Livestock market data for modeling disease spread among US cattle

Ian T. Carroll* and Shweta Bansal

Department of Biology
Georgetown University
Washington, DC 20057, USA

Jason E. Lombard

United States Department of Agriculture
Animal and Plant Health Inspection Services
Veterinary Services, Center of Epidemiology and Animal
Health
Fort Collins, CO 80526, USA

July 3, 2015

1 Abstract

Transportation of livestock carries the risk of importing infectious disease into a susceptible population, leading to costly public and private sector expenditures on disease containment and, hopefully, eradication. Individual animal tracing systems implemented outside the US have allowed epidemiologists and veterinarians to model the risks of livestock transportation and prepare

---

Corresponding author: itc2@georgetown.edu

responses designed to protect the livestock industry. Within the US, data on livestock transportation available to researchers is not sufficient for direct forcing of disease models, but network models that assimilate limited data provide a platform for disease models that can inform policy. Here, we report on a novel data stream with such potential: the information publicly reported by US livestock markets on the origin of cattle consigned at live-auctions. By aggregating weekly auction reports from markets in several states, some providing multi-year historical archives, we obtain an ego-centric sample of edges from the dynamic cattle transportation network in the US. We first illustrate, using over-simplified disease models, how such data are relevant to the outcome of a disease outbreak. Subsequently, we demonstrate how the sample might be used to infer shipments to unobserved livestock markets in the US, although we find the assumptions of edge prediction by generalized linear models too restrictive. We conclude that in combination with statistical models allowing greater dependence between edges, the market data create potential for inference of a complete transportation network model, one which includes the capacity of markets to spread or control livestock disease.

2 Introduction

Livestock operations within the United States (US) must be vigilant against transboundary animal diseases, among which foot-and-mouth disease (FMD) is recognized as a critical threat for cattle producers (USDA-APHIS-VS, 2014). The 2001 FMD outbreak in the United Kingdom (UK) cost their agricultural sector 3 billion, and 5% of the nation's 11 million cattle were culled to control the disease (Thompson et al., 2002). A study on FMD risk to California's 5 million beef and dairy cattle predicts economic losses in the tens of billions of dollars, even for an outbreak artificially terminated at California's border (Carpenter et al., 2011). The potential impact of a full-blown epidemic in the US, putting at risk a 90 million strong cattle herd (USDA\NASS, 2015a), compels us to study the possible routes of disease spread from an initially infected animal holding premises (Shields and Mathews, 2003). Mechanistic models that incorporate livestock transportation are needed to help guide prevention and control of FMD-like diseases, which are known to cause massive burdens on livestock industries, require costly public interventions, raise public health concerns and impact food security

2

(Anderson, 2002).

Studies of past livestock epidemics and disease simulations reveal that network models provide a useful abstraction of data on animal shipments between livestock operations (Vernon and Keeling, 2009). Network models typically emphasize heterogeneity in the number of disease transmitting contacts attributed to infectious nodes, a pattern that emerged strongly during the initial spread of FMD in the UK's 2001 epidemic (Shirley and Rushton, 2005) and one usually absent from simulations where transmission depends on distance alone. Network representations of cattle trade exist for several European livestock industries (Bigras-Poulin et al., 2006; Kao et al., 2006; Nöremark et al., 2011; Dutta et al., 2014; Lentz et al., 2011; Natale et al., 2009), where data for model development is generated from animal tracing systems mandated by the European Parliament (Anonymous, 2000). Availability of these data have allowed for several advances in surveillance and control strategies: for example, (1) identification of "sentinel" livestock premises projected to become infected early during an outbreak in Italy (Bajardi et al., 2012), (2) validation of risk reduction from the standstill rules implemented in the UK after 2001 (Green et al., 2006), and (3) evaluation of targeted movement bans that selectively eliminate network edges based on node centrality in transportation networks (Natale et al., 2009).

Cattle transportation is a vital component of the US livestock industry, but point-to-point data on animal movements is extremely limited (Shields and Mathews, 2003). The US Department of Agriculture (USDA) census collects inventory, production and disposition estimates; assuming these are balanced by in-shipments suggests that 22% of the US herd gets transported across state lines each year (USDA\NASS, 2015b). State agencies collect shipment origin and destination locations on certificates of veterinary inspection (CVIs) for cattle entering their state. One cattle transportation network model has been estimated from a sample of year 2009 CVIs aquired from 48 state agencies (Lindström et al., 2013; Buhnerkempe et al., 2013). The epidemiological network model constructed from these data (Buhnerkempe et al., 2014) represents the state of the art for analysis of a nation-wide epidemic in the US, but limitations in the underlying data are substantial. Given that a mean of only 19.2 (SEM 3.1) percent of shipments onto US beef operations travel over 100 miles (USDA, 2009), shipping within states likely occurs at much higher rates than shipments documented on CVIs. While shorter range shipments can be inferred, they cannot be validated against CVI data, nor can annual variation in the network be estimated without re-

peating a major collection effort, and nor can the type of origin or destination facility be determined from CVIs alone. Finally, and of most relevance to the present study, CVIs are not required for shipments to exempt facilities, including certain federally approved livestock markets (Anonymous, 2013).

As in the UK, where livestock markets had very high centrality in the 2001 foot-and-mouth (FMD) epidemic network (Shirley and Rushton, 2005), and other countries with recent outbreaks of non-endemic FMD(McLaws and Ribble, 2007), cattle in the US are commonly sold at live auction markets between stages of beef or dairy production. In the most recent national survey, over half of beef producers sold non-breeding stock at auction markets in the US (60.7% steers, 58.3% cows), with internet auctions and private treaties being the main alternatives (USDA, 2010). Dairy contributes fewer US cattle shipments, but cows removed from dairy operations are also predominantly sent to auction markets or stockyards (Usda, 2007; Buhnerkempe et al., 2013). Localized but explicitly epidemiological studies have also found direct contacts with livestock markets prevalent in Colorado and Kansas (McReynolds et al., 2014) and California (Bates et al., 2001). The economic reality now, and for the foreseeable future, is that cattle owners regularly buy and sell cattle at particular stages of production and rely on live-auction markets to obtain the best price (RTI, 2007a; Bailey et al., 1995).

