**Summaries of reviews, modeling papers, or empirical analyses on the algorithms or effects of social mixing in networks or assortative mixing by risk level, age, sex on disease prevalence overall or sex-specific prevalence**

Lancichinetti et al. ()

(Sah et al. 2014): Propose a model for generating simple, connected random networks that have a specified degree distribution and level of community structure. They say conveniently that “these networks allow for systematic study of the impact of community structure on the dynamics that may flow on a network”. Perfect. They also compare and contrast previous approaches to generate community-structured networks. Mention Girvan and Newman (2002) which generates community-structure by assigning different probabilities between in and out of community nodes. They say this has been used a lot but limited ability to have different community sizes or heterogeneous degree distributions. Lancichinetti et al. () seem to have developed something where degrees can be heterogeneous so should take a look at that. Other models have been proposed but Sah et al. (2014) say they make strong assumptions about parameters or that the parameters are not directly related to typical parameters like degree distribution and modularity. They say you can use ERGMs but when you squish a bunch together, some difficulties arise (but don’t say what they are?). Lastly, they explain *stochastic block models* where the probability of each link depends only on the community labels of its endpoints so networks must specify number of communities K, their sizes and the group pair probabilities which results in a random graph with a degree distribution that follows a mixture of Poisson distributions, which can be unrealistic. They basically say adapting SBMs for different tasks is a promising area of research which is what they do in this paper.

Their model: can be large (). They present a nice parameterization of Q, the strength of community structure. Model is parameterized by network size (n), degree distribution (expected modularity , number of modules , module size distribution , with mean Algorithm:

1. Assign n nodes to K modules based on the size distribution P(s)
2. Assign degrees, to each node based on and . Then, of these neighbors of v, get # of within group contacts from same class of distribution as but with mean instead.
   1. Ensure draws follow with mean and verify conditions:
      1. for all
      2. A realizable within-degree sequence for each module as defined by Handshake Theorem --- Unclear what this means….
   2. Generate between group degree: by for each node and verifying some criteria
3. Assign between-group edges based on a Havel-Hakimi model and randomize them.
   1. Successively connect high-between degree nodes to each other if not already connected
      1. Then between-group edges are swapped using double-edge swaps (as long as new edge maintains between group identity)
4. Connect within-edges based on Havel-Hakimi and randomize them to remove degree correlations.

(Girvan and Newman 2002): Show community structure is a defining feature of many networks. Review methods of detecting community structure. Introduce new method of detecting community structure based on “edge-betweenness” described in definitions below. Show that this method is very reliable for random networks (constructed by > ). So the graph amounts to Erdos-Renyi graph with equal sized modules but fixed probabilities are used to produce edges between modules. The rest of the article talks about applications. Note that their algorithm is slow for big graphs. Also networks are usually more heterogeneous and big than what this algorithm can do.

(de Almeida et al. 2013): How can you include this “Birds of the feather flock together” proverb into BA model? This paper shows how to develop a scale-free homophilic model. Let “*Similtude*” be the difference in intrinsic characteristics of two node types (I think they were thinking of distance between spatial sites?) . The network starts with sites characterized by which is unchanged in time and assigns randomly a value ranging from 0 to 1. At every time step, a new site with randomly attributed attaches to other pre-existing sites by undirected and unweighted new link. The connections are established by considering jointly the connections between new sites and those having high number of nearest neighbors and high similitude using equation

and this is repeated up to the size desired for the system where t is the time variable. They analyze networks with two types of nodes (0 or 1). Node types are bounded in a sub-interval within the larger interval from 0 to 1. corresponds to a value within that sub-interval BUT I’m not sure how this sub-interval is defined… Seems very confusing and I’m frustrated by it. Seems that you set a and then define the intervals of group 1 as the upper and lower limit of the being used… Then it says that group 0 sub-interval is “composed of sites with out considered”? Whatever that means… Need to figure out if going to use it.

Then, they measure dyadicity and heterophilicity according to Park and Barabasi 2007 (see definitions list) to measure mixing between sites.

With parameters, and m=3, networks are dependent on . and are defined in the following:

and

.

They show that the dyadicity and homophilicity (mixing between 1’s) is dependent on the choice of .

