

## Record Linkage Reconciliation of Arlington Department of Human Services Administrative Data Using Potts Models

Crandell, I<sup>1\*</sup>, Schroeder, A<sup>2</sup>, Higdon, D<sup>1</sup>, and Irwin, M-d<sup>3</sup>

<sup>1</sup>Social and Decision Analytics Laboratory, Biocomplexity Institute of Virginia Tech

<sup>2</sup>Social and Decision Analytics Division, Biocomplexity Institute of University of Virginia

<sup>3</sup>Arlington County Department of Human Services

Situated at the nexus of federal, state, and local governments, the Arlington Department of Human Services (DHS) receives service utilization data from a multitude of different sources. Because of their “no wrong door” policy, customers can sign up for any DHS service from any DHS department. A practical consequence of this is that a single person can appear as multiple records from multiple databases with no unambiguous key between these records. Merging these records requires a probabilistic linkage approach. Classical approaches to record linkage, such as the method of Fellegi and Sunter, consider each possible pair of records between databases and assigning link probabilities to each one. A drawback of considering pairwise links alone is that sometimes the transitive nature of links is violated. In order to better handle such information clashes, we propose a Bayesian linkage method that considers a large set of possible pairs at once. At the heart of this approach is a Potts model representation that tracks which records are assigned to the same individual. This allows us to assign probabilities to the various reconciliations of inconsistent linkage assignments.

\*Corresponding Author:

Email Address: [ian85@vt.edu](mailto:ian85@vt.edu) (I Crandell)