The epidemiological importance of livestock markets arises from their potentially high degree in both contact and epidemic networks—the same epidemic phenomena airports create as hubs for transmission and spread of human influenza (Colizza et al., 2006). When a livestock operation ships infected animals to market, two processes spread the disease: splitting of the original group of infected animals among multiple buyers, and transmission to susceptible animals passing through the same market (Dawson et al., 2014). Both processes acted to give livestock markets high out-degree in an epidemic network for FMD (Gibbens et al., 2001; Shirley and Rushton, 2005), while a less contagious disease would primarily be affected by splitting up infected animals arriving from one premises. Livestock markets can also have high in-degree within the contact network, or a large number of operations from which cattle are sourced (Robinson and Christley, 2007). High in-degree markets are a natural point of surveillance for disease detection; indeed, markets seeking USDA approval are required to provide veterinary inspection of cattle at auction (Anonymous, 2013).

On the premise that transportation of cattle to and from markets is an important class of potential disease-transmitting contacts between farms and

4

other longer-term animal holdings, we studied the potential for data-driven modeling of a market-based contact network for infectious livestock disease in the US. To this end, we collected data on the locations of origin for individual cattle sold at livestock markets that publicly share this information as a form of advertising. Because the data represent an opportunistic sample of livestock transportation, we first report basic trends in the data alongside some potential sources of bias in the sample. We then demonstrate two methods for inference of a network model of contacts between livestock operations from these data: (1) using the sample to estimate degree distributions of an otherwise random contact network, and (2) fitting coefficients to covariates that may predict the presence or weight of unobserved network edges. The results firmly establish that market-bound cattle shipments are dominated by intra-state movements, and are consistent with the possibility that transportation to markets also drives interstate flows. The daily resolution of this data source allows detection of sub-annual variation in trade volume and network degree distributions, and as a continuously updated data stream creates potential for both inter-annual trend and recent-event detection. We demonstrate high potential for inferring the properties of un-observed edges connected to non-reporting livestock markets, and conclude with a discussion of the critical gaps for building a complete epidemic network model that includes markets acting as hubs for the spread of livestock disease.

3   Methods

3.1   Data Collection

As an integral part of livestock trade, stockyards are distributed across all parts of the US with beef or dairy operations, i.e. throughout the US (Fig. 1)(Carroll and Bansal, 2014). Prices obtained at live auction are rapidly publicized to help consignors and buyers decide when and where to trade cattle. In some cases, professional market reporters attend sales and distribute volume and price information through the USDA Market News Service. Where market reporters are unavailable, or to provide additional information, sale reports might be generated by the market itself and publicized on its own website. A subset of markets list specific lots of cattle sold in their sale reports, including a location of origin and other attributes. These data, sometimes labeled "representative sales" as we refer to them here, indicate

5

cattle were transported from the origin to market on, or very near, the sale date.

We aggregated representative sales from several livestock markets and georeferenced each location of origin to a US county or county-equivalent (hereafter "county"). All websites listed in the directory of the Livestock Marketing Association (lma) were manually searched for representative sales. We wrote software to parse data from the subset of websites regularly publishing representative sales and which permit crawling by the Robots Exclusion Protocol. For each lot given as a representative sale, the software attempts to parse the consignor's location along with cattle type (e.g. steer or heifer), number, average weight and price (either per head or per hundred weight). A single animal per lot was assumed whenever the number of cattle could not be parsed. We tuned the parser for each website until two researchers found no data extraction errors in independent spot checks of representative sales. Websites were checked for new reports weekly between June 1, 2015 and variable start dates from June 2014, and parsers returning data of the wrong type (i.e. string or numeric) were promptly corrected.

Acceptable locations of origin are given as the name of a city or other populated place, with or without a state, and can be ambiguous. We matched each location, substituting common abbreviations for full words as needed to obtain a match, to names of populated places in North America using the GeoNames webservice (Anonymous) to identify the encompassing county. The county closest to the reporting market, as determined by the great-circle distance between county centroids (orn, 2011), is recorded as the true location of origin.

3.2   Comparison with Interstate Shipments

Interstate cattle shipments are sometimes found among representative sales, allowing a comparison to cattle shipping data obtained from state ICVI records. For each state with at least one market providing representative sales, we correlated the number of cattle in representative sales originating in other states with the analogous interstate flows reported by Shields and Mathews (2003). The ICVI data sampled shipments occurring the 2001 calendar year, while the majority of available representative sales occurred within the past year. To capture variation over the course of a full year, we aggregated representative sales dated within the year preceding 2015-03-01before calculating correlations, but the comparison necessarily reflects a

6

decade of changes in the livestock system on top of any differences between market shipments and ICVI shipments.

3.3   Analysis of Sampling Rate

Representative sales are not randomly sampled from all cattle shipped to markets, the sample is in-effect stratified by individual sale reports. Estimation that involves any aggregation across sale reports should take into account the sample weight, but the total number of animals shipped to market is unknown in roughly a third of market reports. We analyzed variation in the sample size of representative sales for the remaining sales by fitting a GLMM to the number of head in representative sales, given the total head of cattle sold ("receipts"). Receipts on a given day were extracted from USDA Market News Service reports (USDA\AMS) or, when available, from the market's own report. Inventory and sales of cattle in the county where the market is located (USDA\NASS, 2015a), receipts, and season of year (with four levels) were fixed effects in the model. Positive, integer predictors were first log transformed and standardized. Each market was allowed a random intercept, but the number of representative sales remained overdispersed relative to the fitted model. To avoid possible parameter bias, we added an observation level random intercept to eliminate overdispersion (Harrison, 2014). We fit binomial family GLMMs with logit link function using the 'lme4' package (Bates et al., 2014), and performed Wald χ^2 tests for significance of fixed effects using the 'car' package (Fox and Weisberg, 2011), in R version 3.1.3 (R Core Team, 2015).

3.4   Network Inference: Random Graph

The definition of nodes and the meaning of edges in contact networks are flexible, facilitating data-driven modeling approaches. Models for disease spread among livestock incorporate transportation data as edges in the network of contacts between susceptible and infected individuals (Kao et al., 2006, e.g.). Representative sale data is compatible with a model having two types of nodes: one representing all farms, ranches and other long-term animal holdings located within a given county, and one representing a single market. In order to study the contribution of these data to the inferred topology of the contact network, and its consequence for disease spread, we ignore edges between counties that arise from fence-line contact, private sales,

7

transportation for grazing, and other mechanisms that might transmit disease directly between counties. Observing no indication of market-to-market transportation in the representative sales, we assumed their absence as well. As a result, the only edges we infer from representative sales are between nodes of different types, yielding a bi-partite contact network.

