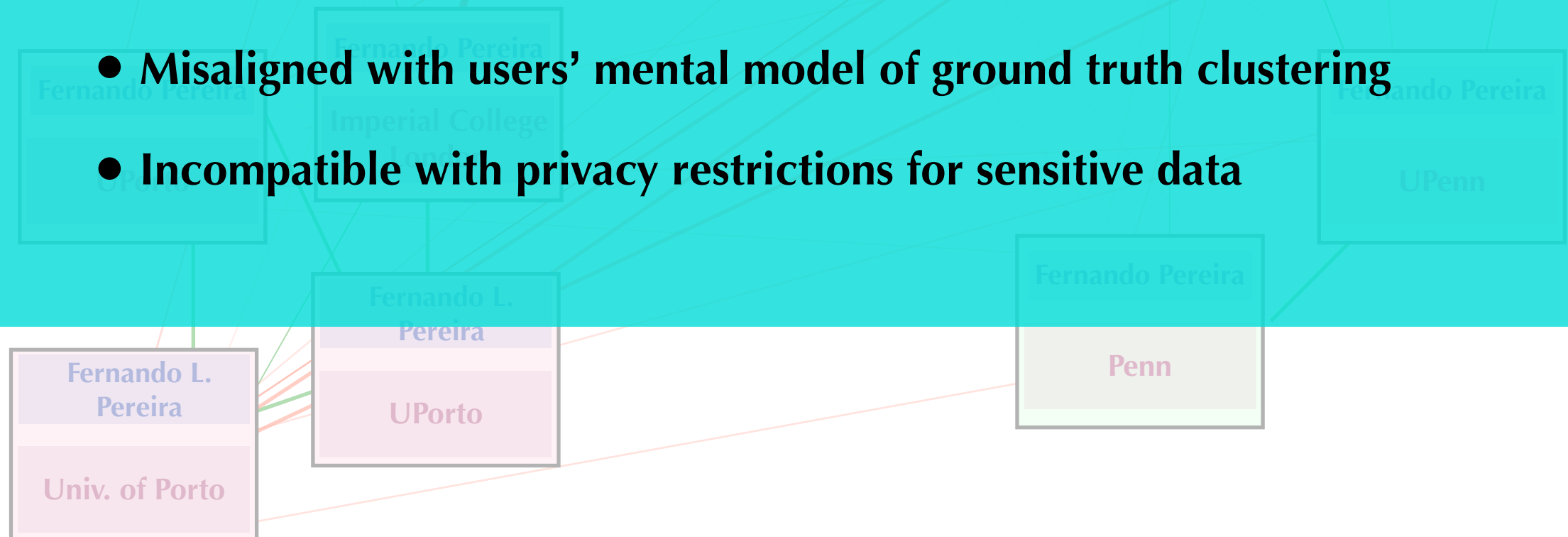


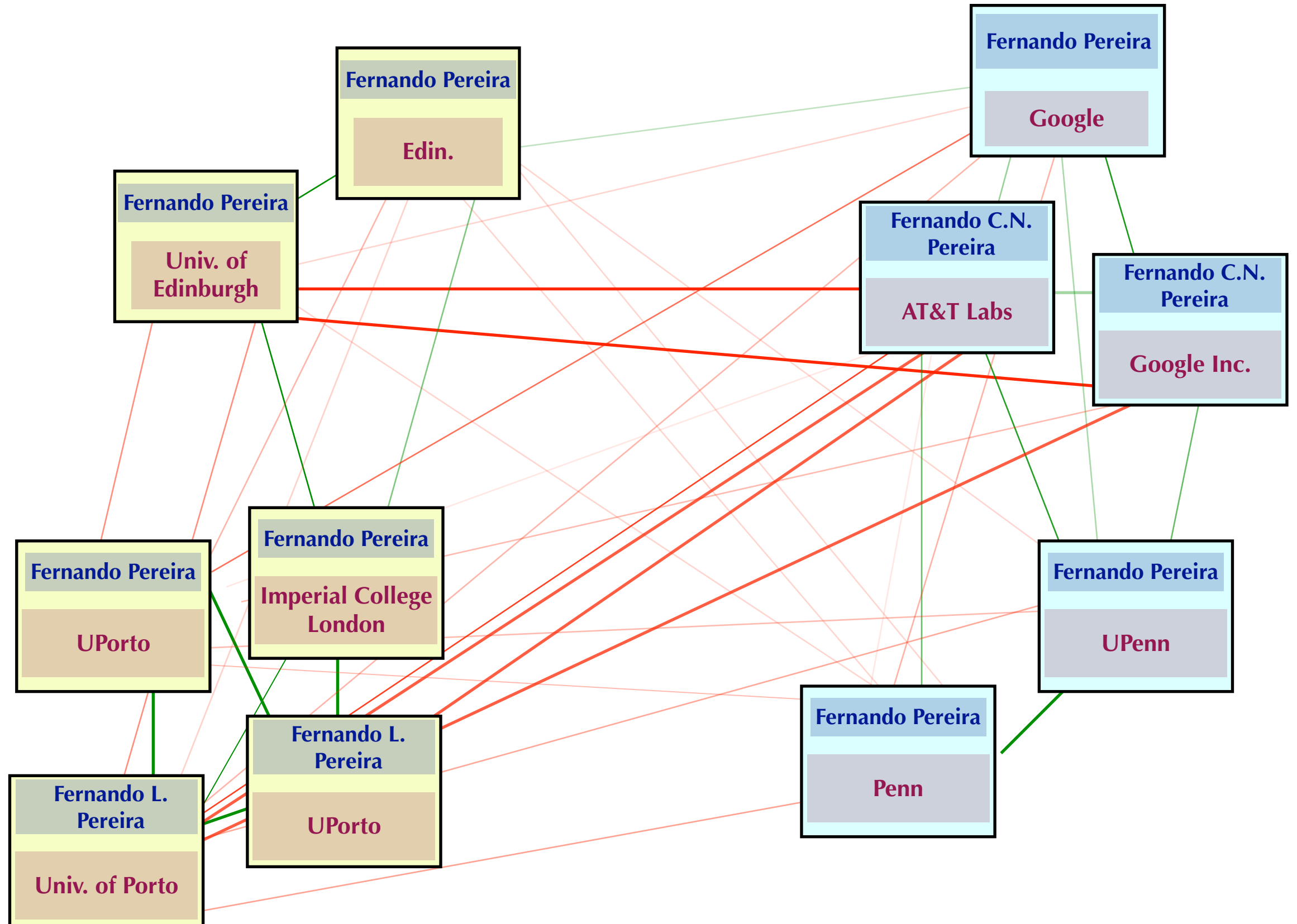


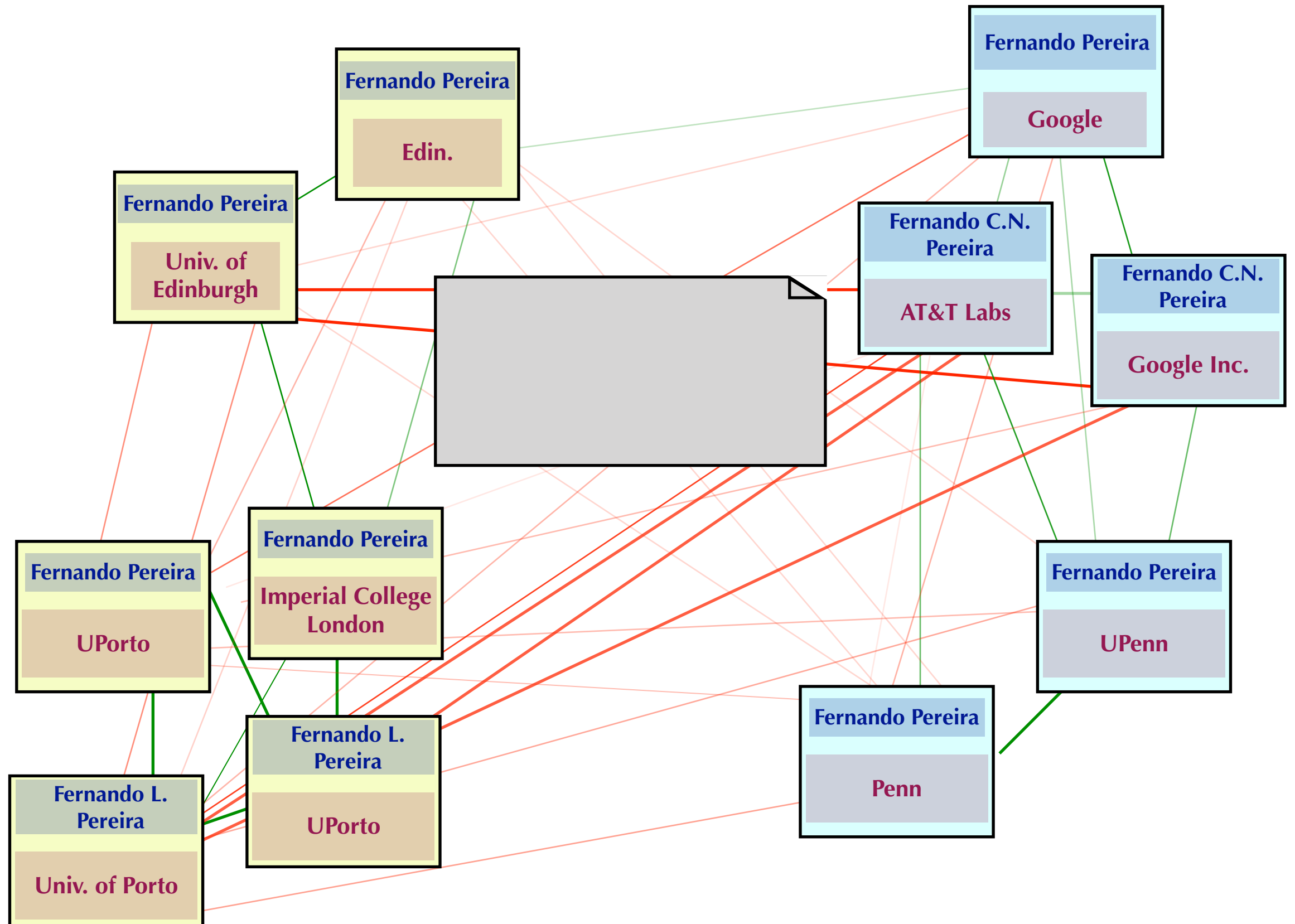


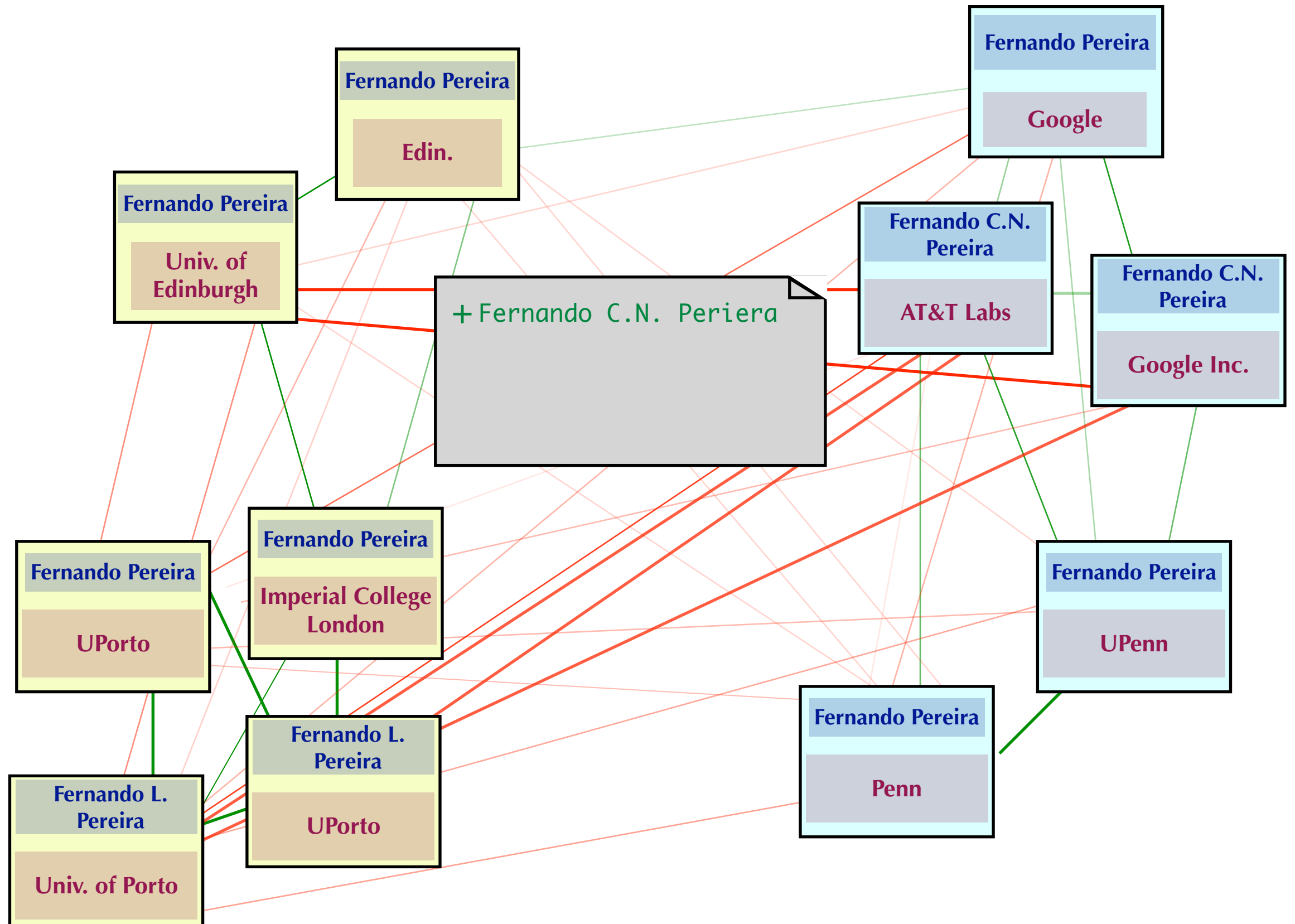
Must-Link & Cannot-Link Disadvantages

