Emily's Thesis Title

Emily Pollock

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021?

Reading Committee:

Steven M. Goodreau, Chair

person 1

person 2

person 3

Program Authorized to Offer Degree:
Biological Anthropology

University of Washington

Abstract

Emily's Thesis Title

Emily Pollock

Chair of the Supervisory Committee:
Title of my chair Steven M. Goodreau
Biological Anthropology

"Here is my abstract"

Table of Contents

List of Figures

List of Tables

Acknowledgments

"My acknowledgments"

Dedication

"My dedication"

Introduction

Anthropologists have long recognized the importance of social connections and behavioral variation among humans and our nonhuman primate relatives. Indeed, the ability for us to participate in distinct but potentially interlocking complex social networks has fueled our evolution as a species and made our uniquely elaborate life possible. Network analysis has often been utilized as a way to visually and quantitatively represent these ties in order to understand their effects on those connected to each other, from kinship, social support and social capital, to the diffusion of information and transmission of disease. These latter networks are crucially important to our understanding of how human biosocial variation influences our health, where the oft-beneficial complex social networks we maintain and navigate every day can also put us at risk of exposure to infection.

transition to STIs

In order to understand the complex patterns by which sexually transmitted infections (STIs) are transmitted throughout populations, we first need to understand the behavior of human relationships and how these behaviors generate the dynamic sexual network across which these types of infections can spread.

This work is guided by the theoretical framework of the human ecology of infectious disease, the investigation of how human behavior, social patterns, and built environments interact with the broader pathogen environment to influence our health. Of particular interest is not just aggregate behavior, but also how variation in individual behavior influences social patterns and alters the landscape through which diseases spread, particularly as this variation relates to biological age. Syndemic theory will also be

used as a guide to understand how variation in behaviors and patterns act synergistically to increase vulnerability and exacerbate existing health disparities of certain population subgroups (Singer et al., 2006).

- Whole other review on chlamydia goes here? Or in Chapter 3? I think some here may make sense to really motivate things. As you see, I've mentioned chlamydia enough times already that knowing the basic epi would be useful. It also really drives home the "biological" aspect of biological anthropology early on.

can I pull some stuff from my PAA abstract about age?

- transition to a history of the evolution of epidemic models (ie. from compartmental where everything is basically independent and exponential through to ERGMs where formation can be quite elaborate but we've never spent much time thinking about dissolution)

Mathematical models are quantitative representations of real-life systems and the processes within these systems important to the outcome of interest. This form of inquiry is particularly useful when classic scientific experiments to understand disease spread or intervention efficacy cannot be conducted for either practical or ethical reasons, or when specific processes or parameter values in a system are unknown. In these situations, we use mathematical modeling as an in-silico laboratory to explore ideas and test hypotheses. Of course, the form and complexity of these models are determined by a variety of factors including the type of question that needs answering and the natural history of the infection of interest, but many types of mathematical models rely on similar underlying assumptions. Without diving too deep into the history of epidemic modeling, here I give a brief overview of the various model forms to highlight some key similarities and differences.

Initial mathematical models for epidemics were deterministic and compartmental in nature. They did not represent people individually, rather they group them into homogenous compartments representing specific states of interest, a portion of which transitioned between compartments at each time step based on a rate. In the most basic models, the compartments are usually "susceptible" and "infected" and the rate of transition from susceptible to infected depends on the rate of contact between the groups and

the size of the infected group relative to the whole population. Additional complexity can be added by adding more compartments or states, like breaking down the state of susceptible and infected into demographic states like race or age groups, adding compartments for vector populations like mosquitoes, or by representing a more complex natural history of the pathogen by including states for groups such as "exposed but not infectious", "recovered", "infected and symptomatic", or "infected and asymptomatic" to name a few. These models were deterministic in nature because the transitions between compartments rely on unchanging rates: the same proportion of each component transitions at each time point and if you run a deterministic compartmental model (DCM) multiple times you will alway have the same result.

Stochastic models grew out of this original framework as a way to capture variability and uncertainty in the systems we wish to study. In this scenario, some or all transitions between states were based on a *probability* of transitioning rather than a set rate, meaning that not the same proportion of a state transitioned at every time step, but on *average*

Notice the assumptions implicit in the way transitions occur in these models - it is memoryless, generating an exponential distribution (or geometric if using discrete time).

- Explain what ERGMs/STERGMs are and why they arose and what are all their various strengths, and then discuss how they have been widely used in epidemic modeling, for HIV/STIs, but even beyond (Sam has a good running list of other applications of EpiModel on the EpiModel page).
- Explain how current approaches to STERGMs have required us to make simplifying assumptions regarding dissolution, even as they've allowed us to do all kinds of awesome things in regards to formation/cross-sectional network structure. And the time has come to try to improve upon these methods within the STREGM framework. Give examples of the various questions that would be improved by doing so (e.g. impacts of EPT in chlamydia)
- And explain that (and perhaps why) it is (currently?) intractable to add in general non-memoryless forms of dependence into STERGMs, and to estimate these from data, so that it is useful to see how far one can get while adding in heterogeneity and complexity but still retaining some form of memorylessness.

Relationship Duration

- Make sure you dive a bit into some of the ways that people have tried to include more complex relationship durations in previous models before. In DCMs, this can include having multiple compartments represent a single state, because the sum of multiple exponentials is not exponential. I can't think of any specific examples where folks did this for relationship length (maybe there are none), but certainly for other types of transition probabilities.

- And then I'm sure there are papers that have included age-specific relational dissolution probabilities in agent-based epidemic models, using Martina's definitions of the terms (that is, they model relationships, but not with a formal statistical model like ERGMs). Try to find a few cases of this to show that it exists. Talk about how where their dissolution probabilities come from – survival analysis.

- Discuss how survival analysis is the obvious way to consider relational dissolution probabilities over time/age, and how a specific duration distribution implies a specific survival curve, and vice versa.

- dynamic networks require information about relationship duration

- why doing a bit better on dissolution/duration, especially by age, will be extra important for thinking about certain interventions for certain relatively short-lived infections (e.g. partner services in chlamydia).

exponential – memoryless survival function, exchangeability

Age-Related Processes * additionally, while births and deaths have been a part of models, only recently are we adding explicit age-dependent formation terms – and age changes over the simulation – what is this effect?

* including age in dynamic models may sound straightforward but as we're going to see adds a surprising

amount of complexity * Talk about why including age-specific rates for anything matters (seems obvious, but still say it. For example, talk a little about what we know about age-specific rates of chlamydia or other STIs; or even of infectious diseases with other transmission types for different reasons related to contacts. And/or talk about obvious fact that relationship patterns change in many ways across the life-course. Perhaps summarize a few modeling papers that bring in age-specific contact patterns in different ways just to make the general point.

Primary Data Source

The empirical behavioral data used in this dissertation are drawn from the 2006-2010 and 2011-2015 waves of the National Survey of Family Growth (NSFG). The NSFG surveys men and women aged 15-44 on many aspects of family life, including but not limited to marriage and divorce, pregnancy, contraception use, infertility, and other aspects of sexual and reproductive behavior. In addition to the demographic information recorded for each respondent and their sampling weights, in this study we use the data collected in section C of the public use files on each respondent's recent sexual partnerships with opposite-sex partners in the last year, with a maximum of three partnerships reported. These data include the century-month of first sexual contact, the century-month of last sexual contact, whether the respondent considers this sexual partnership to be ongoing, and the partnership status (marriage, cohabitation, or other). We limit the combined data set to those respondents who report at least one partnership in the last year. Out of the original 43,303 respondents, our subset contains 32,516 respondents who report on 40,443 sexual partnerships. Due to the study design, all relationships that respondents report as ongoing on the day of interview have right-censored relationship lengths, and there is left-truncation present due to the large number or relationships that started prior to the observation window but continued into it.

This dissertation will….First I will use tools from survival analysis to explore the limitations of the exponential assumption with regard to the distribution of relationship duration across the lifecourse. Second, I will document the ways in which age-related ERGM terms can behave in unexpected ways during certain conditions in simulations with vital dynamics and other important demographic processes and explore a variety of potential adjustments. Lastly….

1

Demography and Dynamic Network Simulations

In mathematical models the choice of model terms depends on the question of interest and the underlying patterns in the data and this is no less true for network models of sexual partnerships developed to understand disease transmission. Several previously published models using ERGMs and EpiModel to simulate epidemics focused on men who have sex with men (MSM) populations in a narrow age range, 18-35 (*cite papers and also double check that this is true*). These models focused on terms related to mixing patterns between races, the propensity to form relationships with individuals relatively close in age, and the likelihood of concurrent partnerships. Because prevalence of both main and casual relationships remained relatively stable over the small age range, the models did not include terms that used age as a predictor of relationship formation. However, in this project, we focus on heterosexual relationships over a larger

age range (15-45). Unlike the narrower MSM models, there are large clear differences in the prevalence of main and casual partnerships over this age range, so we will need to include terms that involve age in our model (see appendix for breakdown of empirical data relevant to the model terms). In addition to influencing the distribution of relationship duration, these differences are likely to be especially important if we want to use this type of model to understand the processes that generate the large observed differences in bacterial STI prevalence by age - one of the broader goals of this dissertation. However, while individual age is straightforward to represent in the model, it introduces several complicating factors for network models largely related to the boundaries imposed by age range and the aging process. This is not the first simulation to incorportae vital dynamics, and there are a few existing corrections relating to these processes already implemented

In this chapter, first I will show the relevant diagnostics demonstrating that the estimated model itself reproduces key network statistics in the absence of dynamic vital processes. Then I will show document how simulations that incorporate individual births, deaths, aging, and sexual debut lead the network to deviate from these key statistics. Third, I will explore the possible underlying causes for these deviations, implement a variety of corrections (some new, some extensions of previously implmented corrections), evaluate the efficacy of each, and outline some possible future directions for improved simulations.

1.1   Some Definitions

Static Diagnostic: Simulates networks based on the STERGM model fit and reports summary statistics for formation model terms. These networks are not dynamic, thus no dissolution or durational statistics are reported.

Dynamic Diagnostic: The dynamic diagnostic simulates the networks over time and tracks cross-sectional summary statistics for the formation model terms as well as dissolution model and relational duration statistics. This is a closed system: nodes do not depart, new nodes do not arrive, and no nodal attributes change across the simulation period (i.e. if you are age 15 when then the diagnostic begins, you remain 15 for the entire run).

Simulation: While the static and dynamic diagnostics are run using simulation methods, when I refer to

the "simulation" I am referring to the simulation using the EpiModel API that controls vital dynamics and other processes like sexual debut and eventually, disease transmission.

Node / Ego: An individual in the network.

Nodal Attribute: A trait of an individual (sex, age, sexual debut status, etc)

1.2 Base Model Overview

The base network models we use for this chapter focus largely on the age effects of relationship formation, with additional terms and structural offsets. For epidemic modeling purposes it will be important to also model race/ethnicity due to the existence of large racial dispartiies in prevalence of several sexually transmitted diseases, but that additional complexity is not the focus here.