One approach to building a network topology that is consistent with representative sales is to estimate only the degree distribution for markets from the data, while specifying all remaining network properties parametrically. With k_i the ith of m sampled values representing market degree, empirical estimates for the market degree distribution generating function, $G_{M,0}(x)$, and the generating function for market "excess degree" (e.g. Newman, 2002, eq. 12), $G_{M,1}(x)$, are

$$G_{M,0}(x) \approx \frac{1}{m} \sum_{i=1}^{m} x^{k_i} \tag{1}$$

$$G_{M,1}(x) \approx \frac{\sum_{i=1}^{m} k_i x^{k_i-1}}{\sum_{i=1}^{m} k_i}. \tag{2}$$

We specify properties of the full network by choice of the algorithm for connectivity and the probability distribution for county degree: we assume a bi-partite configuration model for edges and a Poisson distribution on county degree. In a static network with these properties, the expected size of a disease outbreak in terms of the number, or proportion, of counties affected can be calculated exactly (Meyers et al., 2003). With τ_C and τ_M the disease transmission probabilities from counties and markets, respectively, and the mean county degree equal to λ_C, the epidemic threshold is a fixed value of $\phi = \tau_C \tau_M \lambda_C$ determined by the market degree distribution. The threshold occurs at

$$\phi^{-1} = G'_{M,1}(1). \tag{3}$$

For parameterizations below this threshold, the expected number of counties affected by an outbreak is

$$1 + \frac{\phi G'_{M,1}(1)}{1 - \phi G'_{M,1}(1)}. \tag{4}$$

For parameterizations above this threshold, the proportion of counties in the epidemic is

$$1 - e^{\lambda_C(u-1)}, \tag{5}$$

8

where u is the smallest root of

$$u = (1 - \tau_C) + \tau_C G_{M,1} \left((1 - \tau_M) + \tau_M e^{\lambda_C (u-1)} \right). \tag{6}$$

See Appendix 5 for an explanation of these equations.

Two features of representative sales data conflict with applying this 'random graph' approach to network modeling of livestock disease spread. First, the degree of each market potentially varies from sale to sale, admitting possible temporal variation in the observed degree distribution. We examine the extent of variation in degree by visualizing temporal variation and calculating seasonal estimates of epidemic size. Second, the representative sales may not include all the counties of origin, potentially biasing the observed distribution toward smaller degrees. Using the iNEXT R-package developed for analysis of species accumulation curves (Chao et al., 2014; Hsieh et al., 2014), we calculate complete degree estimates for each sale using extrapolation of the county accumulation curve as a function of the number of cattle in representative sales.

3.5 Network Inference: Edge Prediction

Modeling disease spread on a random graph assumes the absence of clustering within the contact network, ignoring any propensity of livestock operations within different counties to trade cattle at the same two markets. A model that represents additional structural attributes of the transportation network, including clustering, may yield different predictions for the spread of disease, but direct estimation of these attributes requires particular sampling methods (Morris, 2004). For example, a random sample of nodes is not appropriate for estimating clustering coefficients; a first-wave link tracing approach (*sensu* Handcock and Gile, 2010) is needed to avoid underestimating the number of triangles touching each focal node. Edge prediction is any process for inferring the properties of unobserved edges, and allows construction of network models that are more flexible than random graphs.

We applied a regression approach to the problem of edge prediction, using observed edges to estimate whether county and market covariates predict their connectivity. The response variable we examined is the number of cattle shipped from a given county to a market, which we assume to arise from a multinomial distribution given the total number of representative sales observed at that market. This GLM includes the fixed effects of the previous

9

GLMM, with the addition of GCD distance between the centroids of each origin county and the county of each market, the number of registered and bonded market agencies giving an address in each origin county (Carroll and Bansal, 2014), and the product of inventory and sales. Distance and square-root transformed distance were included, and as for all numeric predictors, each was normalized to unit variance with zero mean. In practice, the multinomial model is converted to an equivalent (with respect to the likelihood of the multinomial probabilities) Poisson model with log-link function, requiring an additional fixed effect of market. We evaluate the utility of the model by leave-one-out validation: we fit the model with one market removed, then predict the county of origin for all cattle in representative sales for that market.

4 Results

4.1 Characteristics of Representative Sales

Cattle transported to 48 markets located in 46 counties in 15 states are represented in this analysis (Fig. 1), and data on each movement are freely available online (5). The average number of cattle shipments reported as representative of a given sale increases to a peak of 1000-1500 head in late fall and decreases to a few hundred during summer months (Fig. 2). This seasonal trend persists between the period for which representative sales come from a handful of markets with accessible archives and the period since mid-2014, when we began capturing representative sales posted weekly.

The majority of representative sales are for cattle shipped to a livestock market from within the same state (Fig. 4A&B). Markets selling the majority of out-of-state cattle appear to be clustered in Oklahoma and South Dakota, where it is not uncommon for less than half of representative sale cattle to originate in-state (Fig. 4A). Aggregating across all markets, the proportion of total sales that originate in-state showns no trend in deviations from an average of 0.84 (SD 0.07) (Fig. 4B). Among the states we could compare to a summary of inter-state transportation during 2001, the proportion is at least 70% of representative sales 1. The origin of remaining shipments showed fair to strong correlation with the health-certificate derived data. One of the strongest correlations, 0.87 for South Dakota, arose in the state with the largest number of cross-border representative sales that

10

315 we observed. Shipments into Texas are insufficient for a meaningful comparison, but other low correlations (e.g. 0.44 for Colorado) are unlikely to be sample-size artifacts. Variation in the strength of correlation across sample sizes suggests the presence of real variation in the kind of inter-state shipments sampled by different data streams. The difference could be due to the

320 relative proportions of shipments that are market bound versus non-market bound as well as heterogeneity in state requirements for health certificates.

Uncertainty about the total number of animals shipped to market for a given sale is amplified by unknown sources at the market level and to a lesser extent for each individual report. For roughly one third of market reports,

325 the total receipts is not available for use in weighting the sale's affect on estimates aggregating across sales or markets. In other reports, the number of representative sales affects the proportion of receipts observed(FIXME), with a positive regression coefficient (FIXME). Season is also a significant factor ($\chi_3^2 = 19.3$, $p < 0.001$), with the winter contrast taking the largest

330 coefficient (95% CI (9.2×10^{-2}, 0.27)) by an order of magnitude. Neither inventory nor sales is significant in the GLMM, leaving the random effect of market (SD = 1.6) to explain the remainder of the 61% of variance between market averages of observed proportion. However, the representative proportion remains overdispersed with respect to the model fitted without

335 observation level random intercepts. In other words, variation in the binomial sampling probability predicted by the fitted model for each market underestimates actual variation observed in the representative proportion (Fig. 3). The breadth of the observation level random intercepts (SD = 0.65) reflects unexplained variability in the proportion of cattle reported in the represen-

340 tative sales, but the random intercept for each market remains the greater source of uncertainty.