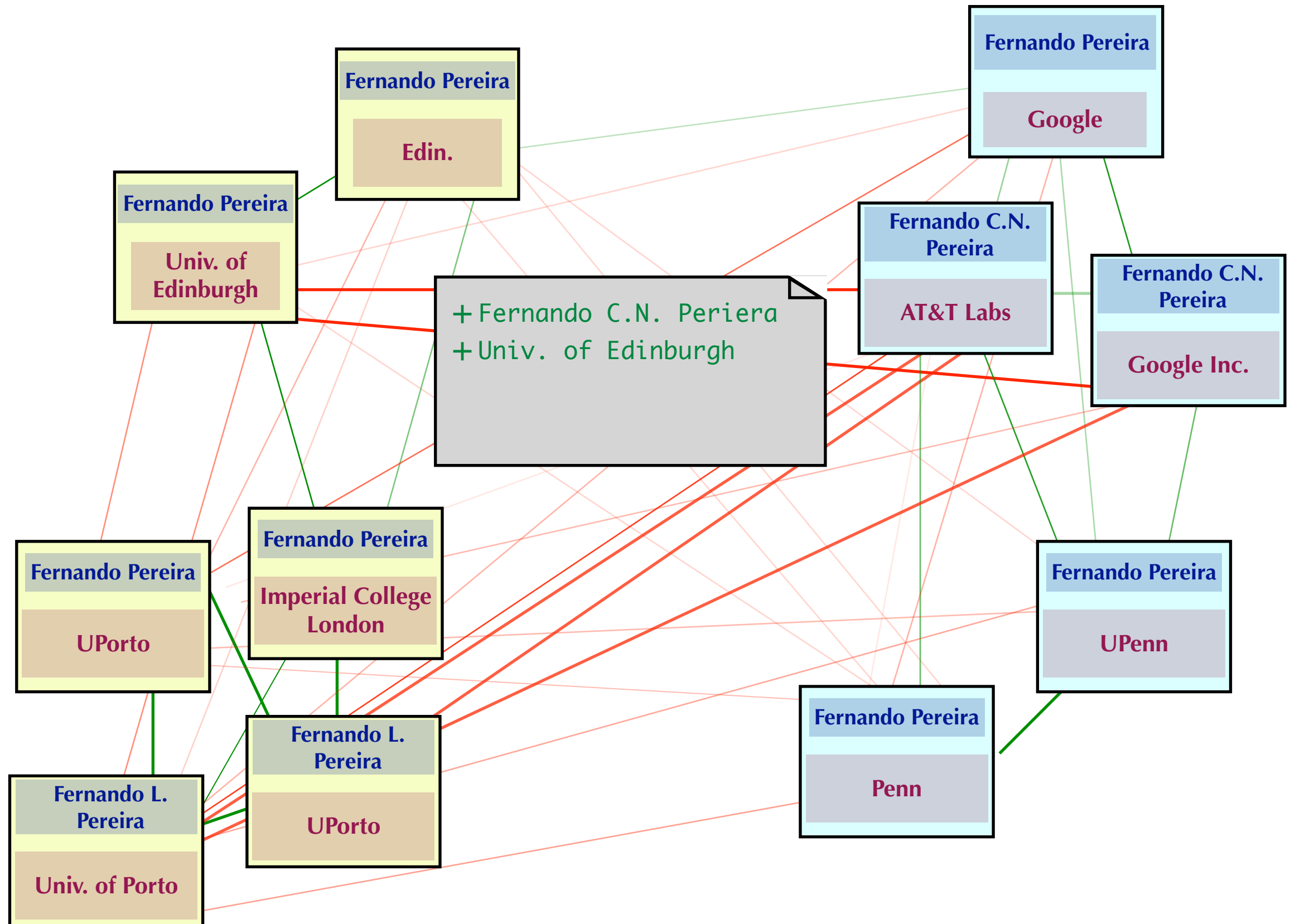
- Users may need to provide many constraints to achieve ground truth
- Misaligned with users' mental model of ground truth clustering
- Incompatible with privacy restrictions for sensitive data