[network terms]

1.3 Closed-System Results

One of the several steps to check model fit and convergence is a dynamic diagnostic. In this diagnostic, we simulate the STERGM for X repetitions of Y time steps and evaluate the cross-sectional network statistics over time. At each time step, ties can form and ties can dissolve based on the model coefficients. If the model is estimated properly and sufficient MCMC intervals are used, the network formation statistics should hover around their estimated targets. In this diagnostic we also evaluate the duration of ties and the rate of tie dissolution to ensure the dissolution targets are met. It is important to note that this diagnostic is an indicator of model performance in a closed system: all nodal attributes are fixed, no nodes exit, and no new nodes enter the population. In this context the models perform exceptionally well. Not only is the overall mean degree of each network met, but mean degree by age in both networks is reproduced well. Additionally, both models reach the target mean cross-sectional relationship duration after sufficient time. Any deviation from these targets as we move to the simulation then, will be related to the introduction of vital dynamics and other processes like sexual debut that alter nodal attributes that model terms.
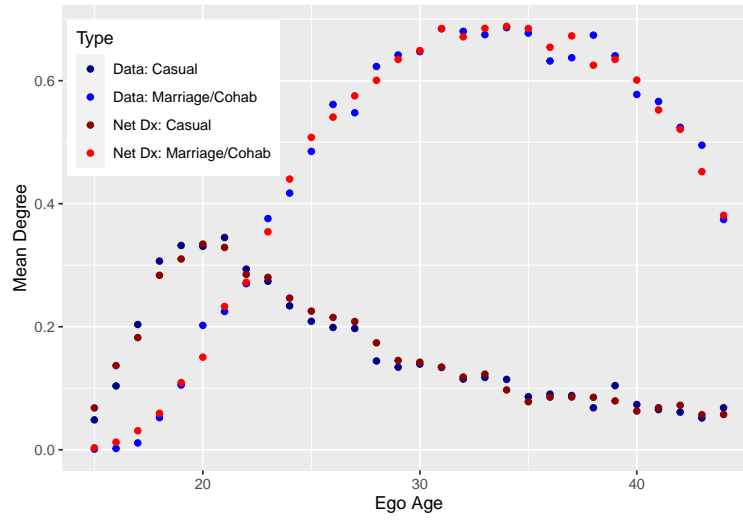
Figure 1.1: Comparison: Egodata vs Diagnostic Mean Degree.
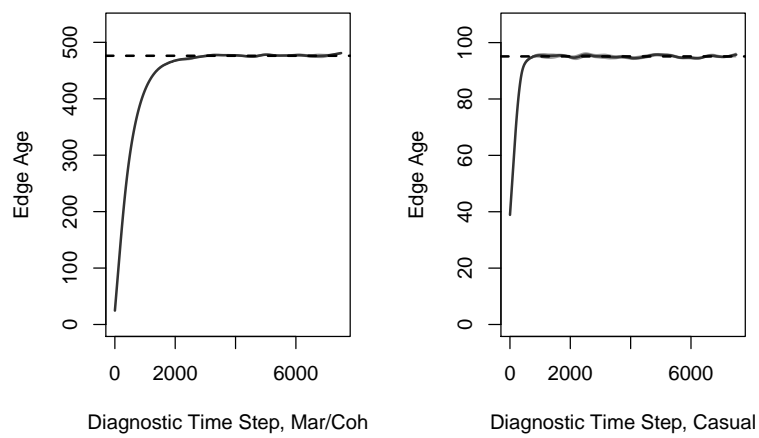
Figure 1.2: Mean Relationship Lengths in Diagnostic
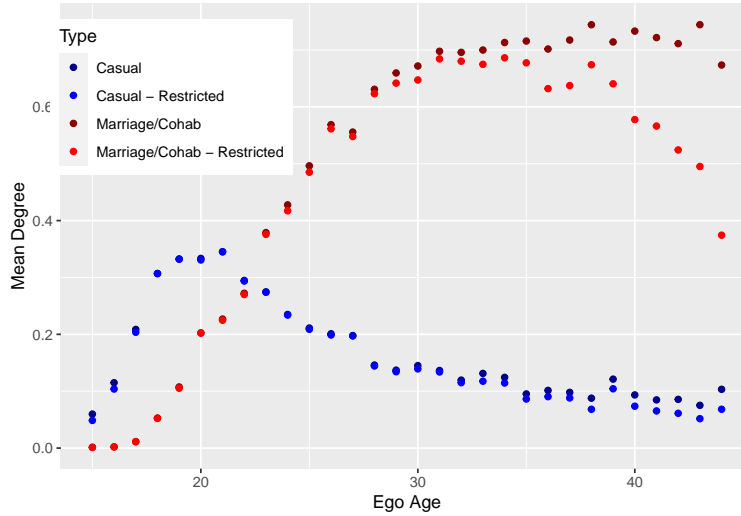
Figure 1.3: Mean Degree by Ego Age and Relationship Type, Restricted and Unrestricted Alters

1.4 Overview of Demographic Processes of Interest

The simulations run using the EpiModel API are distinct from the above closed-system dynamic diagnostic in that in addition to tie formation and dissolution at every time step, a series of modules is run that govern important demographic processes: node departure, node entry, aging, and sexual debut. Nodes automatically depart the model at age 45. This boundary was selected for two main reasons: 1) According to the CDC in their 2018 surveillance report, 97.4% of all chlamydia infections were diagnoses in the 15-44 age range (cite surveillance report) and 2) the National Survey of Family Growth, the empirical data from which we estimate our model, only surveys adults aged 15-44. There are likely other sources of information that we could use to increase the age range, but it did not seem necessary to our questions of interest. Note that implicit in this decision is the elimination of all reported relationships among egos aged 15-45 whose *partners* are outside of this age range. The degree distribution that we actually use to estimate the model (and are trying to maintain during simulation) looks rather different than the original distribution shown above, particularly in the marriage/cohabitation network (see 1.3. We will consider the consequences of this in a later section. In addition to the age boundary at 45, all individuals experience the possibility of dying at each time step. Each node belongs to a class based on their 5-year-age-category

Table 1.1: Mean Degree and Duration Comparison, Targets and Base Simulation

	Mean Degree Target	Base	Pct Off	Mean Duration Target	Base	Pct Off
Marriage/Cohab	0.455	0.431	-5.27	476	365	-23.32
Casual	0.159	0.152	-4.40	95	103	8.42

and their sex, and is evaluated for death at every time step with the probability determined by data from published in U.S. Vital Statistics documents (cite). Given that our age range is relatively young, departures due to background mortality are uncommon relative to the effect of the age boundary on which nodes depart the model. Nodes enter at age 15 at a rate based on the expected number of departures per time step in order to keep the population size relatively stable. Like ASMR, the actual number of entires per time step is stochastic but maintains a population size within 1-2% of the starting size of 50,000 nodes. Each time step in the simulation represents one week, so nodes age by 1/52 per time step. Nodes enter the population at age 15 and are evaluated for sexual debut at each time step, with probability that increases until age 29 to match the age-at-debut distribution as reported in the NSFG. In accordance with the data, some individuals will never debut into the heterosexual population and will therefore never form a tie in these networks.

1.5 Initial Simulation Results

Unlike the dynamic diagnostics, when we run these simulations with demographic processes, several metrics stray from their target values. First, the overall network mean degree, or average number of relationships per person, is too low in both the marriage/cohabitaiton network and in the casual network (by roughly five and three percent respectively). These deviations are not incredibly large, but they are somewhat concerning esepcially becuase Second, the mean degree by age is severally underpresented in both networks for the youngest ages but overrepreseted in the mid-30s. Finally, the mean relationship length is 24% too short in the marriage network but 7% too long in the casual network.
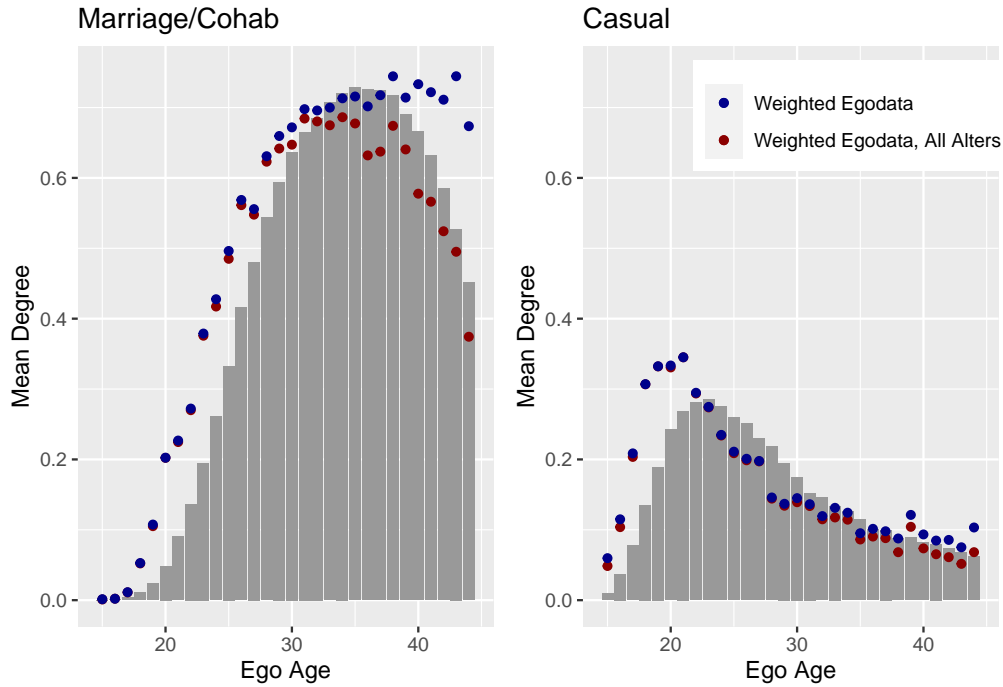
Figure 1.4: Base Simulation: Mean Degree by Age.

1.6   BOUNDARY EFFECTS & EXISTING CORRECTIONS

This section provides an overview of existing corrections for dynamic networks and dynamic demography and outlines the boundary effect issues that we will explore in more detail below.

1. Formation Approximation In many cases, a full STERGM cannot directly be estimated directly due to the computational burden when networks are large, sparse, and have relatively long relational durations (which describes many sexual networks). Instead Carnegie, Krivitsky, Hunter, & Goodreau (2015) introduce an approximation to full STERGM estimation that uses the same data: cross-sectional egocentric network data and information on tie duration. In this approximation, the static ERGM is estimated using standard techniques. Then the edges formation term (which represents the base propensity for ties to form between any two individuals in the network) is decreased by the log odds of the probability of edge persistence, in effect transforming the formation term from *prevalence* of ties in the network to the *incidence* of ties. The explorations below do not attempt to modify this approach, but instead explore

the relationship between the adjustment of the formation coefficient, the probabilty of tie persistence as estimated from the long relationship duration expected in the marriage network relative to the limited observation window per individual as they arrive and eventually age out.

2. Right Boundary Existing Correction: Departure The node departure correction used in the model estimation-to-simulation workflow was developed after the observation that when nodes were removed from the simulation to mimic, for example, background age-specific mortality, the mean degree of the network became lower than expected, as does the mean duration of relationships. The logic is relatively straightforward: the statistical model underlying these network simulations balances the probability of tie formation with the probability of tie dissolution in order to maintain a target number of ties in the network. However, when nodes depart, some additional ties will break due to this process, lowering mean degree and the mean duration of ties. This node death is exogenous to the originally estimated statistical model, and therefore "unexpected". To counter the lowering of relationship duration (and subsequently mean degree) related this excess node death, the expected (endogenous) duration of ties is increased such that the *average* duration is maintained.

The departure correction implemented in previous models has two components: 1) the mortality rate per time step averaged across the entire population, and 2) the rate at which individuals depart the simulation due to the age boundary, calculated as 1/(time steps each node is expected to be observed in the simluation). In the past this approach has successfully....

However, due to the large difference in average probability of breaking a tie upon departer between the marriage/cohab network and the casual network (due to the distribution of mean degree by age), it seems unlikely that the departure correction should be the same for both networks. 3. Left Boundary Existing Correction: Arrival Kritvitsky, Handcock and Morris (2011) - this correction is designed to maintain the target mean degree of the network in the presence of changes in the size of the population. In these simulations, while the exact number of entries at every time step is stochastic, the rate is designed to keep the population size relatively stable so while we do use this correction, we will not futher modify it. However, another arrival-related problem is the issue is that fact that nodes do not enter the population *with* ties to other nodes. While the model coefficients suggest a certain average number of relationships among

these young adults, because the nodes age each week, there is a very limited time frame in which to form relationships and hit these targets.

1.7 Considering the effect of older partners

Because we observed that in the marriage/cohabitation network there is an overrepresentation of relationships among the older egos, and the model coefficients suggests that older nodes are in general more likely to form relationships than younger nodes (with some tapering as age increases), we theorized that when a node aged 45 aged out and broke the tie with their partner (who is likely to be somewhat close in age), that the partner remaining in the simulation very quickly forms a new relationship. However, the tie that dissolved as a result of one partner leaving the simulation due to this age boundary is not a true dissolution, and these newly formed relationships should not actually exist because the remaining partner should not actually be eligible to form a new relationship in the network yet. Perhaps then, these new, short relationships in older ages contribute both to the lower than expected mean relationship duration in the marriage network and the lower than expected mean degree at younger ages.

We consider two ways to address the effect of partners outside the age boundary. First, we prevent egos whose partners have aged out from immediately forming new relationships by adding an offset terms for egos who meet this condition. In this scenario we hope that by preventing new relationships from forming among egos whose previous relationships were terminated artificially by the age boundary, the simulation will better match the data with the restricted alter set and increase the mean relationship length by generating new relationships at earlier ages. In the second scenario, we increase the age at which egos depart the simulation to age 65. While we may not be interested in modeling individuals older than 45 for epidemiological reasons, it may be worthwhile to keep them in the simulation over a longer period of time to avoid the artificial ending of relationships. In this case we hope to match the empirical mean degree distribution among egos with the age-unrestricted alter set. However, because we would be simulating individuals outside the age range in the data we used for estimation, we may run into additional issues.