I’m not entirely clear if/how this would correspond to “intrinsic characteristics” of males and females in our model yet. They go on to show that their model follows a power-law which is good. I think this is a good model to use but I’m really confused by the notation surrounding the parameters.

(Newman 2002): What are the properties of networks assorted by degrees? This is a short paper describing basic findings of Newman 2003 for assortativity based on degree. One interesting finding is that assortative networks (high degree with high degree) are redundant (i.e., removing one highly connected node doesn’t do much to overall network connectivity because they are so clustered with other highly connected nodes).

(Newman 2003)**:** How can you assemble models of discrete assortative networks? Analytical methods using MGF’s are given in part b, but I’m struggling to follow the full derivations. Simulation methods are given in part c, which are still complicated. (1) Choose number of edges, M. Generate M edges, each one identified by the types of vertices that it connects, in some manner such that the fraction of edges between males and females tends to as M becomes large. “In practice, a simple transformation method works well (Newman 1999)”. (2) Count number of ends of edges of each type i, to give sums of the degrees of vertices of each class. Calculate expected number of vertices of each type from rounded to nearest integer and is desired mean degree of type i. (3) Draw vertices from desired degree distribution for vertices of type i, . (4) pair up ends of edges with vertices generated, so that each vertex has the # of attached edges corresponding to its chosen degree. (5) repeat step (3) for each vertex type. What’s important to remember from this paper about assortatively mixed networks? Assortatively mixed networks might percolate at lower edge densities and are more robust to edge removal than random or disassorative networks.

(Bansal, Grenfell, and Meyers 2007)**:** Are realistic contact patterns similar to scale-free or exponential? All found to be more similar to exponential. How good can homogenous mixing models approximate more realistic contact patterns? For highly connected exponential networks, the homogeneous-mixing compartmental models offer reasonable approximation to epidemic size. As the tail of the degree distribution gets fatter, what happens to the predictions from homogeneous mixing models? Predicted disease dynamics from fat tails differ from homogeneous mixing models. What can I return to this paper for? For a good review of earlier ways to model heterogeneity in contact patterns using networks. In addition to questions answered in summary above and a source for network definitions of vocabulary, they also give some analytical expressions for solving epidemics on network models.

(EAMES 2007)**:** Argue that contact-tracing is a powerful public health tool that could in some cases be made more powerful. This modeling paper asks how useful contact tracing is when networks are random-mixing or (dis)assortative? They also investigate targeted interventions by asking, how much better can contact tracing be if we target it towards individuals at higher risk? They use a “pair-wise” SI model that considers pairs of individuals as the basic variable (see vocab). They show you can reduce the overall prevalence some by doing single-step tracing (see vocab), especially when r is close to 1 but for eradication, the space is limited. In contrast, equilibrium prevalence can be greatly reduced by iterative tracing which is introduced into their model as a “hyper-infection” which spreads through the population into “tracing class” [T]. They remain in tracing class for units. With this model, they investigate how disease is concentrated into different sub-groups of the population, finding that as equilibrium prevalence increases, the more spread out the infection is across the population. The lower prevalence is, the more disease is situated in certain parts of the network (i.e., individuals with lots of infected contacts?). They show how iterative tracing, not single-step, can greatly reduce the prevalence, especially among sub-groups, i.e., the ability of core-group to sustain infection is greatly reduced. Next, they study mixing between high and low risk groups. Super important: R0 is maximized with high assortativity because transmission is intense and spreadable among condensed high-risk individuals but PREVALENCE is maximized at intermediate assortativity because enough transmission in high risk but also enough transmission to low risk. Then he proceeds to see if tracing is better in assortative or randomly-mixed populations. Contact tracing requires lower relative effort in assortative networks to eradicate infection. In random mixing, there is a relatively even distribution of infection meaning that targeting is of little worth. The main thing is to figure out where are the core groups.