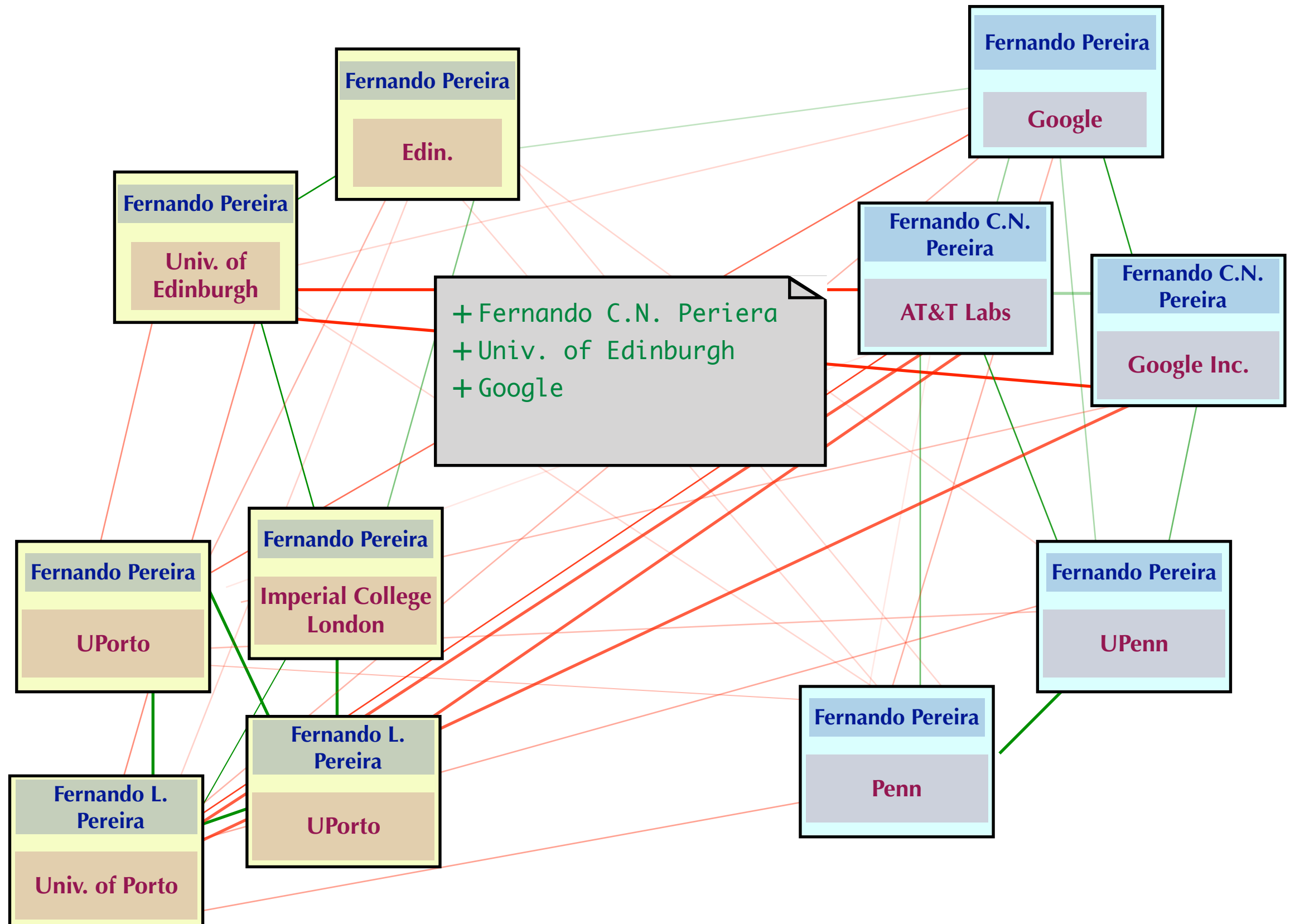


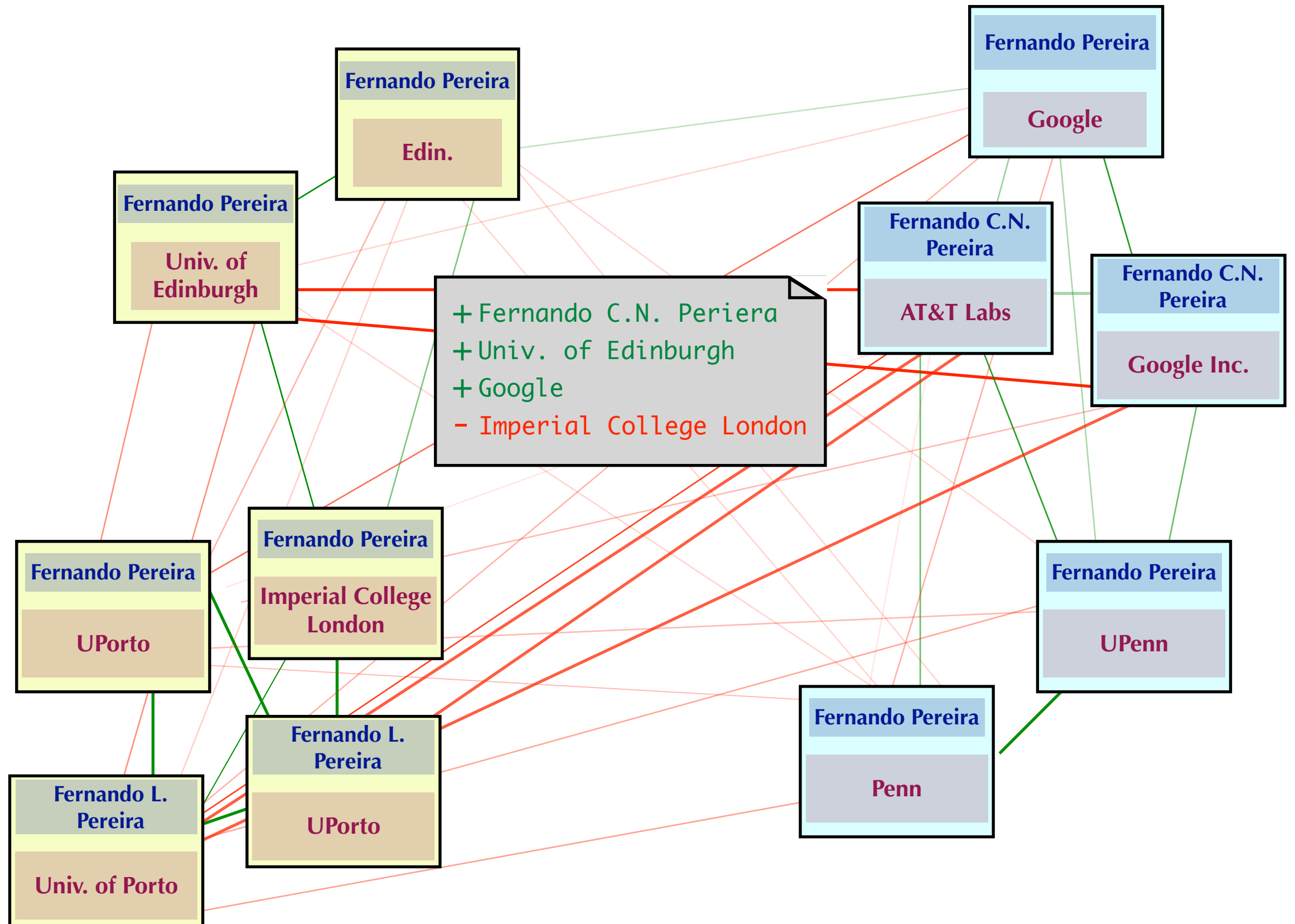


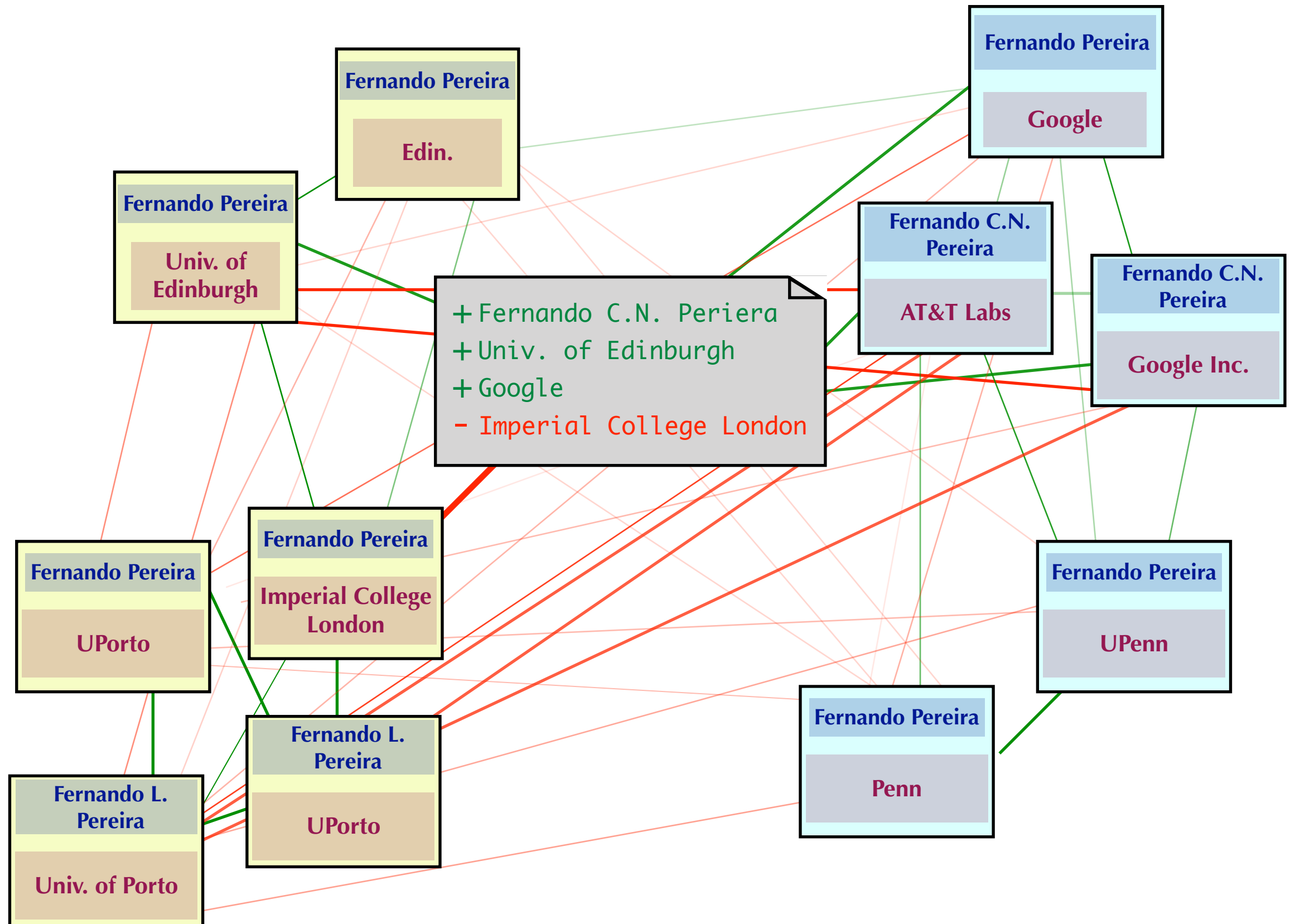


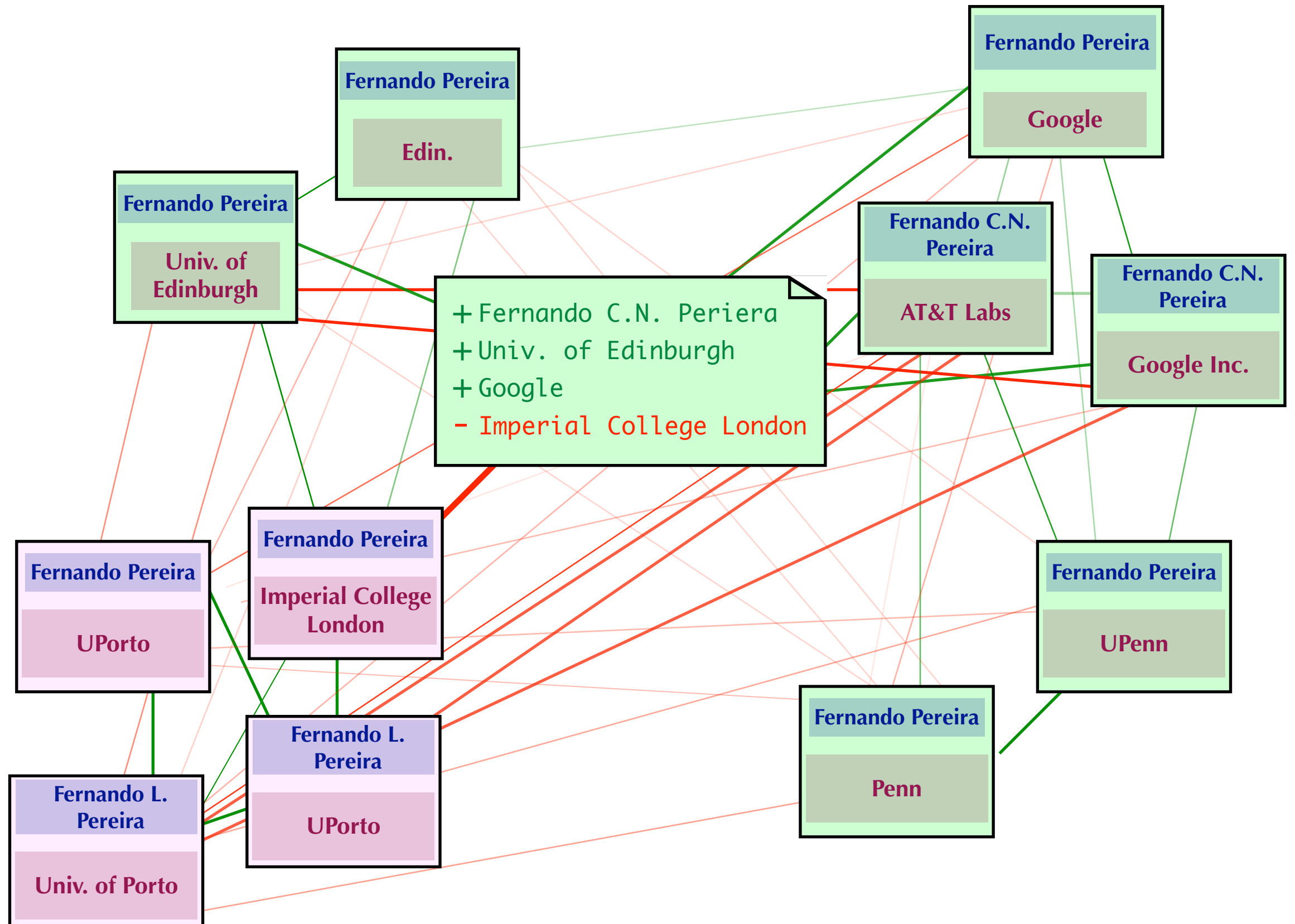


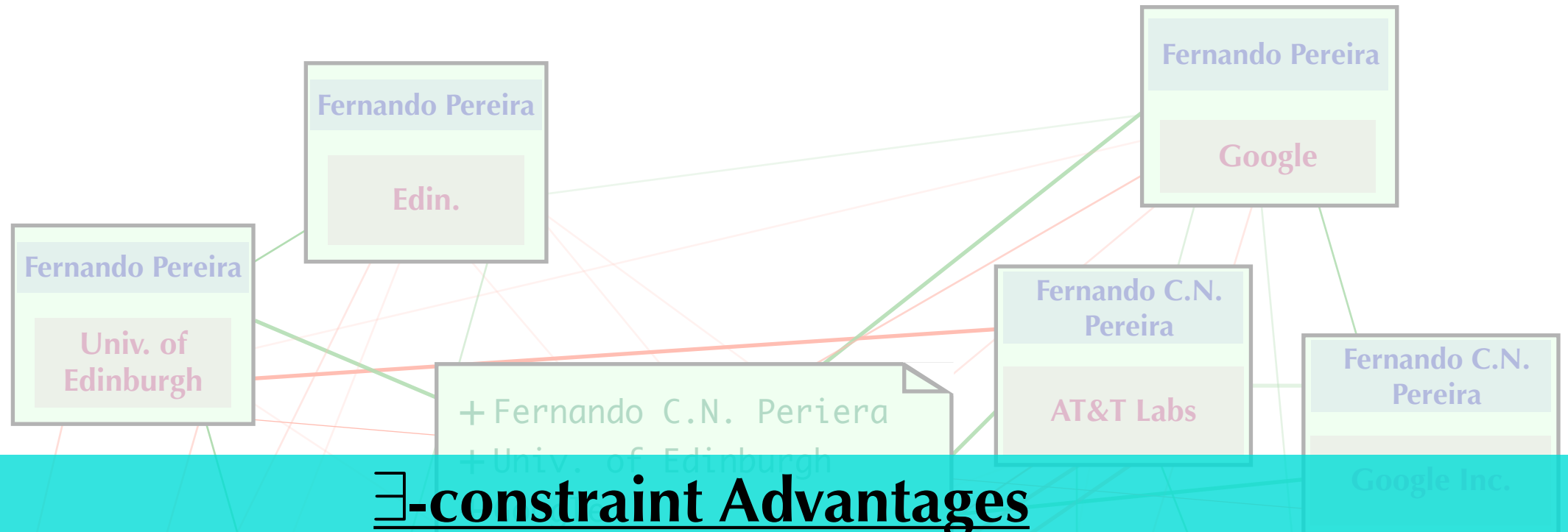






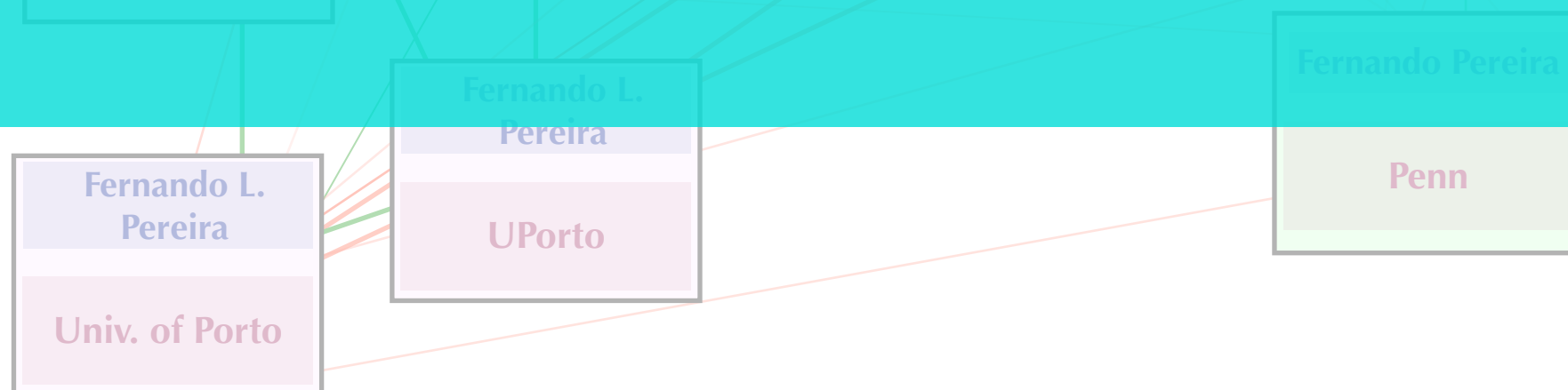






∃-constraint Advantages

- Express user intent more efficiently than point pairs
- More natural for users to express cluster-wise preferences
- Functions even when privacy restrictions prohibit users from seeing raw data



$$\begin{aligned}
& \max \sum_{(u,v) \in E} w_{uv} X_{uv} \\