1.7.1 Offset for Partner Age-Out

This scenario adds an offset term to the formation model ("olderpartner") for egos whose alters are outside of the 15-45 age range modeled in the simulation. During estimation there are already some egos whose partners are outside the age range so they appear to have degree 0 and do not contributed to the expected edge count but are flagged as "olderpartner=1". During the simulation, if an individual ages out while they are in a relationship, the remaining partner gets flagged by the "olderpartner" attribute and are prohibited from forming a new relationship. The probability of becoming available for a relationship on any future time step is equal to 1/expected duration of the relationship type, although in the case of the marriage/cohab network relationships last so long that it's unlikely that a node become available for the rest of their simulation lifecourse (unless the age difference between partners was exceptionally large, which is not impossible).

I don't necessarily expect this to solve the issue of the overall mean degree, but if we prevent relationships that only exist due to the age boundary, perhaps these relationships will be distributed among the younger issues. In turn these relationships would begin earlier, and possibly increase the average mean duration.

Results

The first thing we note is that this offset did not largely influence the overall mean degree in either network, nor did it increase the mean relationship duration in the marriage/cohabitation network (mean relationship length was also unchanged in the casual network, but we did not necessarily expect it to). However, when comparing mean degree by age between scenarios, the offset did correct much of the overrepresentation of relationships at the older ages, and also slightly increased the mean degree in the youngers (these changes are subtle but present). The casual network was largely uninfluenced by this offset, but that would be expected given that older ages are actually less likely to form casual partnerships than younger ages.

Table 1.2: Mean Degree and Duration Comparison, Targets vs Older Partner Offset

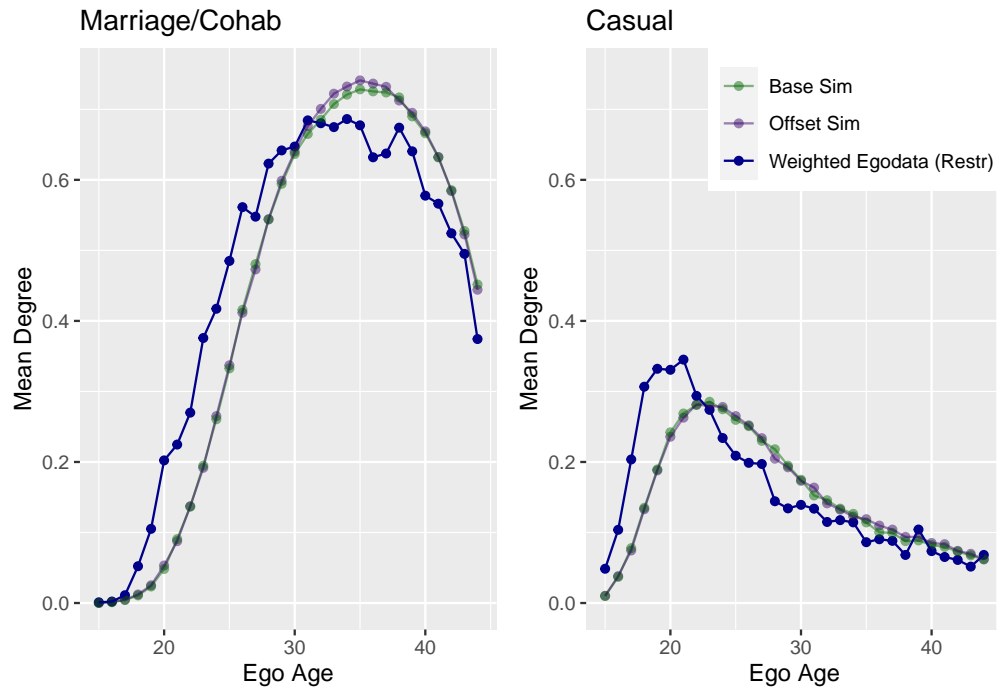|	Mean Degree Target	Base	Pct Off	Mean Duration Target	Base	Pct Off
Marriage/Cohab	0.455	0.434	-4.62	476	364	-23.53
Casual	0.159	0.152	-4.40	95	102	7.37

Figure 1.5: Mean Degree Comparison: Base vs Offset.

Table 1.3: Mean Degree and Duration Comparison, Targets vs Increased Age Boundary

	Mean Degree Target	Sim	Pct Off	Mean Duration Target	Sim	Pct Off
Marriage/Cohab	0.455	0.472	3.74	476	414	-13.03
Casual	0.159	0.141	-11.32	95	104	9.47

Increase Age Boundary

In this scenario, we hope to move the degree distribution closer to the egodata distribution with the age-unrestricted alters (blue dots) – the distribution that better represents reality. This scenario includes the offset for "older partners" but employs it in a slightly different fashion. In the previous scenario, edges dissolved artificially when one of the partners left the model at age 45. We now allow those relationships to continue as they would normally by increasing the age of departure in the model to age 65. However, we use the offset to prevent any nodes older than 45 but not in a relationship from forming new relationships. Only relationships that began prior both partners turning 45 exist.

Results

First off, it is clear that we can easily recover the marriage and cohabitations lost to the age boundary among egos in the 35-45 age range simply by keeping their older partners in the model, even if the data used to estimate the model did not include these partners. However, this approach has consequences. Because the model is targeting a mean degree based on the restricted partner data, the maintenence of relationships among 35-45 year olds increased the overall mean degree beyond the target and also comes at the expense of relationships among the younger ages, the section of the distribution that we already fail to match well. The mean age of relationships has increased, but this is clearly a result of the relationship at older ages, thus only a partial success. The casual network also displays some undesireable qualities similar to the cohab network. The mean degree is too low at the expense of the younger age group and the mean relationship length is unchanged. *Discussion*

conclusion: we keep offset in all future scenarios but not older partners, older partners may be a good
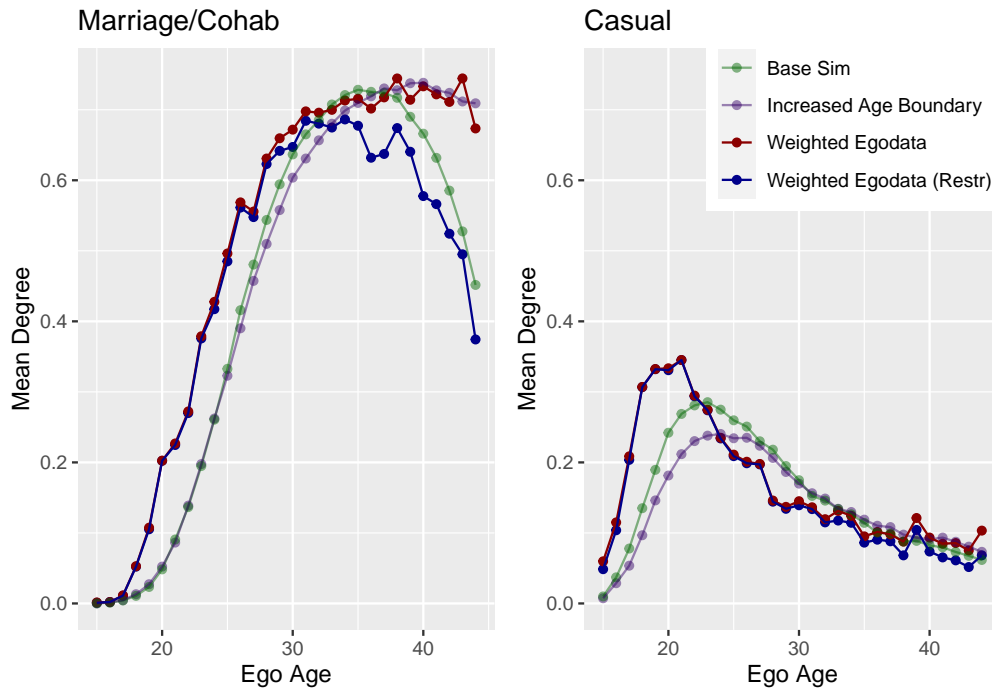
Figure 1.6: Mean Degree Comparison: Increased Age Boundary.

idea in some case but we have to rethink some things especially if the younger ages are going to perform worse

1.8 Relationship Length & The Simulation Window

So far, nothing we have done has largely influenced the issues with relationship duration in these networks. In the marriage/cohab network, the mean duration falls nearly 2 years short of the expected length and in the casual network the mean duration is roughly 3 months too long. There are a few possible reasons that there may be a mismatch between the formation and dissolution coefficients in-simulation that may contribute to these outcomes. Here we explore possible issues related to the window of observation for each node in the simulation and how that inluences the observable mean relationship duration. [will need to re-frame this section if the survival analysis doesn't happen until chapter 2]

From our survival analysis it is clear that the exponential distribution, even when separated into separate
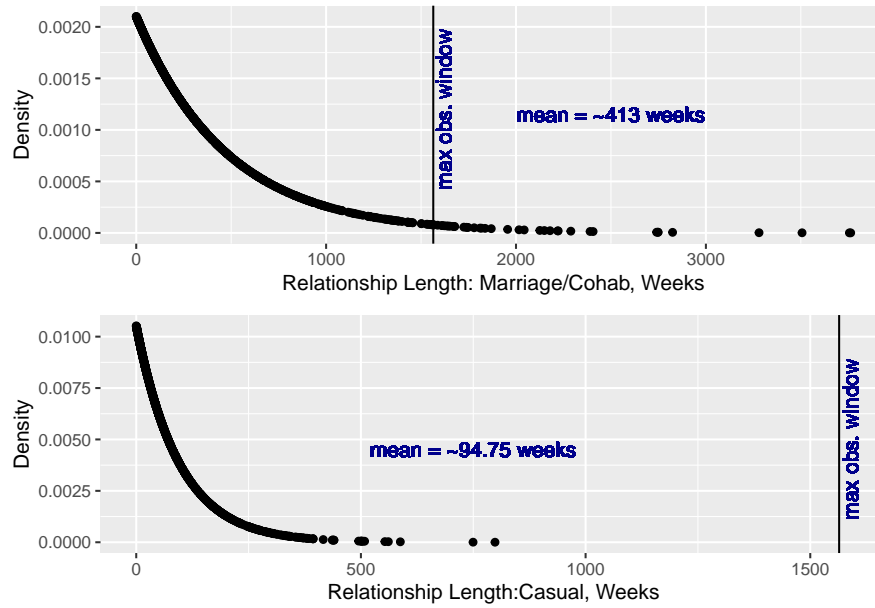
Figure 1.7: Predicted Distribution of Relationship Lengths and Simulation Window

networks by relationship type, had some serious limitations in its ability to represent the full distribution of relationship lengths - both by age and across the whole population. One of the limitations that may pertain to the right tail of the distribution. An exponential distribution with a mean of roughly 95 weeks (the mean cross-sectional length in the empirical data) has a very long right tail extending beyond a normal human lifespan, 15 years. Clearly this tail isn't possible to observe, even less so when you consider that the window of observation in the simulation is equal to the age range of the population, 15-45 (30 years). 1.7 shows the density plot of relationships lengths that are exponentially distributed with a mean of 95 weeks. While 100% of observsations lay within the simulation window, the removal of the tail lowers the mean observerable relationship length based on this distribution (the mean of relationship lengths if you remove the observations that are impossible to occur in the simulation) from 95 weeks to 96 weeks. The mean relationship duration in the casual network is also shown to demonstrate that the simulation window of each node is unlikely to contribute to the variation we see in the mean simulated relationship duration in the same way and as such we will only implment a correction for the marriage network.

In this scenario, we increase the edges formation coefficient by the difference between the log odds of the

Table 1.4: Mean Degree and Duration Comparison, Targets vs Edapprox Correction

	Mean Degree Target	Sim	Pct Off	Mean Duration Target	Sim	Pct Off
Marriage/Cohab	0.455	0.447	-1.76	476	369	-22.48
Casual	0.159	0.152	-4.40	95	103	8.42

target mean duration and the log odds of the observable mean duration in the marriage network. If you recall, this in effect slightly alters the "edapprox" approximation method described above from adjusting the static edges coefficient by the maximum observable mean duration rather than the rather than the target mean. We hope that this wil help edges to form on average earlier in the lifecourse and help us recover missing edges.

