## Paper 1: Data integration for estimation of contact network structure in infectious disease modeling

* Novel and well-studied infectious diseases have a multitude of data sources that describe different elements of the larger process of disease spread. We introduce a system for data integration – bringing these disparate data sets together to make broader inferences about the process of disease spread than can be reached from a single source.
  + Ability to leverage totality of the data we have
* Expand upon existing simultaneous models of contact network and disease spread by switching from ERGM to congruence class models.
  + Allows us to incorporate behavioral survey data, which is not possible in ERGM (best you can do is set starting point for Gibbs sampler).
  + Can set full priors on network terms
* One key element of setting up the model is defining priors. This is a complex question when modeling network structures. We explore strategies for setting building informative priors based on sampled network data, in particular how to define the “information equivalent” of a dependent sample of actors in a network.
* Question: What is our demonstration here?
  + Simulation-based demonstration of parameter recovery?
    - Rate of convergence to truth as sample coverage increases
    - Comparing unrestricted prior, prior restricted to valid degree distributions, posterior (did adding more data improve inference?)
  + Simulation addressing “information equivalent” question?
    - Krivitsky & Kolaczyk paper – adapt their ERGM-based simulation?
  + COVID data from UCSD:
    - proximity data on small sample of individuals (~70)
    - Mixing on-campus and off-campus contacts
    - Contact-tracing data
  + Atrium COVID contact-tracing data
* Notes:
  + Keep it general – allow for wide ranging applications
  + Use SEIR epidemic model

## Paper 2: Bias correction for egocentrically-estimated network terms based on survey data

* While we can theoretically measure several network terms from egocentric data collection (attribute-based mixing, degree distribution), survey data can be biased. This is particularly problematic when the contact network of interest involves sensitive behaviors like sex or injection drug use.
* Mochudi data used for development of BCPP study – we know the levels of concurrent partnership reported in surveys were far too low
  + ~95% reported on lifetime partner. Implemented as 1st ptn, lots of questions, do you have a 2nd most recent, lots of questions…people quit
* However, we know a great deal about the dynamics of HIV transmission, and we have viral genetic data for this sample as well. We hypothesize that incorporating these more objective pieces of information could correct some of the bias in the survey data.
* Simulation study for bias correction on degree distribution
* Application to Mochudi data
* Notes:
  + Uses SIIR epidemic model

## Paper 3: Contribution of data sources

* While integrating multiple data sources can increase the efficiency and precision of estimation of network structure, when planning a study it is useful to know which sources are most informative.
* Gates Foundation gave supplemental funding to add viral genetics to BCPP, which raises the question of how much that improved our ability to estimate contact network structure
* Simulation study: mapping how bias and precision vary with inclusion/exclusion and strength of information from each data source.
* Application: compare BCPP results with and without viral genetics? Or vary weights placed on inputs (strengths of priors) and map out how estimates of network features (and the associated variability) change?
* Notes:
  + Uses SIIR epidemic model
  + Is this where we add non-egocentrically estimated terms?

## Paper 4: Data integration for HIV contact network estimation: San Diego

* Existing methods use general models for disease transmission over the network. We hypothesize that we can gain efficiency by customizing the epidemic model to the disease of interest.
* Choice of epidemic model (4 stages of progression, on/off treatment) – justification and specification.
* Likelihood and MCMC to implement new epidemic model.
* Apply to San Diego data
* Notes:
  + Include dyadic dependent terms: assortativity, clustering

## Paper 5: Validation of HIV model

* Apply methods of paper 4 to Atlanta
* Summarize structures of HIV Network Metastudy networks
* Compare San Diego and Atlanta estimates to HNM – are they at least plausible?
* Notes:
  + Atrium Health might be interested in related questions and developing clinical trial

## Paper 6: Data integration for estimation of features of dynamic contact networks

* To date, all models have assumed a static underlying contact network. In many situations this is not a reasonable assumption.
* Likelihood and MCMC for simple epidemic model with dynamic network
* Application – COVID data?

## Potential Paper 7: Dynamic networks and HIV

* If Paper 5 suggests that we are not getting accurate and/or plausible estimates of dyadic-dependent network properties from HIV model, try extending it to use dynamic contact networks.

Ideas:

* Race/ethnicity, SEP models for SDoH type research
* PCORI, Booz-Allen might be interested in funding related work