& \text{s.t. } X \succeq 0, \\
& \quad X_{ss} = 1, \quad s \in V \cup \Xi \\
& \quad X_{st} \geq 0, \quad s, t \in V \cup \Xi \\
& \quad \sum_{v \in \mathbb{S}_\phi^\xi} X_{v\xi} \geq 1, \quad \phi \in \xi^+, \xi \in \Xi \\
& \quad X_{s\xi} = 0, \quad s \in V \cup \Xi, \xi \in \Xi : s \perp \xi
\end{aligned}$$

$$\begin{aligned}
& \max \sum_{(u,v) \in E} w_{uv} X_{uv} \\
& \text{s.t. } X \succeq 0, \\
& \quad X_{ss} = 1, \quad s \in V \cup \Xi \\
& \quad X_{st} \geq 0, \quad s, t \in V \cup \Xi \\
& \quad \sum_{v \in \mathbb{S}_\phi^\xi} X_{v\xi} \geq 1, \quad \phi \in \xi^+, \xi \in \Xi \\
& \quad X_{s\xi} = 0, \quad s \in V \cup \Xi, \xi \in \Xi : s \perp \xi
\end{aligned}$$

$$\max \sum_{(u,v) \in E} w_{uv} X_{uv}$$

$$\text{s.t. } X \succeq 0,$$

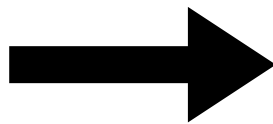
$$X_{ss} = 1, \quad s \in V \cup \Xi$$

$$X_{st} \geq 0, \quad s, t \in V \cup \Xi$$

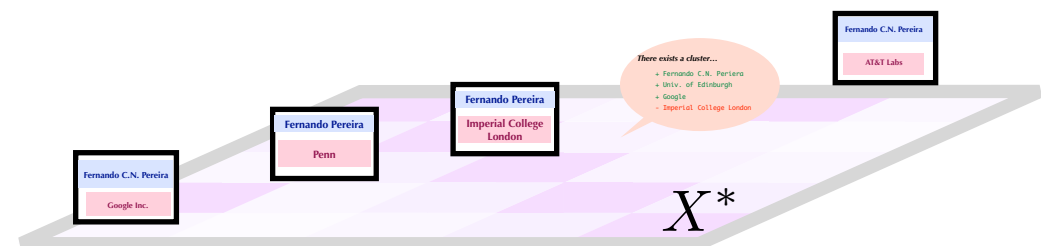
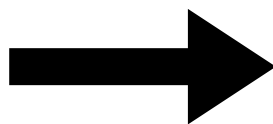
$$\sum_{v \in \mathbb{S}_\phi^\xi} X_{v\xi} \geq 1, \quad \phi \in \xi^+, \xi \in \Xi$$