While not fully considered here, this issue could also be exacerbated by the fact that the mean age of marriage/cohab formation is so much later than the mean age in the casual network. When relationships form earlier on in the life cycle, the average duration is less curtailed by the age boundary of the simulation. So in addition to being much shorter than marriages and cohabitations, and the obverall expected distribution is less influenced by the obvservation window, the representation of casual relationships also benefit from their unique distribution. Results

The boost in the edges coefficient sucessfully increased the overall mean degree of the network to within 2% of the target mean degree. However, very little of the increase in mean degree came from the an increase in the younger ages. Instead, the boost largely only increase the degree at the peak, which was already over-representing relationships in the mid-to-late 30s. Additionally, we only gained about two months in mean relationship duration. This is again likely due to the increase in mean degree in individuals in their 30s rather than across the network more evenly.

1.9 Departure

In this scenario, I re-consider the standard implemented departure correction. As stated earlier, this correction acts to balance out the "unexpected" edge dissolutions due to aging out of the simulation or age-
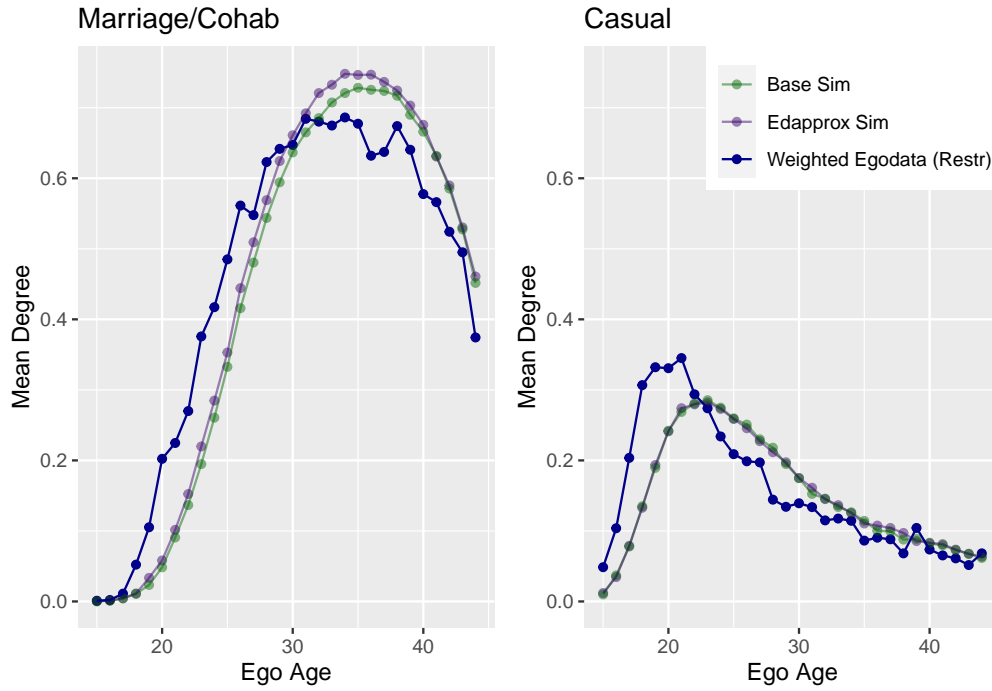
Figure 1.8: Mean Degree Comparison: Edapprox Correction

specific mortality by lowering the probability of "expected" edge dissolution and thereby increasing the "expected" duration of relationships. In the past this correction has been the same for both main (marriage/cohab) and casual partnerships. However, due to the large difference in the likelihood that a nodal departure causes an edge to dissolve between the networks, it is not clear that this correction *should* be same.

The current correction is calculated by adding the likelihood that any one node departs the network due to aging out multiplied by the mean weighted age category-specific mortality rate per time step:

$$drate = \frac{1}{simulation window} + ASMR_w eighted$$

The new correction represents the likelihood that if a node departs, it also dissolves a edge.

$$1 - sum(S * D * P)$$

Table 1.5: Original and Updated Mortality Rates

Original Mortality Rate	Updated Marriage Rate	Updated Casual Rate
0.0006645	0.000749	0.0002807

Table 1.6: Mean Degree and Duration Comparison, Targets vs Edapprox + Mortality Corrections

	Mean Degree Target	Sim	Pct Off	Mean Duration Target	Sim	Pct Off
Marriage/Cohab	0.455	0.455	0.00	476	387	-18.7
Casual	0.159	0.144	-9.43	95	95	0.0

where

S is the vector of survival by age category due to aging out

D is the vector of age category specific mortality, and

P is the vector containing the proportional mean degree in each network by age category

This results in an estimate for the marriahe/cohab network not too far from the original correction, but a much smaller correction for the casual network. This makes sense given the very low mean degree at the oldest ages. Results You should notice that the mean degree of the casual network was lowered by the implementation of this departure correction. However this doesn't mean that we're not on the right track. Instead, look to the mean relationship duration. Because we are no longer overcorrecting for the number of relationships lost to the age boundary, the endogenous expectation of relationship duration is not too long and the cross-sectional mean duration hits the target. The decrease in mean degree is likely due to these relationships that don't last artifically long. The marriage network has also experienced some good improvements. The overall mean degree is now on target, and although the mean relationship duration is still too low, it has increased by almost half a year. However, we can see by the distribution of mean degree by age in both networks that much of the missing relationships are still in the youngest ages. We will focus on this problem in the next section.
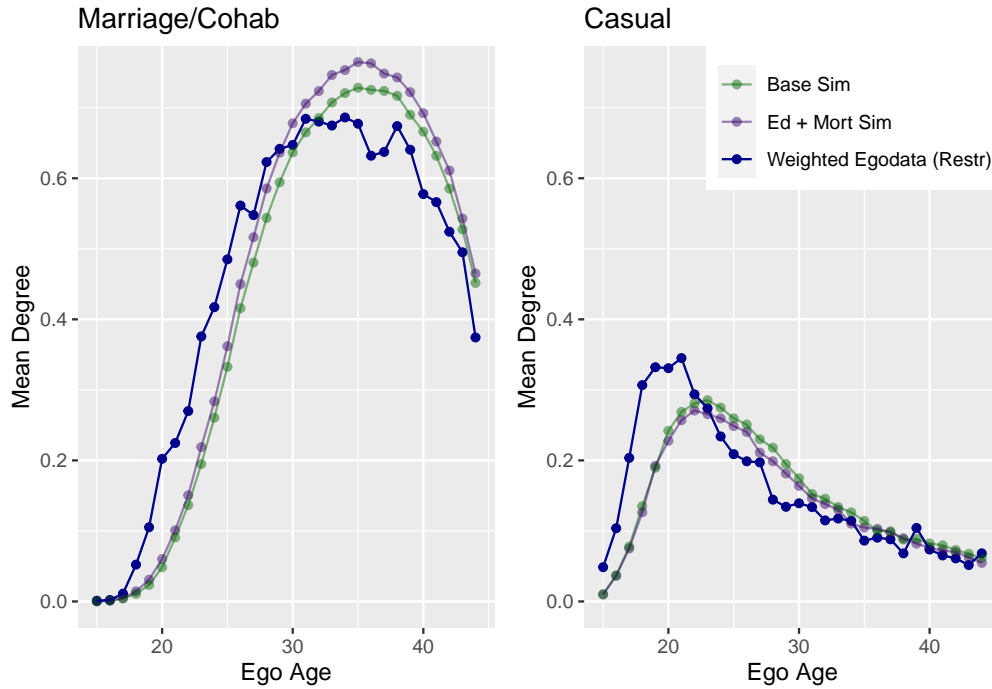
Figure 1.9: Mean Degree Comparison: Departure Corrections.

1.10 Arrival & Sexual Debut

The failure of these networks to adequately form relationships among the youngest ages is yet another form of a boundary problem. The big-picture problem is that 15-year-olds do not enter the population and bring in existing relationships. This creates a problem for the model because the coefficients by age represent the prevalence of relationships at each those ages, and based on these coefficients and the underlying network the algorithm makes and breaks a certain amount of relationships at each time step. It doesn't however, expect to have to form *all* of the relationships among 15-year olds almost *immediately* up on entry, which needs to happen 1) to meet the expected mean degree, and 2) in the presence of aging when there is a very limited window to hit that age target. This also makes large jumps in the expected mean degree challenging, like in the marriage network between age 18 and 25 or in the casual network between ages 15-20. In this section we test two possible approaches to this problem. The first involves manipulating the number of individuals eligible for relationships based on the sexual debut process, and

the second takes a more direct approach to manually calibrate the formation coefficients at certain ages to boost the rate of formation beyond what the ERGM initially estimated.

I am going to contextualize our scenarios and results in two ways. First, in the ability to represent the mean degree distribution from the data (the cross-sectional relationship landscape) and second, in the accuracy of the in-simulation debut rates (i.e. who actually forms a partner at any point in the simulation) - (the longitudinal landscape). Note that the STERGM methods were developed to maintain certain cross-sectional statistics so the longtitudnal correctness isn't the primary goal, but it is important to think about when considering the broader context.

1.10.1 Sexual Debut

[parts of this section too wordy / possibly not necessary]

Representing the sexual debut process is both complex and highly important if we wish to model sexually transmitted diseases in adolesents and young adults. In the U.S., more than 50% of all sexually transmitted bacterial infections such as chlamydia and gonorrhea diagnosed yearly occur among individuals aged 15-24, but not everyone in the age group are sexually active, which concentrates the transmissions into subset of the population and increases the probability of exposure to an STI for those sexually active moreso than at older ages. It is important then, to approximate this process in simulation as faithfully as possible. If too many individuals are able to form sexual partnerships in the model, we risk under-estimating the risk of exposure for those sexually active and conversely over-estimating the risk of exposure if too few are sexually active.

The process by which individuals are labled as "sexually debuted" and thus eligible to form relationships in the model is reasonably straightforward. Individuls enter the model with a 10.6% probability of debut, based on the proportion of 15 year olds in the NSFG who reported having sexual intercourse with a member for the opposite sex prior to age 15. For the rest of the age distribtion (15-44), we used the responses to "have you ever had sexual intercourse with a member of the opposite sex" to generate a cross-section distribution of sexual debut status. From this data we estimated the yearly probability of debut among those who had not, and converted that into a weekly probability. The empirical data and the in-simulation
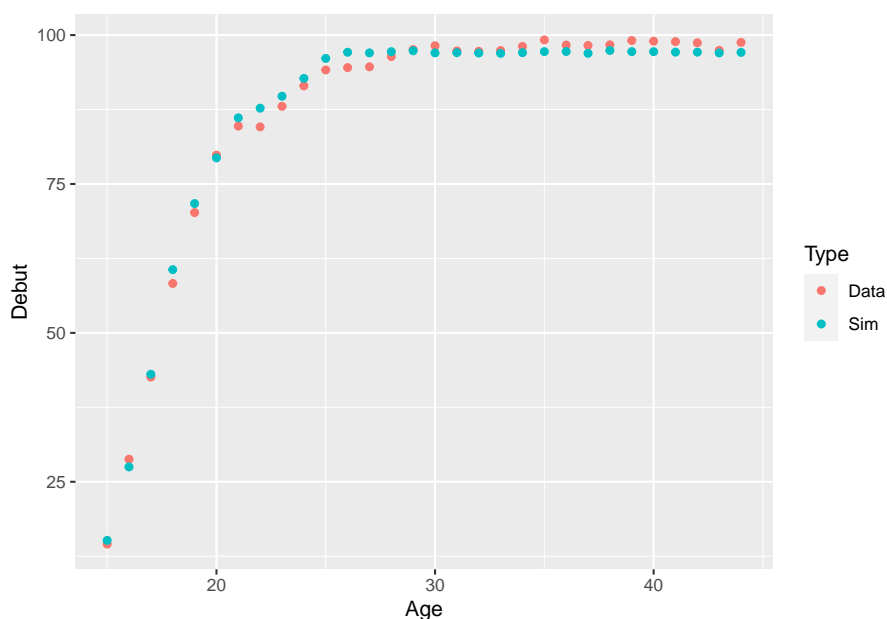
Figure 1.10: Sexual Debut Status: NSFG vs Simulation

distribution of sexual debut from the base model scenario are shown in 1.10. The major caveat to this approach is that it does not mechanistically represent the debut process as well as desired . In real life, a person "debuts" by entering into a sexual partnership. This puts the model in a bit of a catch-22 situation: an individual can't debut in real life until they form a sexual partnership, but in order to form a sexual partnership in the model, they must already be "debuted". We can match the distribution of sexually debuted individuals, but it is possible that the number of individiduals who have "debuted" does not equal the number of individuals who have actually formed a relationship at any point in the simulation.