4.2 Epidemic Size on Random Graphs

At a given sale, livestock markets report receiving cattle from 11.3 (SD 1.5) counties on average, with weekly average market degree showing seasonal

345 variation peaking in late autumn (Fig. 4). While the timing of peak degree closely corresponds to the time of year when the number of representative sales is also greatest, the troughs in degree are flatter, broader and less pronounced than the summertime lows in representative sales. The majority of seasonal variation in the distribution on market degree exists between a

350 winter season (the 42nd through 10th weeks of the year), and the remaining

11

seasons, which have barely distinct empirical cumulative distribution functions (ECDFs; Fig. 5A). For a disease spreading on a bi-partite random graph with these market degree distributions, seasonal variation effects the location of the epidemic threshold with respect to the unknown parameters.
355 Roughly 20% lower values of ϕ, the product of both transmisabilities and mean county degree, prompt an epidemic for the peak time of year relative to off-peak (Fig. 5C). Above the epidemic threshold, the difference between seasons becomes negligible as transmisability increases; it is overwhelmed by the overall high degree of livestock markets. Even with the average excess
360 degree of counties equal to one, nearly two-thirds of counties are affected in the extreme case that every contact between susceptible and infective cattle leads to successful disease transmission (Fig. 5D).

The number of counties included in representative sales is, on average, FIXME of the estimated number of counties extrapolated from county ac-
365 cumulation curves. In over half of sale reports, the estimate is less than 1% greater than the observed number of counties; and among the rest the average increase in degree is just 10% (Fig. 5B). The rarity of singleton counties (i.e. counties with only one individual in representative sales) and the sufficiently high sampling rates (Fig. 3) are responsible for the completeness
370 of the sample for counties of origin. Using the extrapolated values in calculating disease spread on a random graph has the same qualitative effect as the shift from off-peak to peak season market degree distributions. Quantitatively however, the difference between observed and extrapolated market degree distributions has less impact on the estimates of disease spread (Fig.
375 5C&D).

4.3   Edge Predictability

The single most important predictor of counties of origin for cattle shipped to market ought to be distance, and the regression model returns a parameter value for distance within 95% CI (-5.75, -5.70) and for the square root of
380 distance within 95% CI (-1.31, -1.33). However, while these and all other predictors in the model are judged significant, the deviance goodness-of-fit test rejects the assumption that samples are drawn from a Poisson family distribution (χ^2 value numerically zero). Parameter uncertainty cannot be reliably determined for this model.
385 Visualization of the model validation result indicates that a key problem with the regression approach is the effective overestimation of market degree.

12

The proportion of receipts, summed across all sales for each market, that originate in different counties gives an impression of the size and variability of livestock marketsheds (Fig. 6A). Due to aggregation across sales, market degree tends to be higher here than in the weekly average (Fig. 4), but it's not so high that every county within reach of a market is connected. In contrast, the analogous marketshed predicted with the regression approach shows even greater degree; rarely is a sufficiently close county not connected.

5   Discussion

The epidemiological contact network is a fundamental component of models for the spread of diseases, and sale reports publicized by livestock auction markets contribute data to drive inference of these networks for the US livestock system. For this study, we initiated an ongoing process for archiving representative sales as an opportunistic sample of cattle transported from counties with beef or dairy operations to livestock markets distributed across the central US. The study complements previous efforts to summarize transportation of cattle within the US using data derived from veterinary inspection certificates (Shields and Mathews, 2003; Buhnerkempe et al., 2013), but extends our ability to model within-state shipping patterns. We demonstrate how inference of a bi-partite contact network, between nodes representing either cattle holding operations aggregated within a US county or a particular auction market, can allow for new models for the spread of economically disruptive livestock diseases.

Representative sales extracted from livestock market reports provide a reliable sample of cattle shipments and the corresponding potential for disease transmitting contacts. Seasonal variation in the volume of representative sales is consistent with beef cattle production systems, where calves are produced in spring and weaned cattle or yearlings sold to pasturing or feedlot operations in the fall and subsequent spring (RTI, 2007b). The proportion of receipts at a given sale whose origin can be identified is anywhere from a small fraction to around three quarters, and understanding this variability is important for scaling up assessments of transportation networks. Although covariates from agricultural census and season are significant predictors, the dominant source of variability is between markets and can be quantified for future modeling efforts despite having no identified deterministic source. While interstate shipments from representative sales correlate fairly well with

13

health certificate derived data, the former provide data on shipments within a state that were previously unavailable and dominate market directed shipments at typically more than 80% on any week.

⁴²⁵ The market derived data allows estimates for contact networks ranging in complexity from random graphs, which have many analytically tractable properties, to networks with non-trivial clustering, modularity, assortativity and other non-random features. A collection of edge data resulting from sampling random nodes, without tracing any discovered edges, is an ego-⁴³⁰ centric sample that allows straightforward estimation of node, but not edge, attributes (Morris, 2004). From this sample, we find market degree distributions that fit a negative binomial distribution with variance roughly twice the mean, which has more dispersion than a Poisson distribution but less than a power law. We also find seasonal shifts in the degree distribution that lower ⁴³⁵ the epidemic threshold of a random graph during the peak cattle trading season. However, the overlapping marketsheds apparent in this sample suggest high potential for network clustering, which can strongly influence disease spread but are not so easily assessed from an ego-centric sample. Estimating this kind of structural attribute requires inference about unobserved edges, ⁴⁴⁰ and a first analysis shows potential for linear effects of county and market attributes in exponential family likelihoods for edge weight. Extensions to this likelihood that include multiple response variables, particularly market degree and squares in the bi-partite network, may achieve a reliable fit to the representative sales that provides a data-driven, non-random graph for ⁴⁴⁵ livestock disease simulations (Handcock and Gile, 2010).

Despite the increased availability of data on livestock transportation in the US that our study provides, disease models here lag behind the relatively data rich European livestock systems. Research in these systems on the optimal spatial and temporal resolution at which to model contact networks is ⁴⁵⁰ critical for efficient use of limited information available in the US and targeted development of new data streams. In a spatially embedded contact network, each node represents a geographically constrained subpopulation within an interacting metapopulation (Schumm et al., 2014). The constraint should reflect where mixing of susceptible and infected individuals occurs in pro-⁴⁵⁵ portion to their frequencies, but no theory exists for transferring constraints developed in one region (e.g. the UK) to any other (e.g. Pennsylvania). The abstraction of temporally discrete livestock shipments into static edges, representing the potential for disease transmission over time, is better understood (Vernon and Keeling, 2009; Valdano et al., 2014). The additional

14

460 challenge for a contact network distinguishing livestock markets from longer term animal holdings is synchronicity of shipments of cattle between two counties and the same market, although extreme cases of complete within market segregation versus mixing (Dawson et al., 2014) should bracket the range of disease outcomes.