$$X_{s\xi} = 0, \quad s \in V \cup \Xi, \xi \in \Xi : s \perp \xi$$

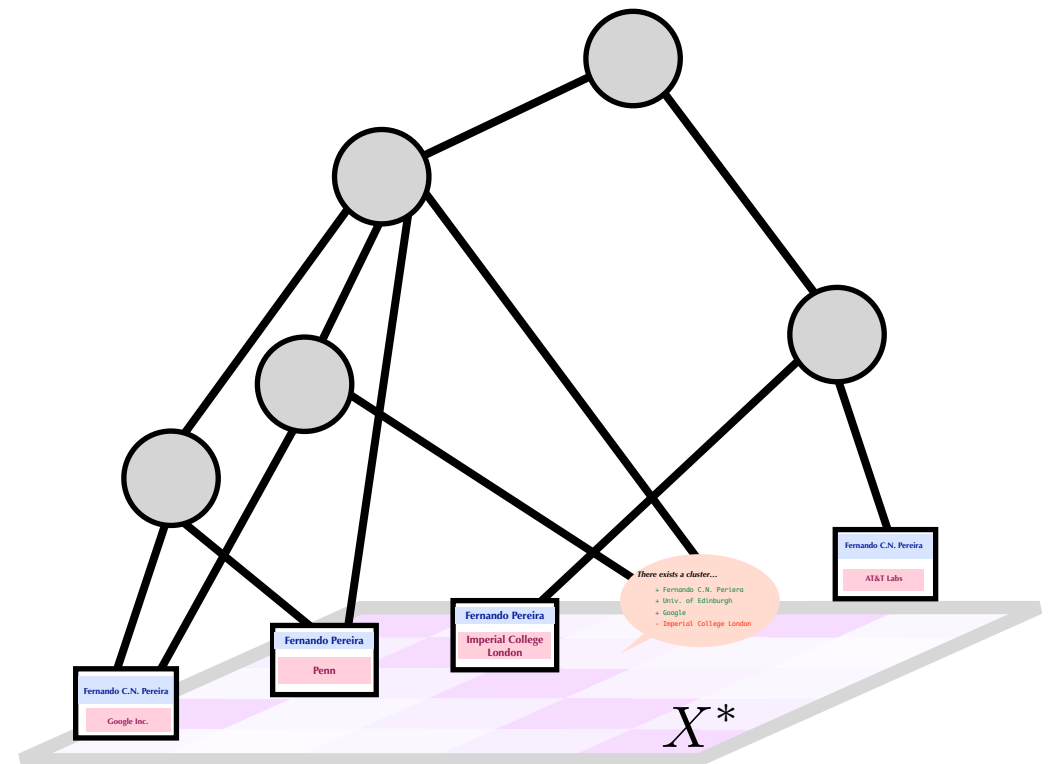
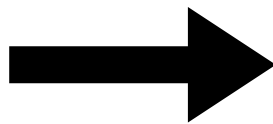
$$\begin{aligned}
& \max \sum_{(u,v) \in E} w_{uv} X_{uv} \\
& \text{s.t. } X \succeq 0, \\
& \quad X_{ss} = 1, \quad s \in V \cup \Xi \\
& \quad X_{st} \geq 0, \quad s, t \in V \cup \Xi \\
& \quad \sum_{v \in \mathbb{S}_\phi^\xi} X_{v\xi} \geq 1, \quad \phi \in \xi^+, \xi \in \Xi \\
& \quad X_{s\xi} = 0, \quad s \in V \cup \Xi, \xi \in \Xi : s \perp \xi
\end{aligned}$$



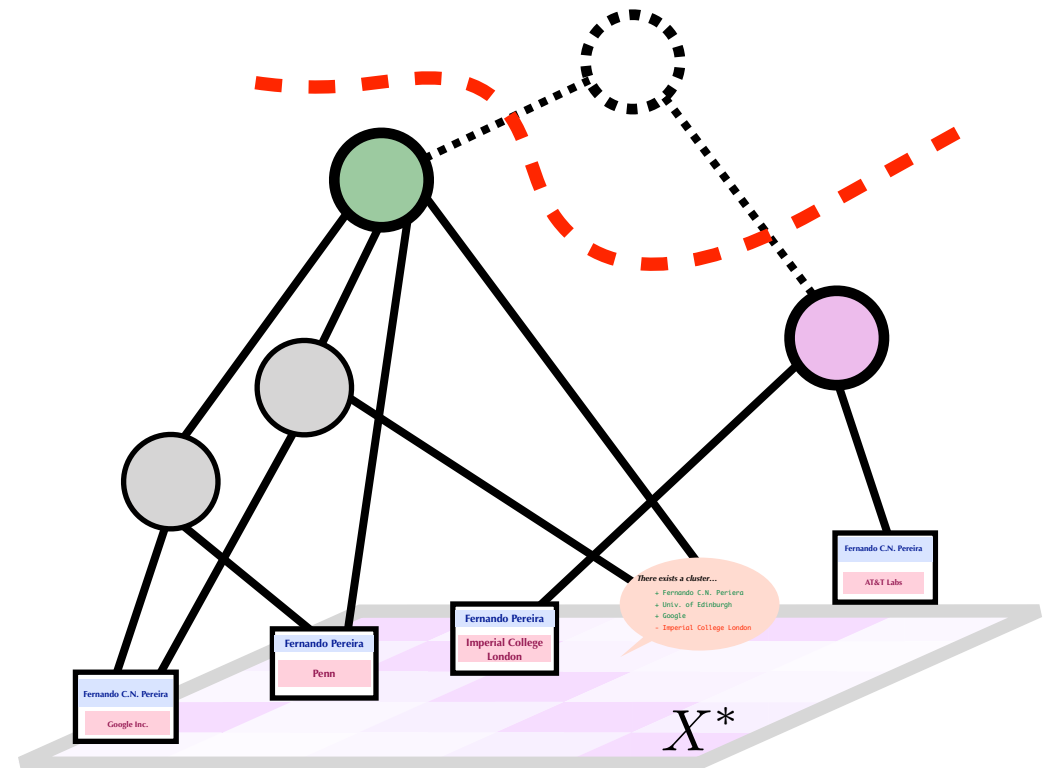
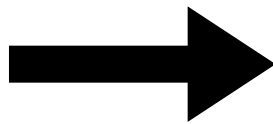
$$\begin{aligned}
& \max \sum_{(u,v) \in E} w_{uv} X_{uv} \\
& \text{s.t. } X \succeq 0, \\
& \quad X_{ss} = 1, \quad s \in V \cup \Xi \\
& \quad X_{st} \geq 0, \quad s, t \in V \cup \Xi \\
& \quad \sum_{v \in \mathbb{S}_\phi^\xi} X_{v\xi} \geq 1, \quad \phi \in \xi^+, \xi \in \Xi \\
& \quad X_{s\xi} = 0, \quad s \in V \cup \Xi, \xi \in \Xi : s \perp \xi
\end{aligned}$$