What if instead we used the "debut" term to model the idea of "sexual eligibility" rather than explictly debut? The concept is straightforward: in most cases, an individual would decide that they are *ready* to begin having sex some period of time prior to actually forming a sexual partnership (or transitioning a relationship from a non-sexual partnership to a sexual one). It is perhaps this underlying trait that we should model instead, allowing us to model the sexual history of individuals in our model more completely. Unfortunately our survey data don't allow us to answer this question directly (i.e. at what age did you decide you were ready for sex vs at what age did you actually start having sex), and the literature

has laregly focused more on individual characteristics and within-partership dynamics that predict sexual debut rather than quantifying the time to readiness or the time from readiness to debut (cite that review paper, others).

In this scenario, we change only the probability of having sexually debuted at entry so that the *effective* debut in simulation matches the proporion of 15 year olds who report having had sexual intercourse. We call this the "eligbility" scenario. All other parameters (i.e. the rate at which debut occurs among the non-debuted) remain the same. As a reminder, in this model setup, the rate of sexual debut does not influence the birth/arrival rate in the model. Eligibilty status however does dictate whether an individual is allowed to form a relationship, and the number of un-debuted persons was jointly estimated in the model, so deviations from the original distribution of eligible individuals will influence the likelihood of tie formation.

Results

The switch to an "eligibility" metric had some dramatic effects on the casual network and moderate effects on the marriage/cohabitation network. In the marriage network, the overall mean degree has increased to about 7% greater than the target, and although we do see increases in the mean degree at younger ages that almost matches the targets, once again, the majority of the degree increase is seen between ages 30-40. The increase in the number of relationships that begin at earlier ages has increased the mean relationship length by roughly one year, but we still fall far short of the target. In the casual network, the increase in available egos for casual relationships has led to an incredible increase both the overall mean degree and in the mean degree in the under-30 population. The mean relationship duration in this network has stayed wihtin 1% of the target, which is good, but wildly over-represents the total number of relationship at most ages. This scenario will come the closest to reproducing the actual debut distbution of the data (1.13), but largely at the expense of the casual network's degree distribution.

Table 1.7: Mean Degree and Duration Comparison, Targets vs Expanded Eligibility

	Mean Degree Target	Sim	Pct Off	Mean Duration Target	Sim	Pct Off
Marriage/Cohab	0.455	0.488	7.25	476	401	-15.76
Casual	0.159	0.227	42.77	95	96	1.05
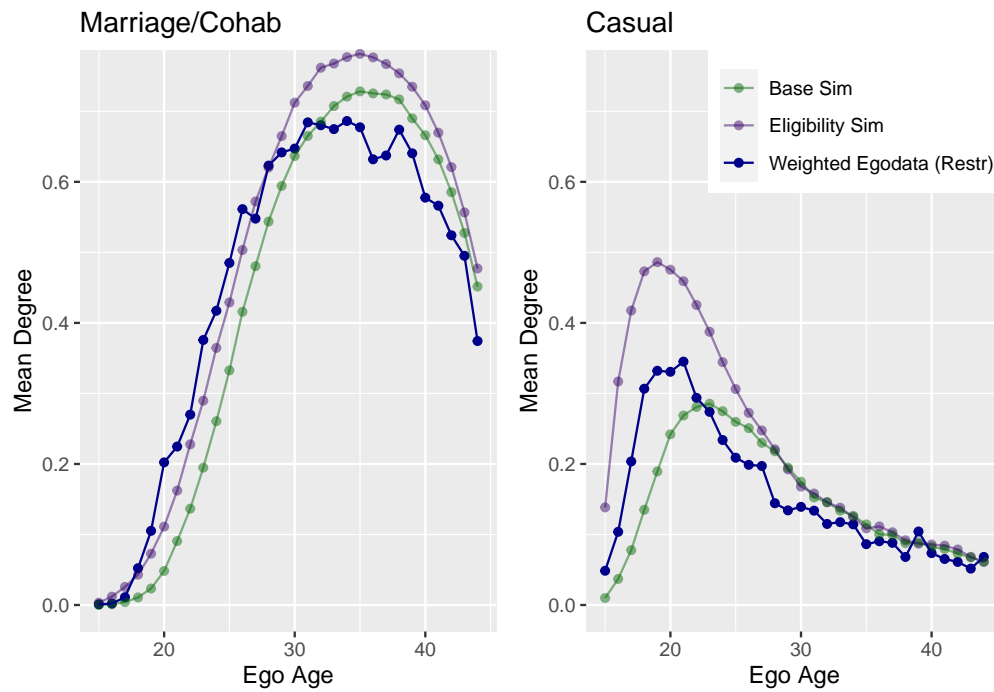
Figure 1.11: Mean Degree Comparison: Eligibility.

Table 1.8: Mean Degree and Duration Comparison, Targets vs Young Formation Boost

	Mean Degree Target	Sim	Pct Off	Mean Duration Target	Sim	Pct Off
Marriage/Cohab	0.455	0.497	9.23	476	413	-13.24
Casual	0.159	0.183	15.09	95	96	1.05

1.10.2 Young Age Formation Boost

In this scenario, we change none of the sexual debut parameters in leiu of adding a nodefactor term for "young persons" in each network and manually modifying 1) the degree of increase in formation, or the "boost", and 2) which ages the boost applies to in each network. This was manually calibrated by trial and error.

There's probably more I can say here. *Results*

- yay mean degree marches through age 25 in the marriage network and through age 20 in the casual network!

- still have problem with over-boosted middle ages and as a result have too high overall mean degree

- but 413 is the mean relationship duration we expect based on an exponential distribution w/ mean 474 with the right tail missing, so I actually think we're doing great here

- effective debut better than default run but not as good as "eligibility" aka trade off between cross-sectional correctness (mean degree) and longitudinal in some cases. Need to think through what this means for the concentration of disease at younger ages more.

- also effective debut not better than simulation with no debut term or process at all LOL
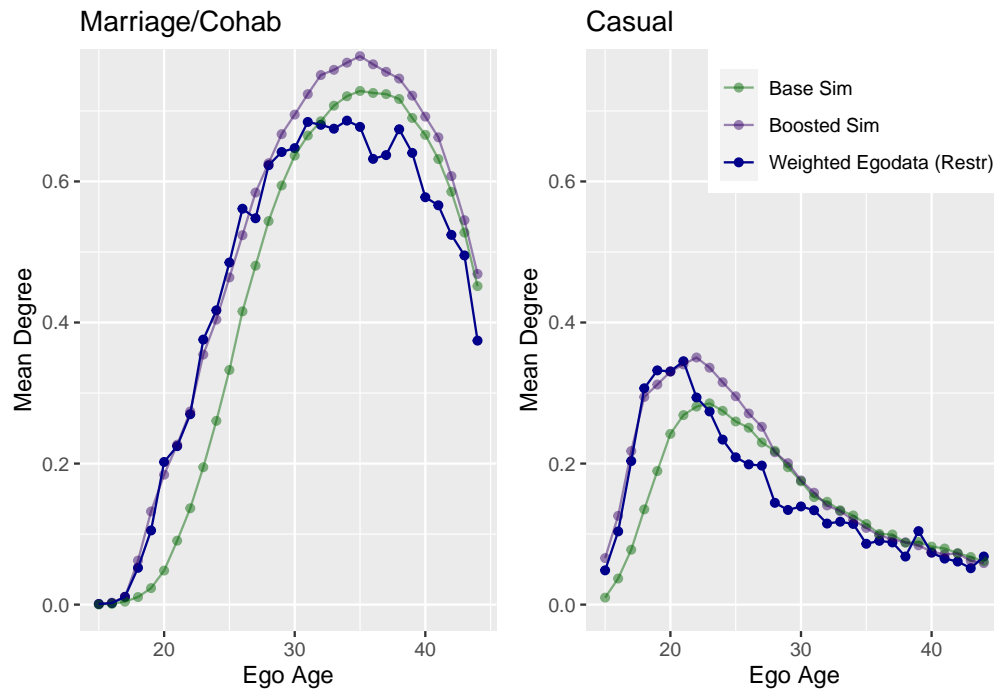
Figure 1.12: Mean Degree Comparison: Young Age Formation Boost.
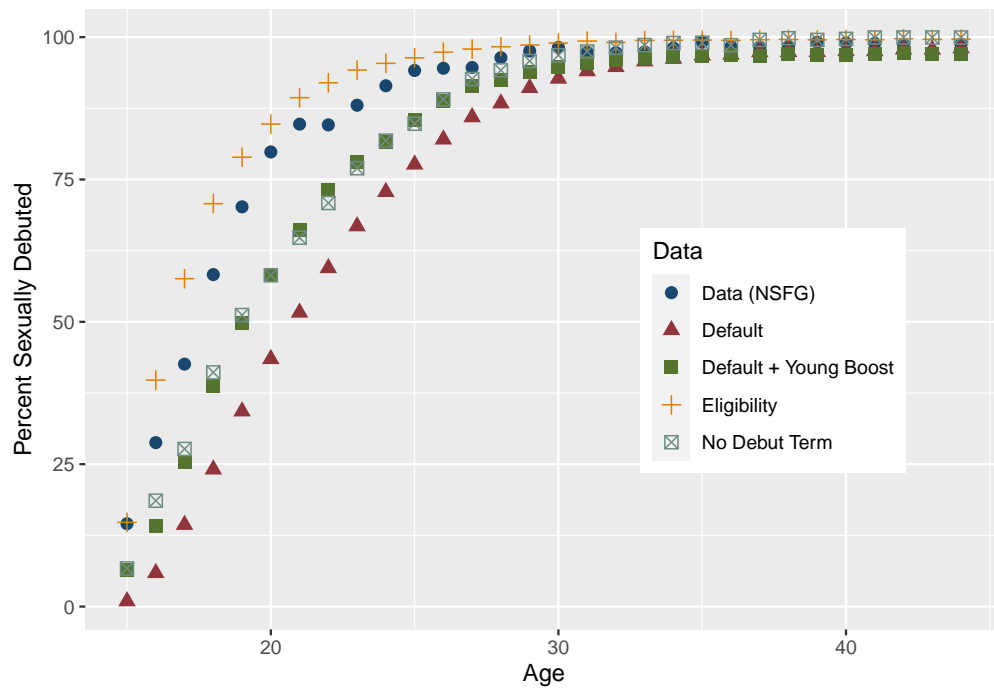
Figure 1.13: Percent Debuted In-Sim vs Data, Various Scenarios

Table 1.9: Mean Degree Comparison Summary Table

	Target	Base	Older Partner Offset	Increased Age Boundary	Sim Window Correction	Sim Window + Departure	Increased Eligibility	Young Age Boost
Marriage/Cohab	0.455	0.431	0.434	0.472	0.447	0.455	0.488	0.497
Casual	0.159	0.152	0.152	0.141	0.152	0.144	0.227	0.183

Table 1.10: Mean Relationship Duration Comparison Summary Table

	Target	Base	Older Partner Offset	Increased Age Boundary	Sim Window Correction	Sim Window + Departure	Increased Eligibility	Young Age Boost
Marriage/Cohab	476	365	364	414	369	387	401	413
Casual	95	103	102	104	103	95	96	96

1.10.3 SUMMARY & DISCUSSION

- no a one-size-fits-all solution
- increased age boundary has some upsides but need to think more through that the actual mean degree target is at that point
- edapprox correction probably not necessary if we're just going to boost some coefs later anyway
- may be other ways to calculate the departure correction but I think this is on the right track
- still concerned about who in the population is sexually active and the under/overestimate of exposure risk
- simulation with no debut term (shown in 1.13) shows that not having a debut term has basically the same results for effective debut as the the young boost scenario, so it's not necessarily even worth having a debut term in the model (yay? oof?)

- several of these approachs have clear upsides and downsides but I think boosting the coefs, and then negatively boosting coefs at older ages might be the best way forward

future work * cross-network terms - probably going to do most analysis on the independent networks but will show both and point at where there are holes (hey by the time this gets finished maybe Chad will have already figured this out)
* need to think about race and sex differences in formation and absdiff(age) by sex if we want to use this for applied work

2

Survival Analysis of Relationship Duration

- main goal here to understand where in the distribution we may not be capturing when we use models based on a memoryless process, and explore some ways to do better within the constraints of feasibility imposed by epidemic models
- so we primarily use the exponential distribution in the SA analyses
- discuss survey data and censoring / truncation issues
- And then flesh out the details of the analyses with a clear narrative and you have the core of a chapter!