465 The eventual purpose of collecting data on livestock transportation is to improve surveillance for disease outbreaks and prevention of control of epidemics. The sources of nation-wide data on US livestock movements contributing to these goals have previously included health certificate records accompanying interstate movements (Shields and Mathews, 2003; Buhn-
470 erkempe et al., 2013) and owner/operator surveys on animal health and management practices for representative animal holdings (Usda-Aphis-Vs, 2009). Future research should aim to combine these source with representative sales data to jointly infer contact networks, because each data source addresses network attributes absent from the others. The primary deficiency
475 of representative sales data is the absence of out-going shipment information, or the destination of cattle purchased at auction. Surveys of livestock operations presently include information on the in-shipment degree, source type and distance, which provides evidence for missing county-node edges. Representative sales also only include market-directed shipments, while health
480 certificate data provides information on network edges that may not have a livestock market at either end. Especially in combination, which we recognize to be a difficult task both conceptually and statistically, inference from multiple data sources will dramatically improve awareness of the network of potentially disease spreading contacts between livestock.

485 # References

???? Livestock Marketing Association: Find an Auction/Dealer.

2011. County-to-County Distance Matrix.

Anderson, I. 2002. Foot and Mouth Disease 2001: Lessons to be Learned Inquiry Report .

490 Anonymous. ???? GeoNames Web Service Documentation.

———. 2000. Regulation (EC) No 1760/2000 of the European Parliament and of the Council of 17 July 2000 establishing a system for the identifica-

tion and registration of bovine animals and regarding the labelling of beef and beef products and repealing Council Regulati.

495 ———. 2013. Traceability for Livestock Moving Interstate. Federal Register 78:2040–2075.

Bailey, D., B. W. Brorsen, and M. R. Thomsen. 1995. Identifying Buyer Market Areas and the Impact of Buyer Concentration in Feeder Cattle Markets Using Mapping and Spatial Statistics. American Journal of Agricultural 500 Economics 77:309–318.

Bajardi, P., A. Barrat, L. Savini, and V. Colizza. 2012. Optimizing surveillance for livestock disease spreading through animal movements. Journal of the Royal Society, Interface / the Royal Society 9:2814–25.

Bates, D., M. Maechler, B. Bolker, and S. Walker. 2014. lme4: Linear mixed- 505 effects models using Eigen and S4. R package version 1.1-7.

Bates, T. W., M. Thurmond, and T. Carpenter. 2001. Direct and indirect contact rates among beef, dairy, goat, sheep, and swine herds in three California counties, with reference to control of po- tential foot-and-mouth disease transmission. Am J Vet Res 62:1121–1129.

510 Bigras-Poulin, M., R. a. Thompson, M. Chriel, S. Mortensen, and M. Greiner. 2006. Network analysis of Danish cattle industry trade patterns as an evaluation of risk potential for disease spread. Preventive veterinary medicine 76:11–39.

Buhnerkempe, M. G., D. a. Grear, K. Portacci, R. S. Miller, J. E. Lombard, 515 and C. T. Webb. 2013. A national-scale picture of U.S. cattle movements obtained from Interstate Certificate of Veterinary Inspection data. Preventive veterinary medicine 112:318–29.

Buhnerkempe, M. G., M. J. Tildesley, T. Lindström, D. a. Grear, K. Portacci, R. S. Miller, J. E. Lombard, M. Werkman, M. J. Keeling, U. Wennergren, 520 and C. T. Webb. 2014. The impact of movements and animal density on continental scale cattle disease outbreaks in the United States. PloS one 9:e91724.

Carpenter, T. E., J. M. O'Brien, A. D. Hagerman, and B. a. McCarl. 2011. Epidemic and economic impacts of delayed detection of foot-and-mouth

16

disease: a case study of a simulated outbreak in California. Journal of veterinary diagnostic investigation : official publication of the American Association of Veterinary Laboratory Diagnosticians, Inc 23:26–33.

Carroll, I. T., and S. Bansal. 2014. Cattle Auction Markets in the USA: a De-duplicated List Compiled from Trade Association and Federal Agency Directories. V1. Harvard Dataverse .

Chao, A., N. J. Gotelli, T. C. Hsieh, E. L. Sander, K. H. Ma, R. K. Colwell, and A. M. Ellison. 2014. Rarefaction and extrapolation with Hill numbers: A framework for sampling and estimation in species diversity studies. Ecological Monographs 84:45–67.

Colizza, V., A. Barrat, M. Barthélemy, and A. Vespignani. 2006. The role of the airline transportation network in the prediction and predictability of global epidemics. Proceedings of the National Academy of Sciences of the United States of America 103:2015–2020.

Dawson, P. M., M. Werkman, E. Brooks-pollock, and M. J. Tildesley. 2014. Epidemic predictions in an imperfect world : modelling disease spread with partial data .

Dutta, B. L., P. Ezanno, and E. Vergu. 2014. Characteristics of the spatiotemporal network of cattle movements in France over a 5-year period. Preventive Veterinary Medicine 117:79–94.

Fox, J., and S. Weisberg. 2011. An {R} Companion to Applied Regression. 2nd ed. Sage, Thousand Oaks {CA}.

Gibbens, J. C., C. E. Sharpe, J. W. Wilesmith, L. M. Mansley, E. Michalopoulou, J. B. M. Ryan, and M. Hudson. 2001. Descriptive epidemiology of the 2001 foot-and-mouth disease epidemic in Great Britain: the first five months. The Veterinary Record 149:729–743.

Green, D. M., I. Z. Kiss, and R. R. Kao. 2006. Modelling the initial spread of foot-and-mouth disease through animal movements. Proceedings. Biological sciences / The Royal Society 273:2729–35.

Handcock, M. S., and K. J. Gile. 2010. Modeling social networks from sampled data 4:5–25.

17

Harrison, X. A. 2014. Using observation-level random effects to model overdispersion in count data in ecology and evolution. PeerJ 2:e616.

Hsieh, T. C., K. H. Ma, and A. Chao. 2014. iNEXT: iNterpolation and EXTrapolation for species diversity. R package version 2.0.

Kao, R. R., L. Danon, D. M. Green, and I. Z. Kiss. 2006. Demographic structure and pathogen dynamics on the network of livestock movements in Great Britain. Proceedings. Biological sciences / The Royal Society 273:1999–2007.