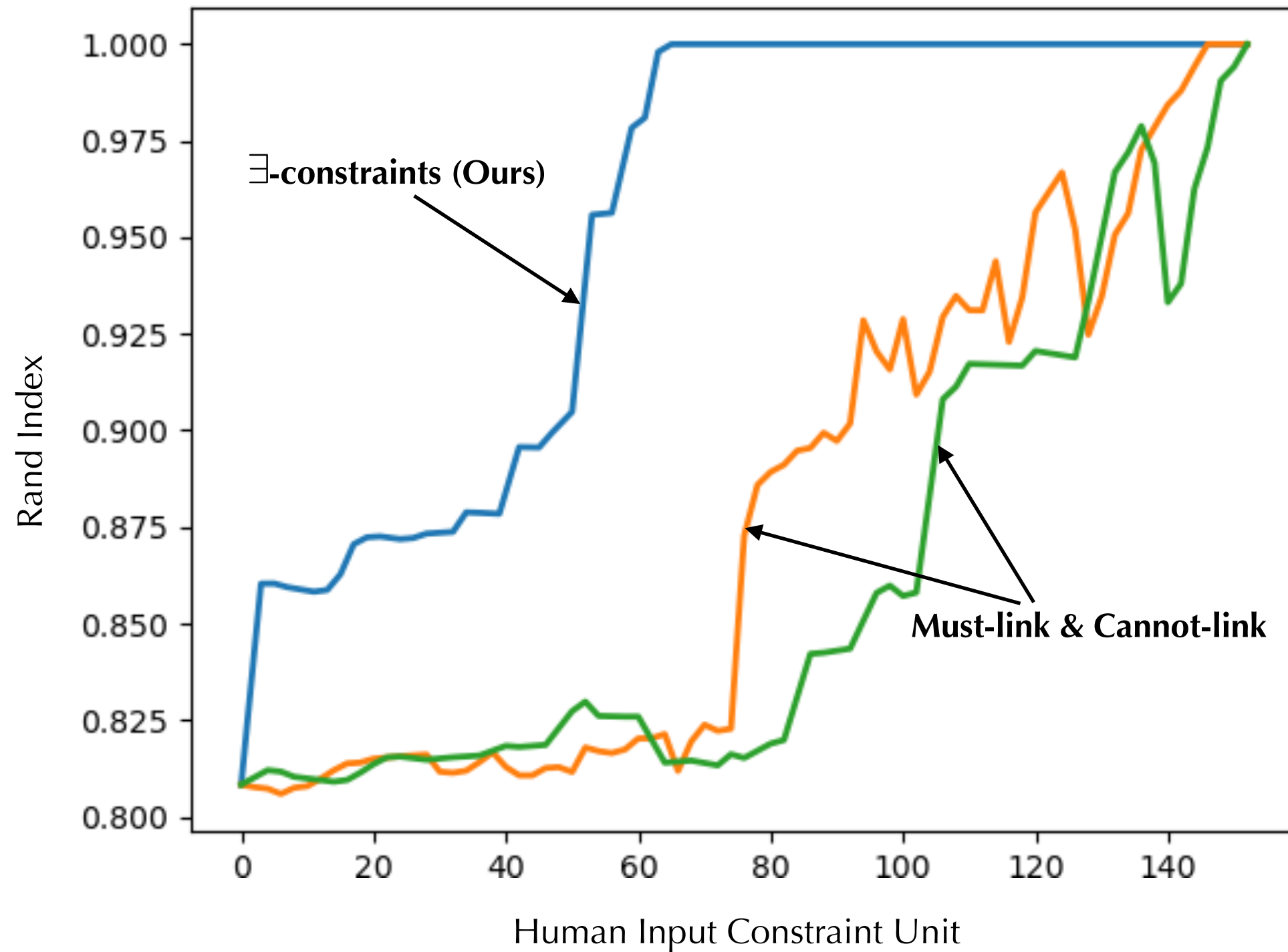
$$\begin{aligned}
& \max \sum_{(u,v) \in E} w_{uv} X_{uv} \\
& \text{s.t. } X \succeq 0, \\
& \quad X_{ss} = 1, \quad s \in V \cup \Xi \\
& \quad X_{st} \geq 0, \quad s, t \in V \cup \Xi \\
& \quad \sum_{v \in \mathcal{S}_\phi^\xi} X_{v\xi} \geq 1, \quad \phi \in \xi^+, \xi \in \Xi \\
& \quad X_{s\xi} = 0, \quad s \in V \cup \Xi, \xi \in \Xi : s \perp \xi
\end{aligned}$$



$$\begin{aligned}
& \max \sum_{(u,v) \in E} w_{uv} X_{uv} \\
& \text{s.t. } X \succeq 0, \\
& \quad X_{ss} = 1, \quad s \in V \cup \Xi \\
& \quad X_{st} \geq 0, \quad s, t \in V \cup \Xi \\
& \quad \sum_{v \in \mathcal{S}_\phi^\xi} X_{v\xi} \geq 1, \quad \phi \in \xi^+, \xi \in \Xi \\
& \quad X_{s\xi} = 0, \quad s \in V \cup \Xi, \xi \in \Xi : s \perp \xi
\end{aligned}$$