The duration of sexual relationships across a population generates the network structure largely responsible for either exposing individuals to or protecting individuals from sexually transmitted infections

(STIs). In addition to dictating this period of possible exposure, relationship durations relative to the pathogen-specific duration of infection are an important driver of how quickly STIs can spread throughout a population. Transmission beyond a pair of actors for infections with short durations relative to relationship lengths is challenging and slow, and it is more likely that an infection will be detected and treated or resolved naturally prior to the dissolution of the relationship. If the duration of infection and duration of relationships are more equal, there is a greater chance that the infection can spread to future partners and throughout the network. When partnerships overlap, transmission pathways increase even among those individuals with few lifetime partners, and this effect is even greater when the duration of overlap is large (Armbruster, Wang, and Morris 2017; Morris and Kretzschmar 1997).

The pattern of relationship durations across the life-course is also important because STIs often have distinct age patterns in terms of prevalence. Individual age is often used as a predictor for risky sexual behavior, but there is additional complexity when considering the effect of age on the duration of relationships across the life-course. Young age likely influences the immediate intentions for relationships (i.e. serious or casual), and the frequency at which individuals form new relationships, but somewhat paradoxically it is also true that the only people who can report extremely long relationships are those who started them at young ages. This also introduces complex sampling issues because most data on relationship durations is collected cross-sectionally or retrospectively – not longitudinally. Given the importance of relationship duration to features of STI epidemiology discussed above, there is growing interest in improving the representation of relational durations in dynamic network models used to study epidemics. This study demonstrates the ways in which the current literature fails to represent this distribution and proposes a new modeling framework to better capture these relationships across the life-course.

One common class of models used to understand network influences on patterns of STI transmission is known as separable temporal exponential-family random graph models (STERGMs). These models are governed by two expressions: one that represents the set of processes that influence the formation of relationships, and a comparable one for dissolution (Krivitsky and Handcock 2014). The current standard practice for the dissolution models in this modeling framework assumes that once a relationship begins, its persistence is governed by a constant hazard. This memoryless process is a convenient simplifying

assumption, adopted because most hypotheses being explored relate to processes impacting network formation or cross-sectional structure. However, it is unlikely that this assumption faithfully represents the distribution all relationship durations we observe across a wide range of ages.

Several recent models have begun to address this issue by splitting out relationships into two categories: the first, marriages and cohabitations or main partnerships, and the second, persistent or casual partnerships. These are then modeled as separate networks simultaneously. By structuring the model in this fashion, each network has a hazard of dissolution specific to its type. (These models often have a third network for one-time sexual contacts which last only one time-step, but this network is not the focus of our study). While these models are indeed able to reproduce the mean relationship lengths drawn from empirical data, it remains unknown how well these strategies reproduce the full distribution of lengths observed. In particular, the memoryless assumption means that the modal length of main partnerships remains near zero across all ages, which basic intuition says is not true and descriptive data analysis confirms. Other work has considered disaggregating relational durations by a single demographic attribute of their members related to a hypothesis or prevention modality being explored, but again with no further effort to capture the full distribution, particularly by age (Goodreau et al. 2017; Jenness et al. 2017).

In this ongoing study, we seek to understand the changing distribution of relationship duration over the life-course using data from the National Survey of Family Growth, to evaluate which features the above dissolution assumptions are capable of replicating and which they cannot. We then introduce an alternative framework designed to more faithfully represent these data and the different demographic and data-collection processes that impact them in an age-structured population over time. We use tools from both event history analysis and network analysis to answer the following questions: First, under what circumstances, if any, can an exponentially distributed time-to-event model reasonably approximate empirical relationship duration data? Second, does it make sense to lump marriages and cohabitations into one network with the same dissolution probability? And third, can we better capture the age-wise relationship distribution by using one network where (1) relationships can transition between states (e.g. from a cohabitation into a marriage) rather than modeling several types separately, and (2) where relational formation probabilities depend on current relational status.

Methods

First the empirical relational duration data (using the start and end dates of all reported relationships in NSFG) will be investigated using histograms (overall and stratified by age category). Then, due to the issues of right censoring and left truncation as a result of survey design in NSFG, a reference survival curve will be constructed from the empirical data using a Modified Kaplan-Meier model following (Burington et al. 2010) and using the R package 'survival'. Next, exploratory parametric models will be estimated from the data (with corrections for right-censoring and left-truncation) using a variety of distributions (namely and latent mixture components) to gain intuition about the underlying generative processes. Initial models will be covariate-free (representing the effects of relationship duration only on the chances of survival) and additional models will begin to examine the influence of age on relationship duration, including (but not limited to) the ego's age category, the reported current age difference between ego and alter, and the ego's age category at the beginning of each reported relationship. All models without latent components will be fit using the R package 'flexsurv' and the likelihood functions for all models with latent components will be personally developed and models will be fit using the maxLik package (Jackson 2016; Henningsen and Toomet 2011).

From PAA abstract: In our preliminary work, we first checked the assumption that relationship duration can be modeled by a simple memoryless process, and then explored some natural extensions to this framework. In order to generate the reference distribution, we fit a Kaplan-Meier model using a modified estimator to account for both right-censoring and left-truncation following Burington et al (2010). We then fit several parametric models (all adjusted for the above sampling issues): first a simple exponential model to represent the memoryless process assumption, then a Weibull distribution and Gamma distribution, all with and without additional covariate attributes. Model fit was evaluated primarily using the Akaike Information Criterion (AIC) and visuals to understand which relationship lengths were represented better than others, given our ultimate goal of adapting these into dynamic network simulations. Below is a selection of explored models; parametric models with covariate categories are displayed in color, with their Kaplan-Meier reference curves plotted in black. All parameters in these fitted models are statistically significant ($p < 0.001$).
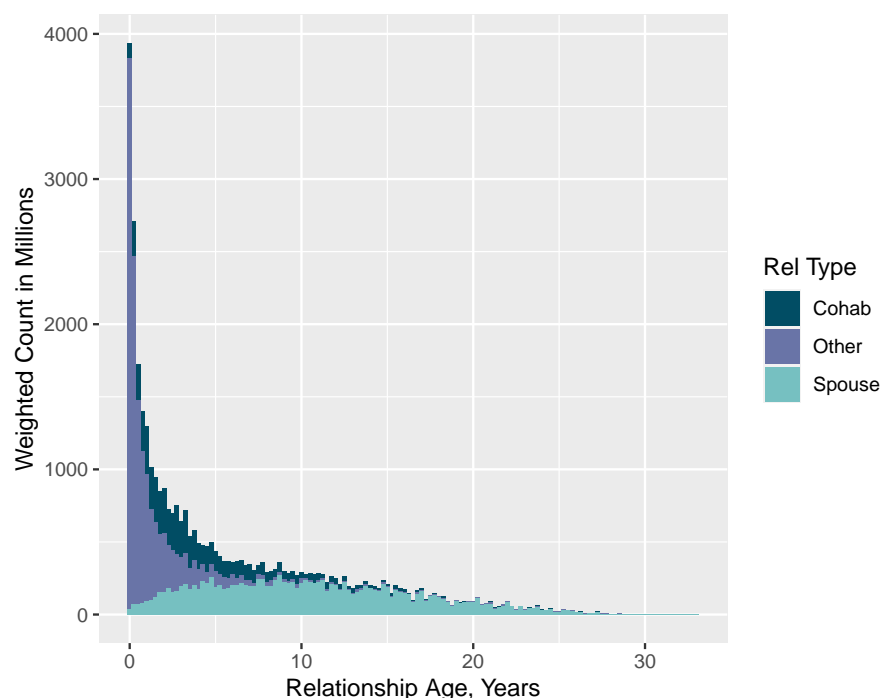
Figure 2.1: All Relationships either Current or Ended in the Last Year

Initial Histograms Results

The first takeaway is that an exponential distribution alone is not sufficient to capture the relationship distribution – it overestimates the survival of short relationships and underestimates the survival of long relationships (top left figure, below). The Weibull and Gamma perform better and capture more of the short relationships, suggesting that there is important heterogeneity in the data, but like the first models they also fail to capture the longest relationships. The age category of the reporting individual is not explanatory across all age categories (top right). This is perhaps not surprising, in that the age distribution of relationship lengths is at least partly an emergent property rather than a causal one. That is, no individual can have a relationship that has lasted longer than they have been sexually active, so the range of relationship lengths for young age categories is relatively small. Meanwhile, the older age categories are challenging to represent because the possible range of relationships is so much larger, and are likely influenced not only by dissolution probabilities but also by the changing formation probabilities over the
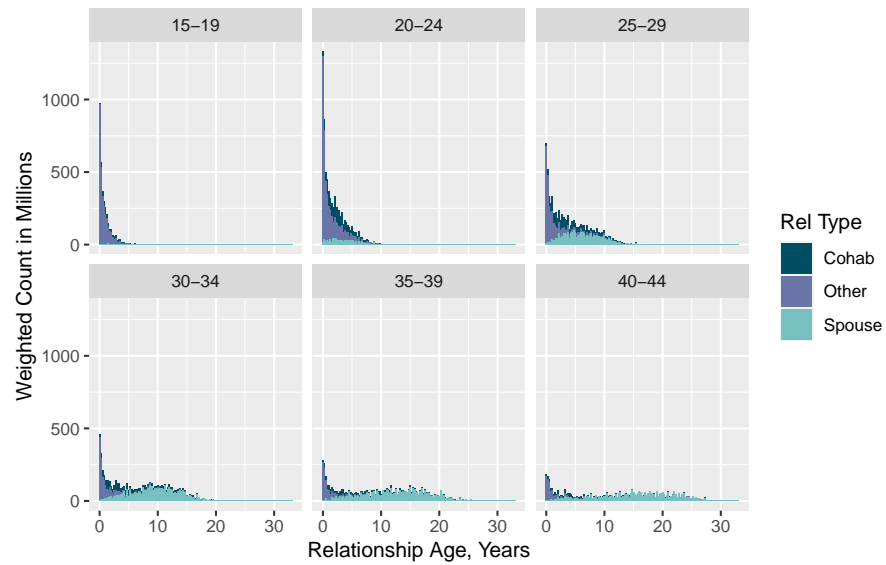
Figure 2.2: All Relationships either Current or Ended in the Last Year, by Age Category

lifecourse – that is, older people in long-term relationships do not start new relationships at the same rate as others, and thus have relatively few relationships that are short.