Lentz, H. H. K., M. Konschake, K. Teske, M. Kasper, B. Rother, R. Carmanns, B. Petersen, F. J. Conraths, and T. Selhorst. 2011. Trade communities and their spatial patterns in the German pork production network. Preventive Veterinary Medicine 98:176–181.

Lindström, T., D. a. Grear, M. G. Buhnerkempe, C. T. Webb, R. S. Miller, K. Portacci, and U. Wennergren. 2013. A bayesian approach for modeling cattle movements in the United States: scaling up a partially observed network. PloS one 8:e53432.

McLaws, M., and C. Ribble. 2007. Description of recent foot and mouth disease outbreaks in nonendemic areas: exploring the relationship between early detection and epidemic size. Canadian Veterinary Journal 48:1051–1062.

McReynolds, S. W., M. W. Sanderson, A. Reeves, M. Sinclair, A. E. Hill, and M. D. Salman. 2014. Direct and indirect contact rates among livestock operations in Colorado and Kansas. Journal of the American Veterinary Medical Association 244:1066–74.

Meyers, L. A., M. E. J. Newman, M. Martin, and S. Schrag. 2003. Applying network theory to epidemics: Control measures for Mycoplasma pneumoniae outbreaks. Emerging Infectious Diseases 9:204–210.

Morris, M. 2004. Network epidemiology: A handbook for survey design and data collection. Oxford University Press.

Natale, F., A. Giovannini, L. Savini, D. Palma, L. Possenti, G. Fiore, and P. Calistri. 2009. Network analysis of Italian cattle trade patterns and evaluation of risks for potential disease spread. Preventive Veterinary Medicine 92:341–350.

18

Newman, M. 2002. Spread of epidemic disease on networks. Physical Review E 66:016128.

Nöremark, M., N. Hå kansson, S. S. Lewerin, A. Lindberg, and A. Jonsson. 2011. Network analysis of cattle and pig movements in Sweden: measures relevant for disease control and risk based surveillance. Preventive veterinary medicine 99:78–90.

R Core Team. 2015. R: A Language and Environment for Statistical Computing.

Robinson, S. E., and R. M. Christley. 2007. Exploring the role of auction markets in cattle movements within Great Britain. Preventive veterinary medicine 81:21–37.

RTI. 2007a. GIPSA Livestock and Meat Marketing Study. Report Volume 1 1.

———. 2007b. GIPSA Livestock and Meat Marketing Study. Report Volume 3 3.

Schumm, P., C. Scoglio, Q. Zhang, and D. Balcan. 2014. Global epidemic invasion thresholds in directed cattle subpopulation networks having source, sink, and transit nodes. Journal of theoretical biology 367C:203–221.

Shields, D. A., and K. H. Mathews. 2003. Interstate Livestock Movements pages 1–22.

Shirley, M. D., and S. P. Rushton. 2005. Where diseases and networks collide: lessons to be learnt from a study of the 2001 foot-and-mouth disease epidemic. Epidemiology and infection 133:1023–32.

Thompson, D., P. Muriel, D. Russell, P. Osborne, a. Bromley, M. Rowland, S. Creigh-Tyte, and C. Brown. 2002. Economic costs of the foot and mouth disease outbreak in the United Kingdom in 2001. Revue scientifique et technique (International Office of Epizootics) 21:675–687.

Usda. 2007. Dairy 2007 Part II pages 1991–2007.

USDA. 2009. Beef 2007-08: Part II. NAHMS .

———. 2010. Beef 200708: Part IV. NAHMS .

19

Usda-Aphis-Vs. 2009. Analysis of Bovine Tuberculosis Surveillance in Accredited Free States. Tuberculosis .

USDA-APHIS-VS. 2014. Foot-and-mouth disease response plan .

USDA\AMS. ???? Livestock, Poultry, and Grain Market News.

USDA\NASS. 2015a. 2012 Census of Agriculture.

———. 2015b. Meat Animals Production , Disposition , and Income 2014 Summary .

Valdano, E., L. Ferreri, C. Poletto, and V. Colizza. 2014. Analytical computation of the epidemic threshold on temporal networks. arXiv preprint page 19.

Vernon, M. C., and M. J. Keeling. 2009. Representing the UK's cattle herd as static and dynamic networks. Proceedings. Biological sciences / The Royal Society 276:469–476.
Epidemic Size on Bi-Partite Random Graphs
Description of Released Data

20

Table 1: Correlations with ERS on between state movements, also show within-state proportion and other state-numbers (number of markets, reports, movements, etc)