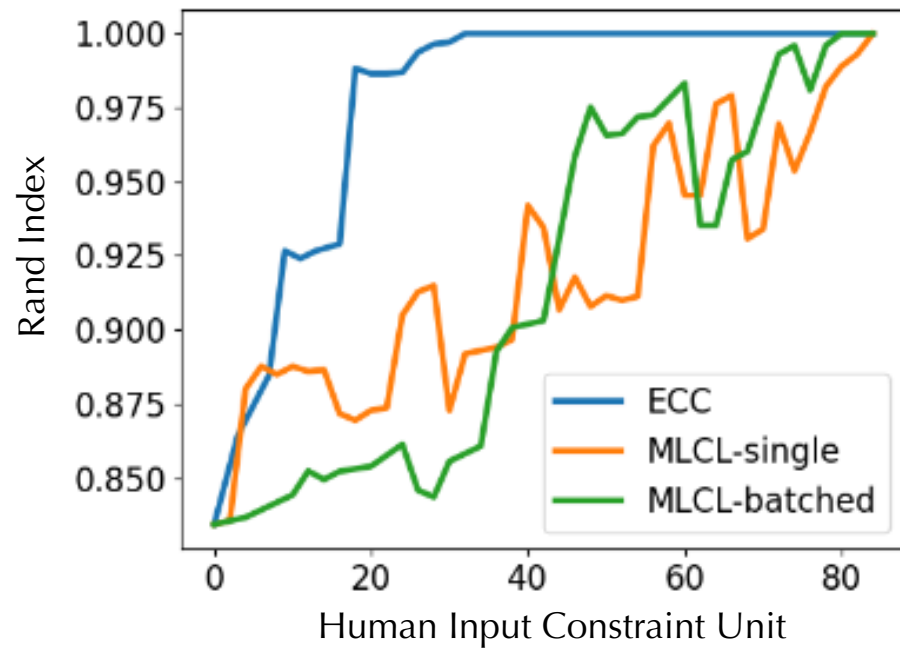


Empirical Efficiency of Feedback

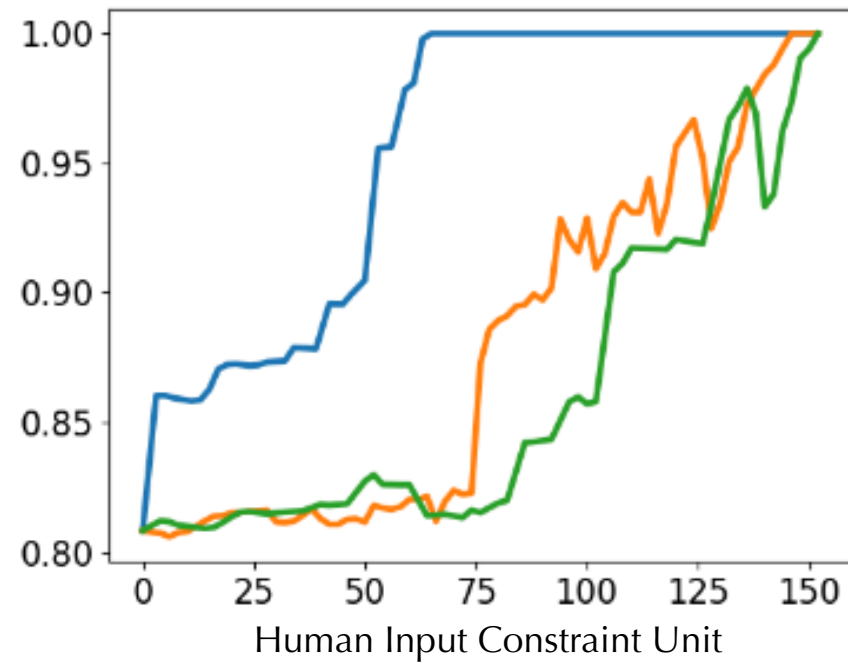


Empirical Efficiency of Feedback

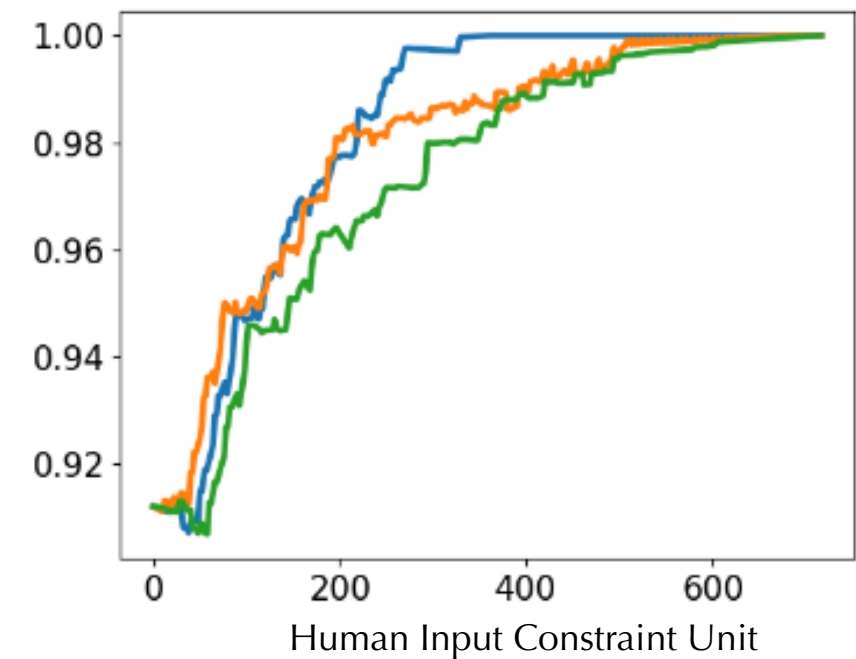
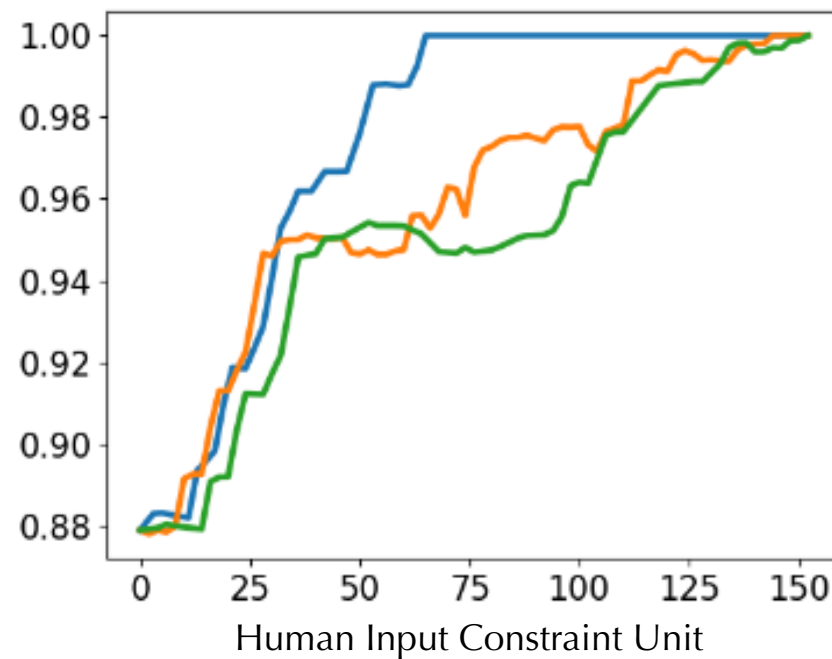
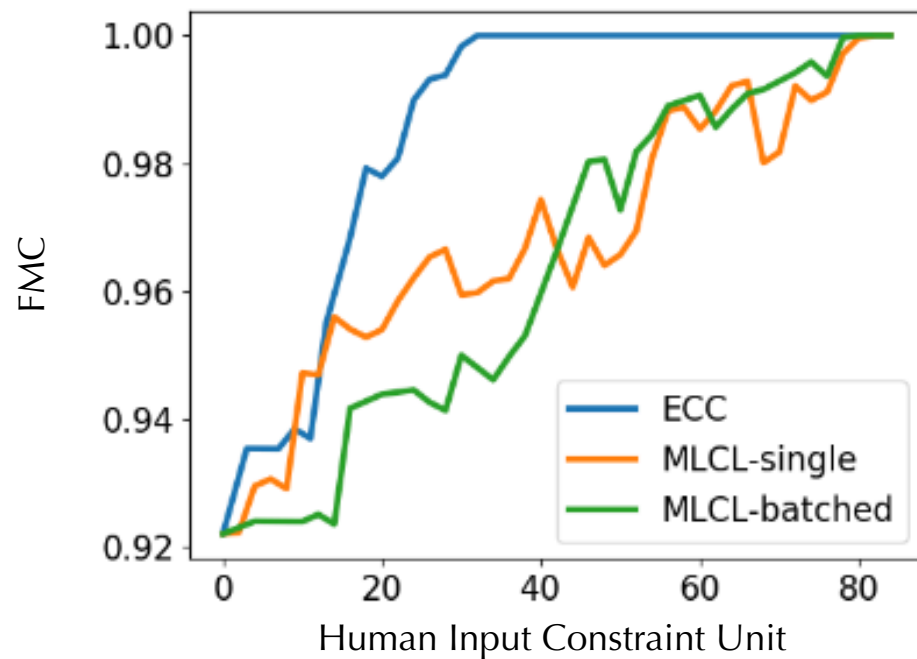
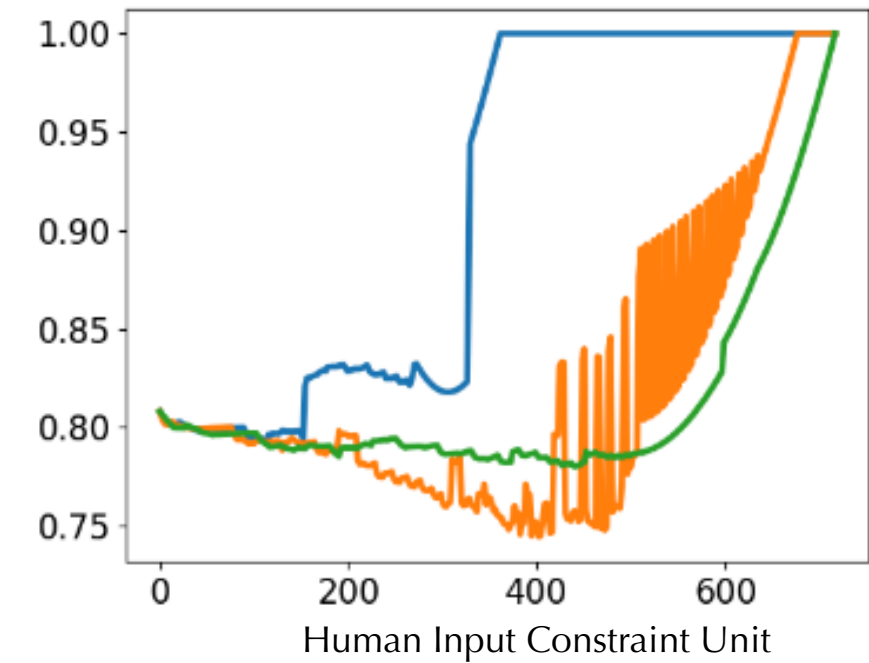
PubMed



QIAN

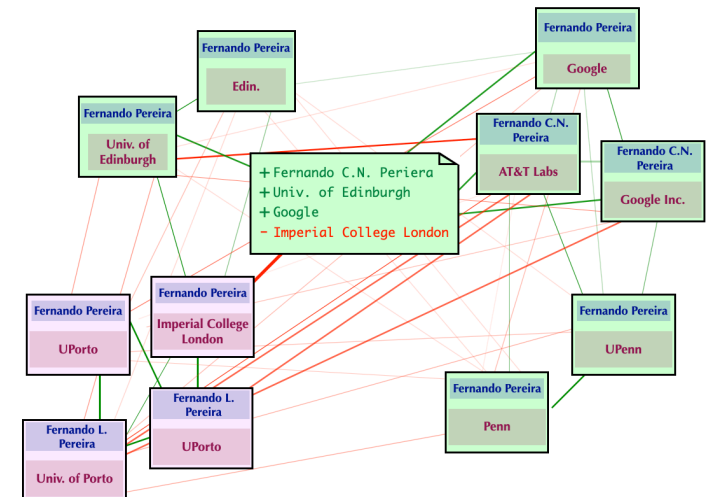


SCAD-zbMATH



Novel Form of Feedback for Interactive Clustering

Users express feedback using statements about the existence of clusters having (and not having) particular discrete features.



Inference Algorithm

Jointly cluster data points and Ξ -constraints using a correlation clustering objective.

$$\begin{aligned}
 & \max \sum_{(u,v) \in E} w_{uv} X_{uv} \\
 & \text{s.t. } X \succeq 0, \\
 & \quad X_{ss} = 1, \quad s \in V \cup \Xi \\
 & \quad X_{st} \geq 0, \quad s, t \in V \cup \Xi \\
 & \quad \sum_{v \in S_\phi^\xi} X_{v\xi} \geq 1, \quad \phi \in \xi^+, \xi \in \Xi \\
 & \quad X_{s\xi} = 0, \quad s \in V \cup \Xi, \xi \in \Xi : s \perp \xi
 \end{aligned}$$