Survival of Relationships
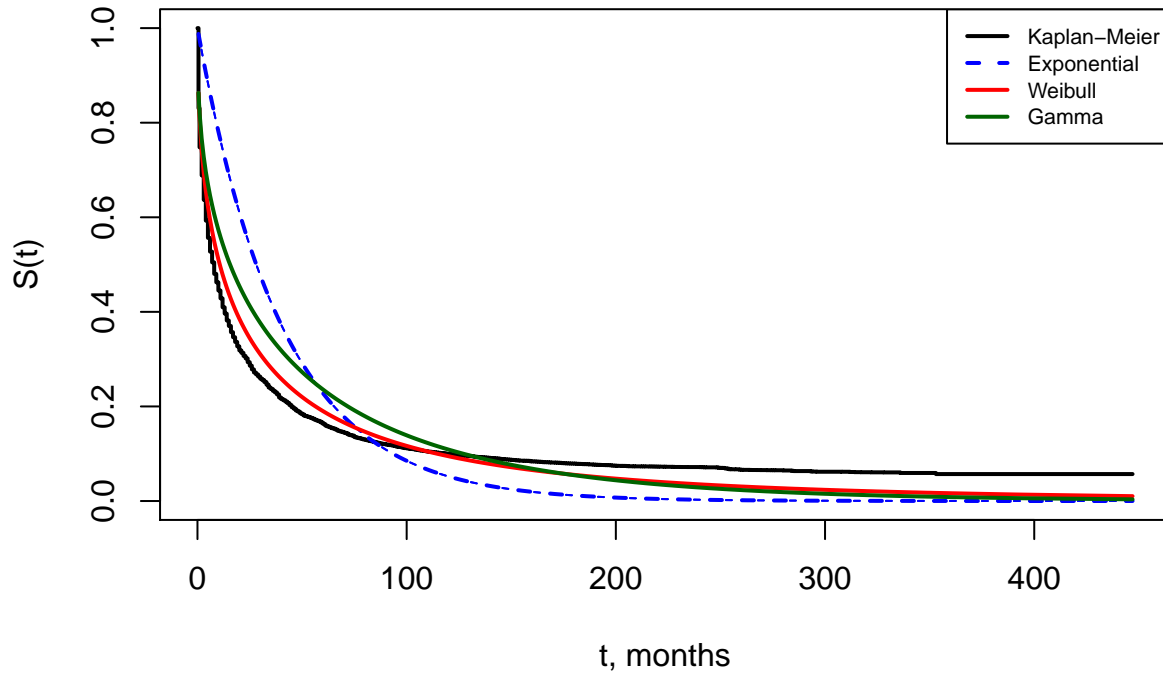
The next two models test how appropriate it is to group relationships defined as marriages and cohabitations into the same dissolution model, as has been done in recent STERGMs. We see clear evidence that marriages and cohabitations have distinct hazards of dissolution and the combined marriage and cohabitation curve, like the simple exponential for all relationships, dramatically fails to capture both the shortest and longest relationships of these types (bottom right and bottom left figures, respectively. These results are similar to other work in family demography that has shown significant differences in the risk of dissolution between cohabitations and marriages due to variation in joint lifestyles (van Houdt and Poortman 2018). These results suggest to us that previously developed STERGM dissolution models that only capture the mean relationship length are not appropriate approximations of the data, and that cohabitation represents a distinctly separate type of relationship from marriages and other casual relationships and should be treated as such in our networks.
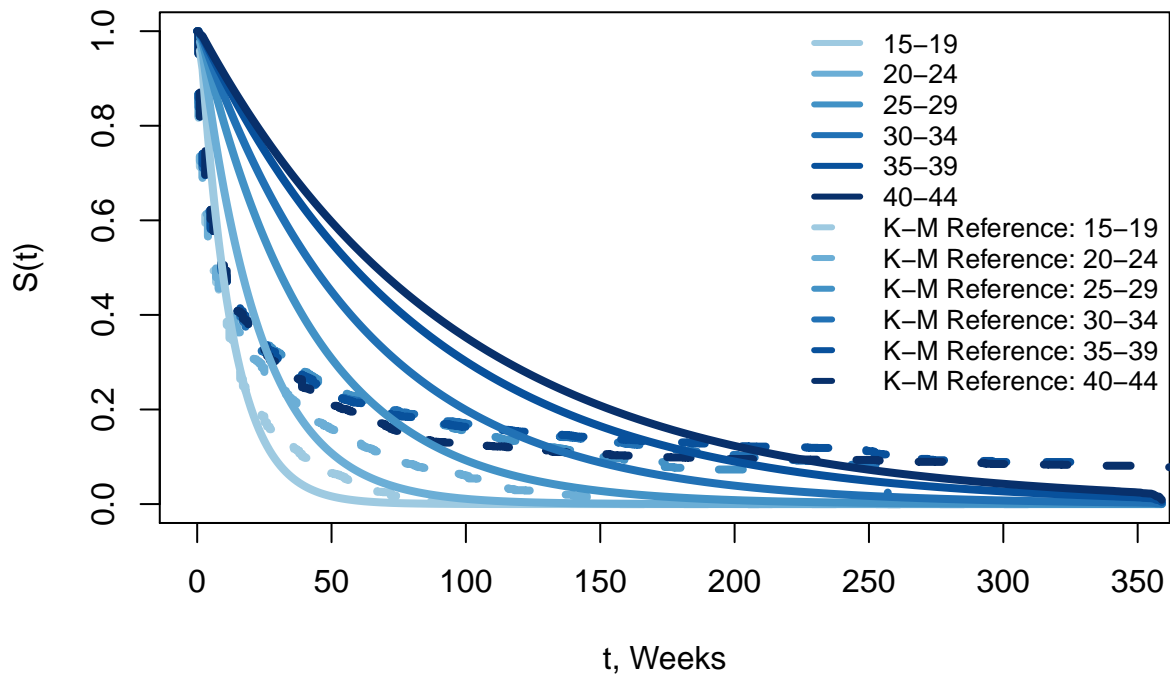
Figure 2.3: Kaplan-Meier vs. Constant Hazard by Current Age Category
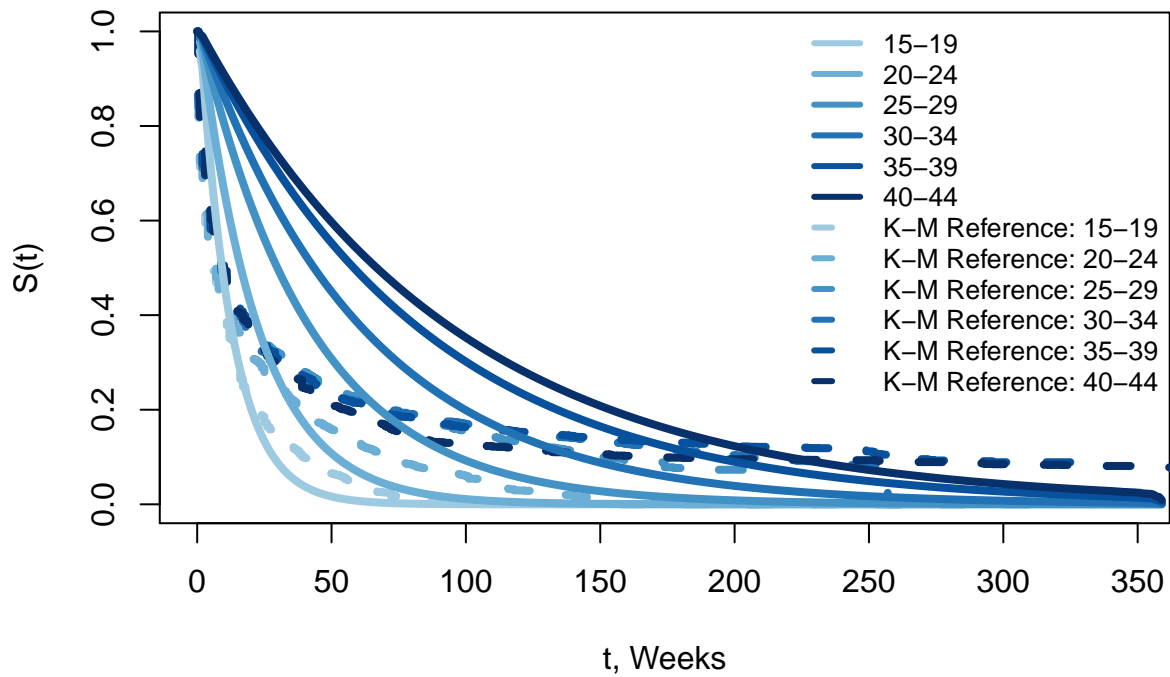
Figure 2.4: Kaplan-Meier vs. Constant Hazard by Current Age Category

3

Chlamydia, Acquired Immunity, & Expedited Partner Therapy?

copying over some text from diss proposal:

C. trachomatis is an obligate intracellular bacterium transmitted through sexual contact among humans. Chlamydial infections are most often asymptomatic. Untreated infections in women are an additional public health concern because they can lead to a variety of sequalae including pelvic inflammatory disease, scarring of ovaries and fallopian tubes, ectopic pregnancies, chronic pain, and infertility. Repeat infections are common and are an additional risk factor for the development of the above sequelae (Brunham and Rey-Ladino 2005). There is a great deal of uncertainty regarding the natural history of chlamydia, but the

duration of infection for untreated individuals is generally thought to be up to 6 months for men and a year or more for women (Golden et al. 2000; Satterwhite et al. 2013). Chlamydia is usually treated with azithromycin or doxycycline, and unlike other common STIs like syphilis and gonorrhea, true antibiotic resistance is rare (Kong et al. 2015).

Chlamydia is the most common reportable disease in the United States and incidence, particularly adolescents and young adults aged 15-29, is increasing nationwide. The Centers for Disease Control and Prevention (CDC) estimates that half of all new STI infections (including gonorrhea, syphilis, and others) occur in those aged 15-24 despite them making up only a quarter of the sexually active population. Even in places like King County, Washington, where overall rates have remained stable, longstanding acknowledged disparities in prevalence by race are marked and continue to increase (2015 SKCPH STD Report). These rates are particularly distressing in light of the fertility consequences of long-term infection and reinfection: it is estimated that in King County, over 60% of non-Hispanic Black women have had at least one chlamydia infection by age 34 (a rate 5x higher than non-Hispanic White women) and 1 in 500 of non-Hispanic Black women develops chlamydia-associated tubal factor infertility over their life-course (Chambers et al. 2018).

The United States has some of the highest STD rates in the industrialized world, and despite this, funding for public health programs dedicated to these issues has largely declined (CDC 2016 STD Report). As a result, few health departments are able to offer traditional partner notification services, where a patient who tests positive for an STI gives the contact information of their recent sex partners to the health department, and the department then contacts their partners with the hope that these partners will then get tested and, if necessary, treated. Expedited partner therapy (EPT) was developed with this scenario in mind (See figure 2). Under EPT, a patient who tests positive, upon receipt of their own treatment, receives either additional antibiotic pills for their recent sexual partners or prescriptions for treatment that their partners can fill. The patient then is expected to hand-deliver either the treatment or prescription to their partner(s), who take the medicine at their own discretion and without the need for a positive lab test. By using these actors to essentially leverage their sexual network in reverse, this system hopes to decrease the time to treatment for all possible infected partners and increase the total number of partners treated.

It can also reduce re-infection among the index patients if the partnerships are ongoing. There have been several clinical trials of EPT across the US (and Europe), including Washington State. These trails demonstrated that relative to traditional referral practices, EPT provision increased the proportion of partners who were ultimately treated, reduced the number of individuals who were re-infected at follow-up, and was less costly if at least 30% of partners were treated via EPT (CITE). Despite these results and a growing body of evidence in support, widespread implementation of EPT has been slow and there are still many questions to be answered.

also – EPT as a tool for health equity, not just effective population-level decrease in prevalence – can be more effective in high-prevalance groups?

Annals of Internal Medicine Article High Incidence of New Sexually Transmitted Infections in the Year following a Sexually Transmitted Infection: A Case for Rescreening - Peterman et al

Arrested Immunity Hypothesis One of the paradoxes in era of modern public health is that chlamydia incidence has actually increased overall in the presence of mass control programs. In Sweden, Norway, Finland and Canada the rates initially decreased but then resumed increasing, and in Australia, United States, and the United Kingdom the rates never stopped increasing even after program initiation, although this second pattern has been attributed to the challenges of implementing control programs consistently throughout a large population (Brunham and Rekart 2008). These areas now experience incidence rates higher than rates prior to introduction of control programs. Additionally, a regression analysis using data from family planning clinics in Region X of the United States (Alaska, Washington, Idaho, and Oregon) found that, after controlling for any changes in demographics, sexual behaviors, and increased sensitivity of clinical tests, there was a remaining 5% 'true' and unexplained annual increase in chlamydia positivity from 1997-2004 (Fine et al. 2008). In response to these and other examples of unabated chlamydia infection in the presence of control programs, Brunham and Reckart have proposed the arrested immunity hypothesis (Brunham and Rekart 2008). Under this hypothesis, early detection and treatment of chlamydia interrupts the development of acquired immunity, making treated individuals particularly vulnerable to reinfection almost immediately after treatment. While we have no natural history studies of chlamydia infection in humans that address the development, duration, and extent of immunity, there is growing

evidence beyond rodent models and trends in incidence that partial immunity can develop and play a role. Rodent models of chlamydial infection suggest that a high proportion are able to resolve their primary infection and are temporarily resistant to infection. Rodents that then eventually become reinfected with chlamydia have a shorter duration of disease, lower pathogen load and decreased inflammatory response (Rank et al. 2003). However, it has also been shown that treatment early in the course of infection interrupts the development of this protective immunity (Su et al. 2002). There is also some indirect evidence in humans. A 2010 review article acknowledged that in several studies of infection status among couples, the rates of discordance (i.e. one partner is infected while the other is not), are higher for chlamydia than for gonorrhea and that this discordance increases with age, providing indirect evidence for some level of protective immunity to chlamydia that increases with age, likely due to exposure over time. There is little immunity that develops to gonorrheal infection due to high levels of antigenic variation (Batteiger et al. 2010). Recent modeling using data from both the UK and United States has demonstrated that at least some immunity to chlamydia following natural clearance is necessary to generate observed patterns in incidence (Omori, Chemaitelly, and Abu-Raddad 2019). These questions are particularly relevant in the context of expedited partner therapy, where the goal is to interrupt transmission by treated individuals and their partners as quickly as possible. However, due to the arrested immunity of those treated quickly, if the timing of delivery and uptake of partners is not sufficient, the initially treated is likely at higher risk of reinfection than under the standard referral scenario. If sufficient numbers of partners are treated effectively and quickly and transmission throughout the network is greatly diminished, then EPT may be able to overcome the effects of this arrested immunity.