(Perkins, Ferrari, and Hudson 2008)**:** Used a natural wild rodent – parasite system where there is no sex-bias in prevalence but a strong sex-bias in transmission has been observed (Ferrari 2004). In this study, asked to what extent is the observed overall prevalence and prevalence by sex in this system due to observed sex-assortativity? Also, asked what strength of male-biased transmission predicts empirical patterns of macroparasite prevalence and sex ratio of infected hosts? Found disassortative mixing by sex (-0.13) which was somewhat surprising given the breeding biology of mice. With these disassortative networks, they simulated outbreaks with chain binomial model where I is infected neighbors in the network and is the sex-specific shedding rates. *Note: I wonder how difficult this is in larger networks*. They determined by and and so by increasing c, you increase . They found that without male-bias in shedding, the overall prevalence and prevalence by sex in this system does not match up with field results. But when you increase c all the way up to 10, then simulation results match with field results… Thus, they conclude that heterogeneity at the individual level (differences in shedding) not mixing-patterns at the population-level for macroparasite dynamics analyzed here. They say future work entails figuring out whether network characteristics (I assume centrality) are associated with parasite burden, which could have implications for their findings.

**Other relevant papers**

* Malloy & Reed (1995): configuration models
* (Read et al. 2012; Danon et al. 2012): review of contact patterns relevant to respiratory transmitted infection
* (Mossong et al. 2008): age-structured contacts
* (Dodd et al. 2016): age-sex prevalence in TB and finding that males drive transmission
* Felipe et al (2005): deterministic age-sex-structured model with differences in transmission rates
* Newman 1999: “Monte Carlo Methods in Statistical Physics”
* Garnett & Anderson (1996): STDs and sexual behavior: insights from math models, assortativity results?
* Barlow et al. (1997): assortative mixing, a limiting factor in HIV spread
* Edumunds WJ, OCallaghan, and Nokes 1997: Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections
* Garnett GP et al. (1996): Sexual mixing patterns of patients attending sexually transmitted disease clinics
* Boguna, Pastor-Satorras, Vespignani (2003): Epidemic spreading in complex networks with degree correlations.
* Hethcote & York 1984: modeled STI in male/female/high/low risk groups

**Definitions**

**Degree-based differential equations:** Models track numbers of individuals in each disease state for each degree class (there are 3k differential equations for a SIR model). See Pastor-Satorras & Vespignani (2002) for details on this model.

**Dyadicity and heterophilicity**: Ways to measure homophily in networks *(as calculated by Park and Barabasi 2007):*Assume two states s with values s=1 or s=0 where there are and of each in the desired network ( and the total number of links, M, satisfying . Note that and but subtler constraints exist due to the network structure and number of each type (see more info below). If s is *distributed randomly* among N sites, the *expected values* of and :

Where is the average probability that any two sites are linked. If and display significant deviations, it implies that s=1 is not distributed randomly. To quantify these deviations, they use the definition introduced by Park and Barabasi 2007:

The dyadicity (D) can be thought of as the connections between 1-1 pairs and the heterophilicity (H) can be thought of as the connections between 1-0 pair. If D>1, then 1’s tend to link more closely among themselves than expected for a random configuration. If D<1, then 1’s link less densely amongst themselves. Analogous quantitative interpretation can be done with H.

The different *possible* configurations of D and H networks can be visualized using a phase diagram and depends on the network size and type (e.g., ER or SF). The phase boundary is stationary (system size independent).

**Edge betweenness and community structure:** Can be used to detect community structure (Girvan and Newman 2002). Focus on edges that are least central, i.e., “most between” communities. Progressively remove edges from original graph with these steps: 1. Calculate betweenness for edges. 2. Remove edge w/highest betweenness. 3. Recalculate for all edges affected by removal. 4. Repeat until no edges remain.

**Hierarchical clustering:** Can be used to detect community structure in networks but is flawed for nodes that are loosely connected to anyone. Calculate some weight between two vertices in the network which represents some sense of how close they are. Add edges between unconnected nodes until the network is connected. Choosing weights usually based on number of paths between nodes.

**Iterative tracing**: when infected partners of any diagnosed cases being sought for treatment and further tracing (e.g., HIV today).