Dest.	p	head	n	r
WY	0.88	5149	28	0.71
SD	0.72	87369	39	0.78
CO	0.85	10523	23	0.44
MO	0.82	5581	17	0.34
NV	0.96	148	28	0.66
ND	0.98	3146	3	0.85
MT	0.92	12199	16	0.85
KS	0.89	1610	38	0.68
MN	0.95	484	37	0.44
TX	1.00	12	33	0.29
OK	0.70	1777	18	0.73
NM	0.85	2887	17	0.99
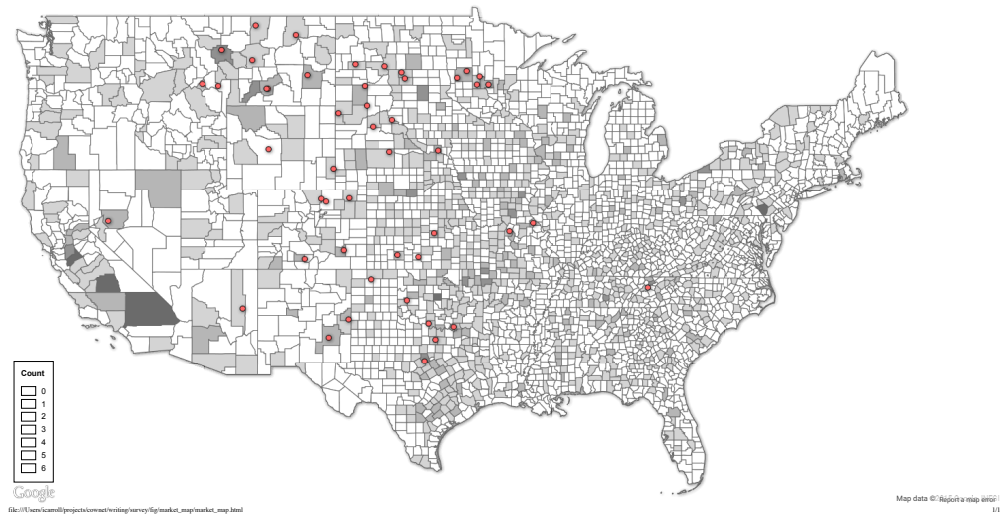
Figure 1: Cattle auction markets in the US, pinning those sampled in this study. Using GIPSA's list until further validated.
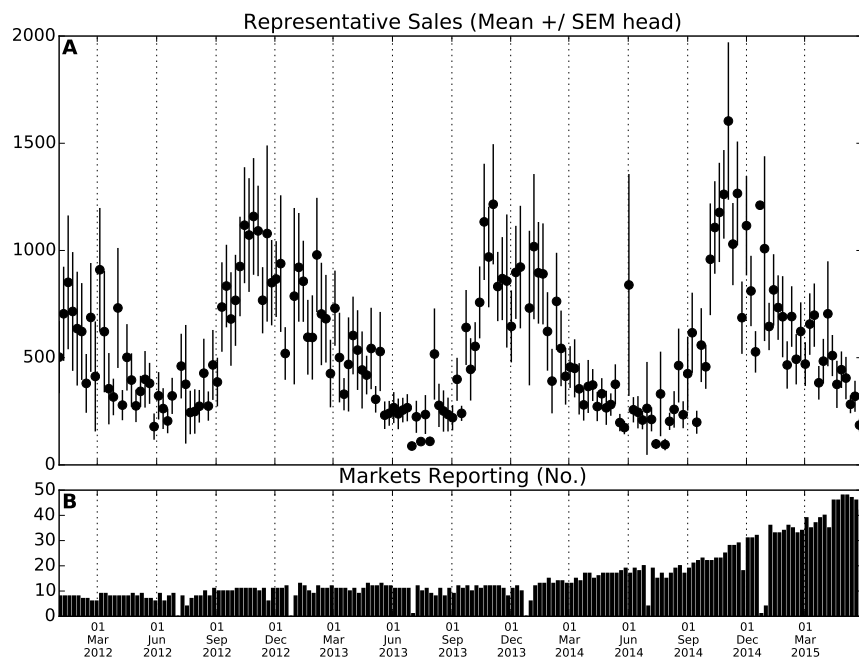
Figure 2: Time series of the average number of representative sales observed, and the number of markets reporting each week.
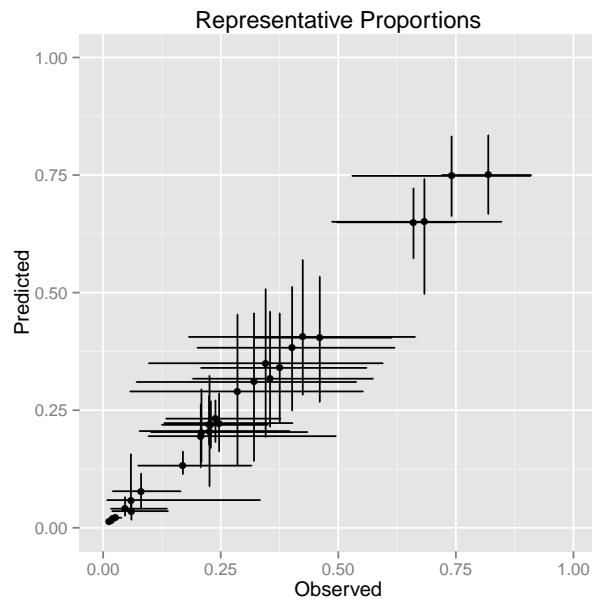
Figure 3: Observed versus predicted representative proportions, grouped by market. Horizontal bars show SD of observed ratios of head in representative sales to receipts. Vertical bars show predicted binomial probabilities, leaving out the observation level random intercept but including the market level random intercept.
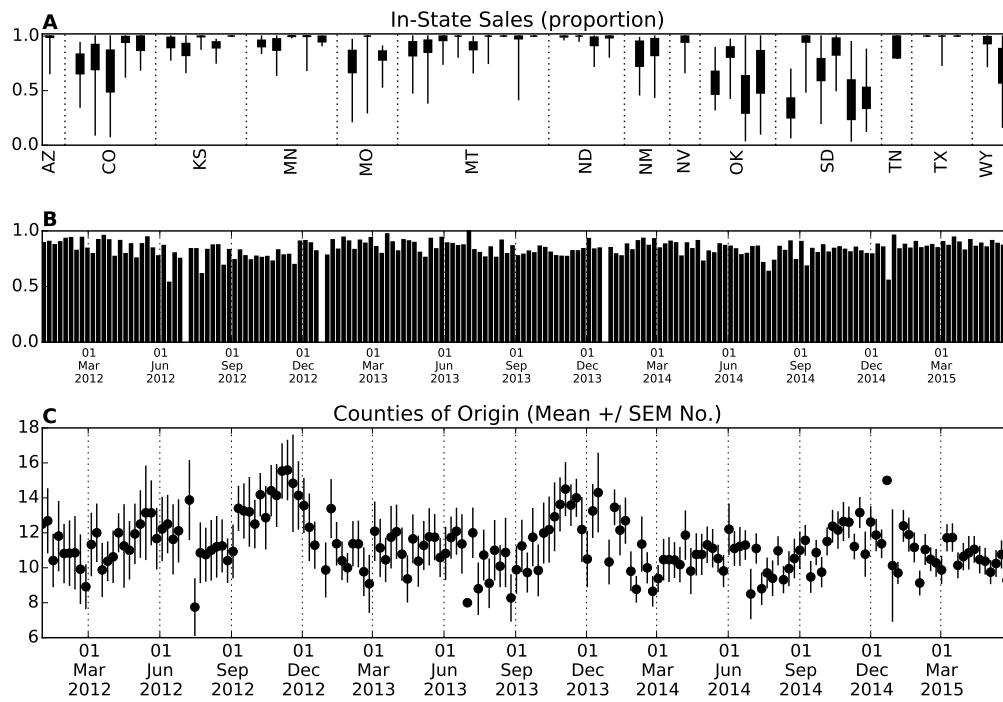
Figure 4: Time series of the average number of counties of origin across markets and the proportion of total representative sales originating in-state.
