Key Points from Day 2 (Wednesday, January 18) [DRAFT]

We must grapple with an inherent trade-off between statistical accuracy and privacy preservation.

1. The more we want our data to tell us, the more noise will be introduced into a given synthetic data set.
2. Using differential privacy (DP) synthetic data algorithms efficiently and effectively requires that we (1) clearly define the minimum statistics we want to preserve from the original data, and (2) ensure some available validation check is in place.
3. Privacy concerns are inherent in all graphed data. Node-private algorithms offer one way of facilitating more precise analysis of social networks while simultaneously privatizing graphical network data.

Maintaining data privacy has instrumental value for both the individual and their collective network.

1. Potential threats to the privacy of large-scale data collection and surveillance can vary by time and degree of impact.
2. Mechanisms for mitigating privacy threats will require weighing the shorter versus longer term potential harms, as well as the material and immaterial potential harms.

Key Points Re: MAPPING@Brown

1. We should use a data minimization approach for the initial phases of MAPPS to demonstrate feasibility of our approach, prevent the collection of unneeded data, and help assuage concerns (e.g., from Brown leadership, NSF) related to data privacy.
2. Thinking of data as sensitive vs. non-sensitive may not be the most useful framing. Rather, we should think of all data as potentially sensitive and potentially identifiable, and develop our data collection tools accordingly.
3. For very precise statistics (e.g., number of contacts per day), we already have the tools available to generate very precise statistics (e.g., number of contacts per day) using differential privacy techniques. For less precise statistics, there remain questions and opportunities for how best to apply DP algorithms, missing data techniques and other approaches to improve privacy while maximizing data accuracy.