Conclusion

We conclude.

A

The First Appendix

A.1 Data and Model Terms

include plots for the other terms - absdiff(sqrtage), concurrent, debuted etc

several general trends in relationship formation (finish write-up and cite) –

- individuals often select partners that are not their exact age
- this difference in partner ages often increases over the life course (i.e. adults usually have wider age differences between their partners than do adolescents)
- it is common for men to partner with younger women (although the sex differences in relationship
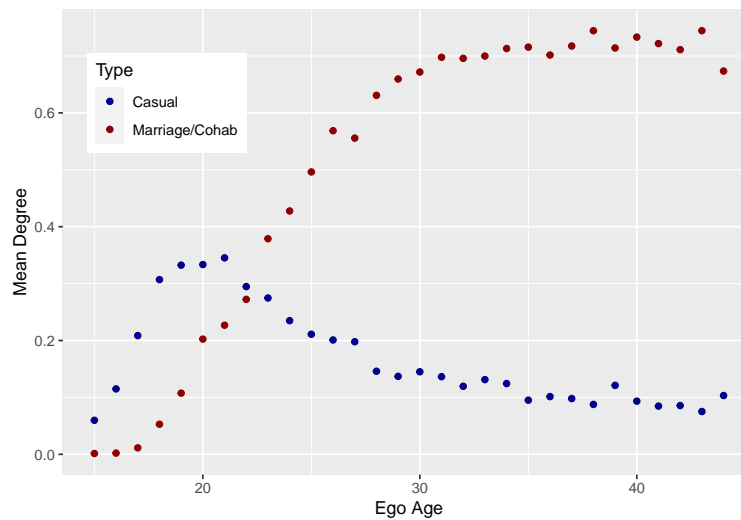
Figure A.1: Mean Degree by Ego Age and Relationship Type.

formation are not explored in this model, it's important to note that in a more realistic model the effect of aging out would disproportionately affect the women whose partners age out before them

(include model terms and coefs and explain terms) (full description of EpiModelHIV module flow w/ parameters in appendix, brief overview here)

Cross network terms - we're going to avoid them due to complications

B

The Second Appendix

more technical stuff in here?

Colophon

This document is set in EB Garamond, Source Code Pro and Lato. The body text is set at 11pt with *EBGaramond*(3).

It was written in R Markdown and *LaTeX*, and rendered into PDF using huskydown and bookdown.

This document was typeset using the XƎTeX typesetting system, and the University of Washington Thesis class class created by Jim Fox. Under the hood, the University of Washington Thesis LaTeX template is used to ensure that documents conform precisely to submission standards. Other elements of the document formatting source code have been taken from the Latex, Knitr, and RMarkdown templates for UC Berkeley's graduate thesis, and Dissertate: a LaTeX dissertation template to support the production and typesetting of a PhD dissertation at Harvard, Princeton, and NYU

The source files for this thesis, along with all the data files, have been organised into an R package, xxx, which is available at `https://github.com/xxx/xxx`. A hard copy of the thesis can be found in the University of Washington library.

This version of the thesis was generated on 2020-10-12 19:53:59. The repository is currently at this commit:

The computational environment that was used to generate this version is as follows:

```
- Session info --------------------------------------------------------------
 setting  value
 version  R version 3.6.1 (2019-07-05)
```

```
os        macOS Catalina 10.15.3
system    x86_64, darwin15.6.0
ui        X11
language  (EN)
collate   en_US.UTF-8
ctype     en_US.UTF-8
tz        America/Los_Angeles
date      2020-10-12
```

```
- Packages ------------------------------------------------------------------
package      * version     date       lib
ape            5.3         2019-03-17 [1]
assertthat     0.2.1       2019-03-21 [1]
backports      1.1.9       2020-08-24 [1]
bookdown       0.20.2      2020-08-06 [1]
broom          0.5.2       2019-04-07 [1]
callr          3.4.3       2020-03-28 [1]
cellranger     1.1.0       2016-07-27 [1]
cli            2.0.2       2020-02-28 [1]
coda           0.19-3      2019-07-05 [1]
codetools      0.2-16      2018-12-24 [1]
colorspace     1.4-1       2019-03-18 [1]
crayon         1.3.4       2017-09-16 [1]
data.table     1.12.8      2019-12-09 [1]
DBI            1.1.0       2019-12-15 [1]
ddaf         * 0.0.0.9000  2020-10-07 [1]
DEoptimR       1.0-8       2016-11-19 [1]
desc           1.2.0       2018-05-01 [1]
```

```
deSolve      * 1.27.1    2020-01-02 [1]

devtools     * 2.3.1     2020-07-21 [1]

digest         0.6.25    2020-02-23 [1]

doParallel     1.0.15    2019-08-02 [1]

dplyr        * 1.0.2     2020-08-18 [1]

ellipsis       0.3.1     2020-05-15 [1]

EpiModel     * 1.7.5     2020-01-07 [1]

ergm         * 3.10.4    2019-06-10 [1]

evaluate       0.14      2019-05-28 [1]

fansi          0.4.1     2020-01-08 [1]

farver         2.0.3     2020-01-16 [1]

flexsurv       1.1.1     2019-03-18 [1]

forcats      * 0.4.0     2019-02-17 [1]

foreach        1.4.7     2019-07-27 [1]

fs             1.5.0     2020-07-31 [1]

generics       0.0.2     2018-11-29 [1]

ggfortify    * 0.4.7     2019-05-26 [1]

ggplot2      * 3.3.2     2020-06-19 [1]

ggpubr       * 0.2.2     2019-08-07 [1]

ggsignif       0.6.0     2019-08-08 [1]

ggthemes     * 4.2.0     2019-05-13 [1]

git2r          0.27.1    2020-05-03 [1]

glue           1.4.1     2020-05-13 [1]

gridExtra    * 2.3       2017-09-09 [1]

gtable         0.3.0     2019-03-25 [1]

haven          2.1.1     2019-07-04 [1]

here         * 0.1       2017-05-28 [1]

hms            0.5.0     2019-07-09 [1]
```

```
htmltools          0.5.0      2020-06-16 [1]

httr               1.4.2      2020-07-20 [1]

huskydown        * 0.0.5      2020-08-06 [1]

iterators          1.0.12     2019-07-26 [1]

jsonlite           1.7.0      2020-06-25 [1]

kableExtra       * 1.1.0      2019-03-16 [1]

km.ci              0.5-2      2009-08-30 [1]

KMsurv             0.1-5      2012-12-03 [1]

knitr              1.29       2020-06-23 [1]

labeling           0.3        2014-08-23 [1]

lattice            0.20-38    2018-11-04 [1]

lazyeval           0.2.2      2019-03-15 [1]

lifecycle          0.2.0      2020-03-06 [1]

lpSolve            5.6.13.3   2019-08-19 [1]

lubridate          1.7.4      2018-04-11 [1]

magrittr         * 1.5        2014-11-22 [1]

MASS               7.3-51.4   2019-03-31 [1]

Matrix             1.2-17     2019-03-22 [1]

memoise            1.1.0      2017-04-21 [1]

mitools            2.4        2019-04-26 [1]

modelr             0.1.4      2019-02-18 [1]

mstate             0.2.11     2018-04-09 [1]

muhaz              1.2.6.1    2019-01-26 [1]

munsell            0.5.0      2018-06-12 [1]

mvtnorm            1.0-11     2019-06-19 [1]

network          * 1.16.0     2019-12-01 [1]

networkDynamic   * 0.10.0     2019-04-05 [1]

nlme               3.1-140    2019-05-12 [1]
```

```
pillar              1.4.6      2020-07-10 [1]

pkgbuild            1.1.0      2020-07-13 [1]

pkgconfig           2.0.3      2019-09-22 [1]

pkgload             1.1.0      2020-05-29 [1]

prettyunits         1.1.1      2020-01-24 [1]

processx            3.4.3      2020-07-05 [1]

ps                  1.3.4      2020-08-11 [1]

purrr             * 0.3.4      2020-04-17 [1]

quadprog            1.5-7      2019-05-06 [1]

R6                  2.4.1      2019-11-12 [1]

RColorBrewer      * 1.1-2      2014-12-07 [1]

Rcpp                1.0.5      2020-07-06 [1]

readr             * 1.3.1      2018-12-21 [1]

readxl              1.3.1      2019-03-13 [1]

remotes             2.2.0      2020-07-21 [1]

rlang               0.4.7      2020-07-09 [1]

rmarkdown           2.3        2020-06-18 [1]

robustbase          0.93-5     2019-05-12 [1]

rprojroot           1.3-2      2018-01-03 [1]

rstudioapi          0.11       2020-02-07 [1]

rvest               0.3.4      2019-05-15 [1]

scales              1.1.1      2020-05-11 [1]

sessioninfo         1.1.1      2018-11-05 [1]

srvyr               0.4.0      2020-07-30 [1]

statnet.common      4.3.0      2019-06-02 [1]

stringi             1.4.6      2020-02-17 [1]

stringr           * 1.4.0      2019-02-10 [1]

survey              4.0        2020-04-03 [1]
```

```
survival          2.44-1.1   2019-04-01 [1]

survminer       * 0.4.5      2019-08-03 [1]

survMisc          0.5.5      2018-07-05 [1]

tergm           * 3.6.1      2019-06-12 [1]

testthat          2.3.2      2020-03-02 [1]

tibble          * 3.0.3      2020-07-10 [1]

tidyr           * 1.1.1      2020-07-31 [1]

tidyselect        1.1.0      2020-05-11 [1]

tidyverse       * 1.2.1      2017-11-14 [1]

trust             0.1-8      2020-01-10 [1]

usethis         * 1.6.1      2020-04-29 [1]

vctrs             0.3.2      2020-07-15 [1]

viridisLite       0.3.0      2018-02-01 [1]

webshot           0.5.1      2018-09-28 [1]

withr             2.2.0      2020-04-20 [1]

xfun              0.16       2020-07-24 [1]

xml2              1.3.2      2020-04-23 [1]

xtable            1.8-4      2019-04-21 [1]

yaml              2.2.1      2020-02-01 [1]

zoo               1.8-6      2019-05-28 [1]

source

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.2)

Github (rstudio/bookdown@f9cf1ac)

CRAN (R 3.6.0)

CRAN (R 3.6.2)

CRAN (R 3.6.0)
```

```
CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.1)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

local

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.2)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.2)

CRAN (R 3.6.2)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.2)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.2)
```

```
CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.2)

CRAN (R 3.6.2)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.2)

CRAN (R 3.6.2)

Github (benmarwick/huskydown@a909835)

CRAN (R 3.6.0)

CRAN (R 3.6.2)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.2)

CRAN (R 3.6.0)

CRAN (R 3.6.1)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.1)

CRAN (R 3.6.1)
```

```
CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.1)

CRAN (R 3.6.2)

CRAN (R 3.6.2)

CRAN (R 3.6.0)

CRAN (R 3.6.2)

CRAN (R 3.6.0)

CRAN (R 3.6.2)

CRAN (R 3.6.2)

CRAN (R 3.6.2)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.2)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.2)

CRAN (R 3.6.2)

CRAN (R 3.6.2)

CRAN (R 3.6.0)
```

```
CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.2)

CRAN (R 3.6.0)

CRAN (R 3.6.2)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.2)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.2)

CRAN (R 3.6.2)

CRAN (R 3.6.2)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.2)

CRAN (R 3.6.2)

CRAN (R 3.6.0)

CRAN (R 3.6.0)

CRAN (R 3.6.2)

CRAN (R 3.6.2)

CRAN (R 3.6.2)

CRAN (R 3.6.0)
```

```
CRAN (R 3.6.0)

CRAN (R 3.6.0)
```

```
[1] /Library/Frameworks/R.framework/Versions/3.6/Resources/library
```

References

Carnegie, N. B., Krivitsky, P. N., Hunter, D. R., & Goodreau, S. M. (2015). An Approximation Method for Improving Dynamic Network Model Fitting. *Journal of Computational and Graphical Statistics*, *24*(2), 502–519. http://doi.org/10.1080/10618600.2014.903087

Singer, M. C., Erickson, P. I., Badiane, L., Diaz, R., Ortiz, D., Abraham, T., & Nicolaysen, A. M. (2006). Syndemics, sex and the city: Understanding sexually transmitted diseases in social and cultural context. *Social Science and Medicine*, *63*(8), 2010–2021. http://doi.org/10.1016/j.socscimed.2006.05.012