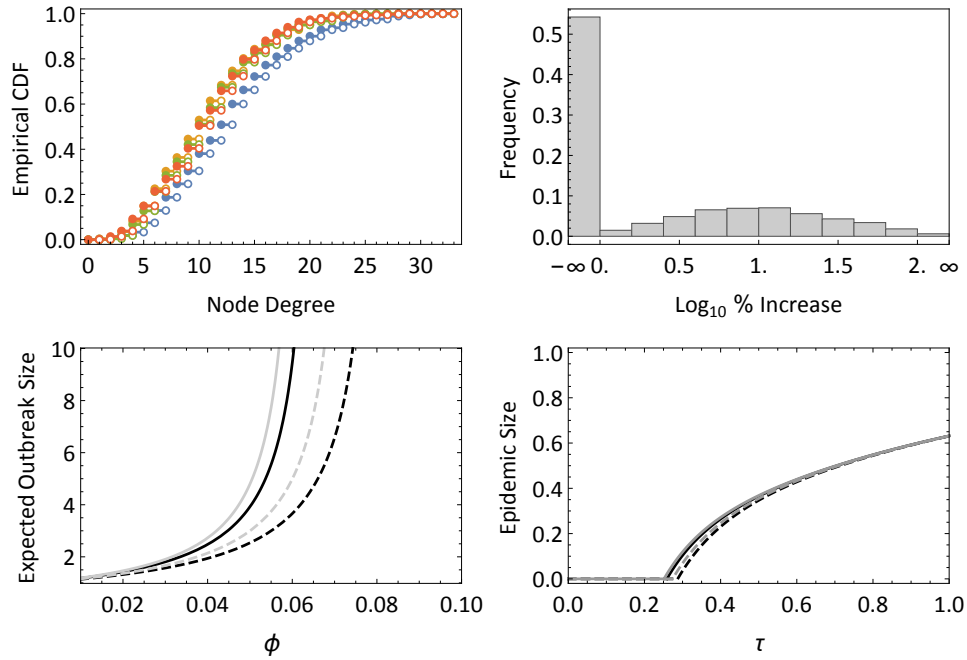
Figure 5: A) Raw degree distribution for sales occurring with four season (blue: nominally winter). B) Percent increase in extrapolated degree relative to sampled degree, with extrapolation up to the total receipts. C) Expected outbreak size for the winter "peak" season (solid) and remaining "off-peak" seasons (dashed), including the observed (black) and extrapolated degrees (gray). ϕ is a product of τ_M, τ_C, and λ_C. D) Proportion of counties affected in the four cases shown in panel C but above the epidemic threshold, assuming $\tau_M = \tau_C = \tau$ and $\lambda_C = 1$.
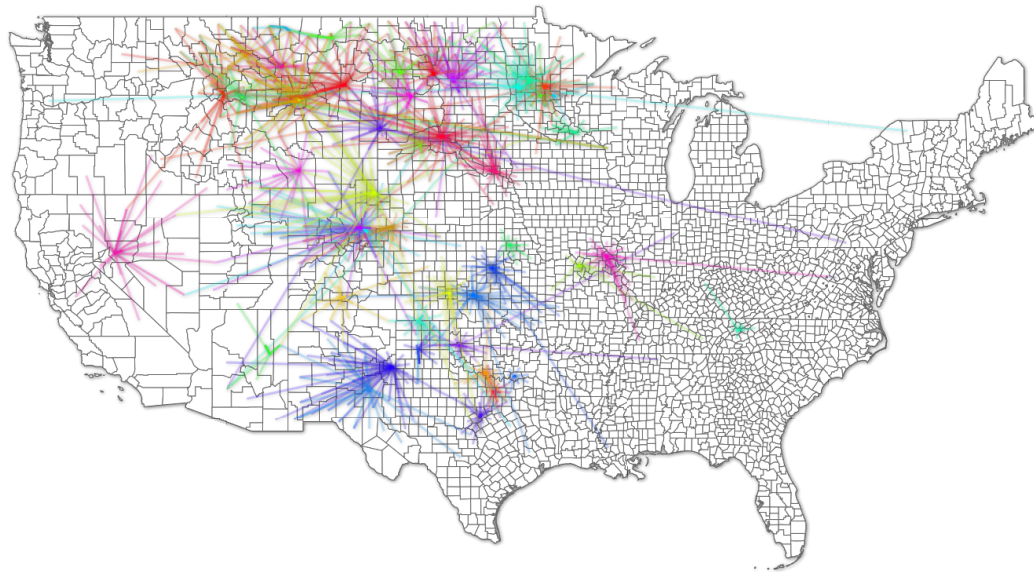
(a)
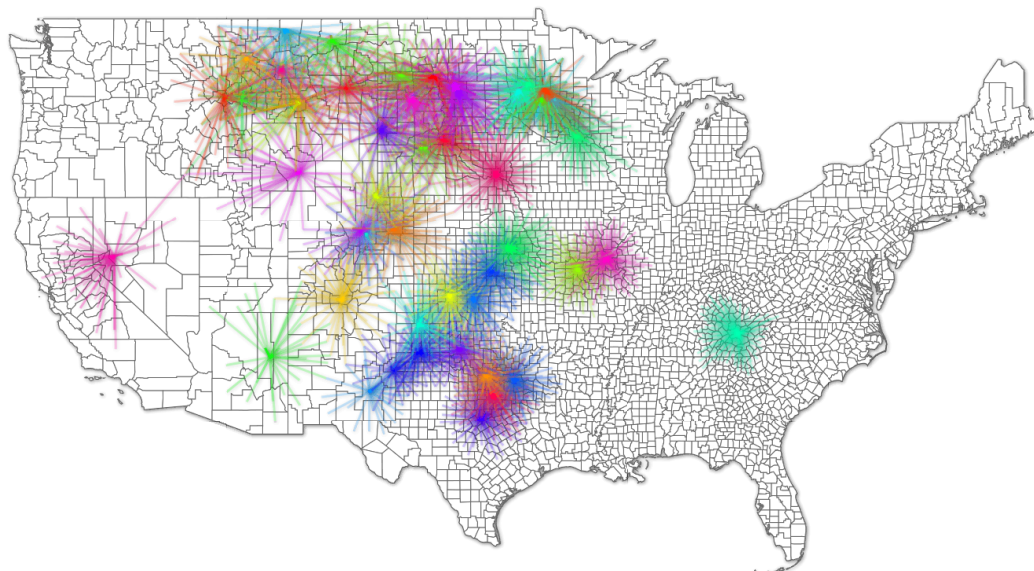
(b)

Figure 6: The proportion of cattle listed in each market's representative sales by county of origin (a), and the multinomial probability of origin by county predicted for each market by the GLM, fitted while excluding the focal market (b). Color corresponds to an individual market, and is linearly scaled between a maximum proportion or probability of 0.8 (transparency at 0%) and a minimum of 0.004 (transparency at 80%)

27

.