Realistic Commodity Flow Networks to Assess Vulnerability of Food Systems

Abhijin Adiga[1(✉)], Nicholas Palmer[1], Sanchit Sinha[1,2], Penina Waghalter[3], Aniruddha Dave[1], Daniel Perez Lazarte[1], Thierry Brévault[4,5], Andrea Apolloni[6,7], Henning Mortveit[1,8], Young Yun Baek[1], and Madhav Marathe[1,2]

[1] Biocomplexity Institute and Initiative, University of Virginia, Charlottesville, USA
abhijin@virginia.edu
[2] Computer Science, University of Virginia, Charlottesville, USA
[3] Yeshiva University, New York City, USA
[4] BIOPASS, CIRAD-IRD-ISRA-UCAD, Dakar, Senegal
[5] CIRAD, UPR AIDA, 34398 Montpellier, France
[6] Université de Montpellier, CIRAD, Montpellier, France
[7] CIRAD, UMR ASTRE, 34398 Montpellier, France
[8] Department of Engineering Systems and Environment, University of Virginia, Charlottesville, USA

Abstract. As the complexity of our food systems increases, they also become susceptible to unanticipated natural and human-initiated events. Commodity trade networks are a critical component of our food systems in ensuring food availability. We develop a generic data-driven framework to construct realistic agricultural commodity trade networks. Our work is motivated by the need to study food flows in the context of biological invasions. These networks are derived by fusing gridded, administrative-level, and survey datasets on production, trade, and consumption. Further, they are periodic temporal networks reflecting seasonal variations in production and trade of the crop. We apply this approach to create networks of tomato flow for two regions – Senegal and Nepal. Using statistical methods and network analysis, we gain insights into spatiotemporal dynamics of production and trade. Our results suggest that agricultural systems are increasingly vulnerable to attacks through trade of commodities due to their vicinity to regions of high demand and seasonal variations in production and flows.

1 Introduction

1.1 Background and Motivation

With rapid population growth, shrinking farm acreage and intensive agriculture, society has come to critically depend on long distance flows of agricultural commodities [25]. This phenomenon has led to availability of a variety of commodities round the year. However, it has also made our food systems increasingly susceptible to threats such as invasive species [11], food contamination [14],

R. M. Benito et al. (Eds.): COMPLEX NETWORKS 2021, SCI 1015, pp. 168–179, 2022.
https://doi.org/10.1007/978-3-030-93409-5_15

extreme weather events [22], and even pandemics like COVID-19 [23]. For example, trade networks act as conduits, enabling the rapid dispersal of pests and pathogens through crops, livestock, packaging, propagating material, etc. In the US alone, the annual economic impact of biological invasions is estimated to be over $120B [20]. Therefore, modeling food systems in all their complexity and understanding their vulnerabilities is critical to ensure food security, biodiversity, health and economic stability.

As the complexity of the systems that are part of society and everyday life continues to grow, the need for more models with increasing levels of resolution and fidelity is continuously growing. Network models, be they simple, hierarchical, and/or multi-scale, have become ubiquitous all throughout science and applications in order to keep up with the demands for analysis, discovery, and support for policy formation. As is the case with many built infrastructures, food flows naturally yield to multi-scale network representations [6,15,19,21]. Depending on the problem being studied, nodes of the network represent locations such as operations, markets, cities, counties, states, or countries, connected by transportation infrastructure, and edges representing flow (commodity specific or aggregated). This work focuses on the *construction and analysis of realistic representations of production, trade, and consumption of agricultural crops that can be applied to epidemiological processes* such as invasive species spread, food poisoning, and biological warfare. Epidemiological models are being increasingly applied in the context of invasive species spread [4,6,15]. Such models can inform policy makers on a variety of aspects such as forecasting, causality, invasion source detection, surveillance, interventions, and economic impact.