**Pair-wise (pair-counting) model**: Count the number of pairs of individuals in each disease class. More simple than full stochastic simulation on networks and because of this cannot capture large-scale dynamics (e.g., giant component?) but can capture local network structure. For example can capture networks with clustering or spatial networks. Assume underlying mixing remains fixed. Infection parameter and denote as the number of connections between S and I where new infections can occur. Evolution of new infections then is but changes over time as well . But triples can change over time as well.. so they introduce a moment closure approximation. Single-step contact tracing introduced as and similarly, for [I] where fraction f of contacts of an index case receive treatment at the same time as the index patient. Seems to have been developed in Keeling (1999) but also used in Eames 2007.

**Percolation-theory methods:** Based on generating functions and only require degree distribution of the network, average transmissibility T of the pathogen. These methods are very general and very tractable. They provide excellent final-state predictions but do not predict the dynamics of an outbreak. Examples are Newman 2002; Meyers et al. 2003, 2005, 2006

**Probability generating function model:** Tracks global epidemiological dynamics in complex random networks (Volz, referenced in Bansal 2007).

**Single-step tracing**: a proportion of the partners of index cases is treated concurrently with the index patient, thus resulting in the recovery of pairs of individuals; this is an appropriate model for certain STD situations when patients and their partners attend a clinic together

**Full citations**

Bansal, Shweta, Bryan T Grenfell, and Lauren Ancel Meyers. 2007. “When Individual Behaviour Matters: Homogeneous and Network Models in Epidemiology.” *Journal of the Royal Society Interface* 4 (16). The Royal SocietyLondon: 879–91. doi:10.1098/rsif.2007.1100.

Danon, Leon, Thomas A House, Jonathan M Read, and Matt J Keeling. 2012. “Social Encounter Networks: Collective Properties and Disease Transmission.” *Journal of the Royal Society Interface* 9 (76). The Royal Society: 2826–33. doi:10.1098/rsif.2012.0357.

de Almeida, Mauricio L, Gabriel A Mendes, G Madras Viswanathan, and Luciano R da Silva. 2013. “Scale-Free Homophilic Network.” *European Physical Journal B* 86 (2). Springer-Verlag. doi:10.1140/epjb/e2012-30802-x.

Dodd, Peter J, Clare Looker, Ian D Plumb, Virginia Bond, Ab Schaap, Kwame Shanaube, Monde Muyoyeta, et al. 2016. “Age- and Sex-Specific Social Contact Patterns and Incidence of Mycobacterium Tuberculosis Infection.” *American Journal of Epidemiology* 183 (2): 156–66. doi:10.1093/aje/kwv160.

EAMES, K T D. 2007. “Contact Tracing Strategies in Heterogeneous Populations.” *Epidemiology and Infection* 135 (3): 443–54. doi:10.1017/S0950268806006923.

Girvan, M, and M E J Newman. 2002. “Community Structure in Social and Biological Networks.” *Proceedings of the National Academy of Sciences* 99 (12). National Academy of Sciences: 7821–26. doi:10.1073/pnas.122653799.

Mossong, Joël, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, et al. 2008. “Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases..” Edited by Steven Riley. *PLoS Med* 5 (3). Public Library of Science: e74. doi:10.1371/journal.pmed.0050074.

Newman, M E J. 2002. “Assortative Mixing in Networks.” *Phys. Rev. Lett.* 89 (20). American Physical Society: 208701. doi:10.1103/PhysRevLett.89.208701.

Newman, MEJ. 2003. “Mixing Patterns in Networks.” *Physical Review E* 67 (2). American Physical Society. doi:10.1103/PhysRevE.67.026126.

Perkins, S E, M F Ferrari, and P J Hudson. 2008. “The Effects of Social Structure and Sex-Biased Transmission on Macroparasite Infection.” *Parasitology* 135 (13): 1561–69. doi:10.1017/S0031182008000449.

Read, J M, W J Edmunds, S Riley, J Lessler, and D A T Cummings. 2012. “Close Encounters of the Infectious Kind: Methods to Measure Social Mixing Behaviour.” *Epidemiology and Infection* 140 (12). Cambridge University Press: 2117–30. doi:10.1017/S0950268812000842.

Sah, Pratha, Lisa O Singh, Aaron Clauset, and Shweta Bansal. 2014. “Exploring Community Structure in Biological Networks with Random Graphs.” *BMC Bioinformatics* 15 (1). BioMed Central. doi:10.1186/1471-2105-15-220.