1.2 Challenges

Inferring or estimating commodity flow networks is a major challenge as there is hardly any data available on commodity-specific flows. Even if available, the spatial and temporal resolutions of such datasets are not adequate. For example, datasets on inter-country commodity-specific trade are available [8] at yearly resolution. The same holds for production, which is typically available at the state level for a country. Besides, the availability of data differs by country or study region, thus posing a hurdle to generalizing the construction framework. To cope with such challenges, simple models for production and commodity flow have been used. For example, production systems have been mostly weather driven, and host crop production can be modeled using simple regression models with environmental variables as input parameters to sophisticated mechanistic models. But increasingly, farmers are relying on protected cultivation methods, and are thus able to extend the cultivation period to offseasons. Distance between production and consumption area need not be a driver of flows as trade depends on other factors such as crop type, environment, trade & transport infrastructure and pricing. Hence, traditional spatial interaction models might not be able to characterize commodity flows. It is important to incorporate knowledge of growing periods, whole-sale market locations, fine-resolution estimates of crop

production, imports, and exports. Fusion of multi-type datasets, misaligned in space and time and validation are major challenges.

1.3 Contributions

Framework. In this work, we develop a general framework to construct high-resolution temporal networks that capture the production, trade, and consumption of agricultural crops. The framework is developed in the context of invasive species spread through multiple pathways, but is generic enough to be applied to other problems such as mentioned above. We use a number of datasets (at grid-level, administrative unit-level, and qualitative) including crop production, growing seasons, market locations, trade, imports and exports, spatial data capturing human activities, and human population all fused together to obtain multi-scale networks with node and edge attributes.

Application. The number of biological invasions at global scale is steadily increasing. The spread of the South American tomato leafminer [2] over the last 15 years is exemplar of such intercontinental biological invasion events. The pest has been responsible for devastating tomato crops globally. To study its spread, we apply our network generation framework to develop temporal attributed networks representing tomato production and trade for two different regions: Senegal (SN), representative of the spread in West Africa and Nepal (NP), representative of the spread in South and Southeast Asia. The pest invaded Senegal and Nepal in the last decade, and the has been preparing for its impending invasion. The dynamics of tomato production widely differ in these three regions due to climate and trade infrastructure making for an interesting comparison. Secondly, the datasets available for each region vary leading to different construction approaches and assumptions.

Analysis and Summary of Results. We rigorously analyze the resulting networks using statistical methods and structural analysis. Using decision trees, we identify the factors that drive trade in SN network and NP network. We assess the potential for invasion of high-production areas in a number of ways. We investigate the relationship between production areas and characteristics of nearby localities. Analysis of trade flows in conjunction with production shows that production at source is a primary driver of flow. In many cases, areas with high amounts of imports not only have high population, but also have reasonably high production. Such areas are susceptible to invasions.

1.4 Related Works

Datasets on trade of agricultural commodities are seldom available. The United Nations maintains country-to-country trade data for several commodities: FAO-STAT [8] and ComTrade [5]. Ercsey-Ravasz et al. [6] construct an international food trade network using ComTrade by aggregating product codes corresponding to food and assess the vulnerabilities of the network to attacks. Nath et al. [18] and Suweis et al. [24] use production and trade data from FAO and demographics

information to assess the resilience and the reactivity of the coupled population food system.

Multi-pathway models that include long-distance trade have been applied in the invasive species literature [4,15,19,26]. Due to lack of commodity flow data, spatial interaction models such as the gravity model are applied in many cases [4,15]. Nopsa et al. [19] analyze rail networks of grain transport in the context of mycotoxin spread. This work is motivated by the challenges faced in modeling the spread of the South American leafminer [10,15,26].

Here we give some example works from other domains that share features with the models introduced in this paper. Human contact networks for epidemiology have a natural multi-scale organization being composed as a union of networks constructed over multiple types of locations such as workplaces, school, stores, and places of worship. The modeling by for example Mistry et al. [16] and Voigt et al. [27], directly model and calibrate such component networks using data on contact rates and and contact duration form the person-person contact network. Eubank et al. [7] construct the same type of network but use the approach of synthetic populations: by fusing demographic data, highly detailed geo-spatial data on residences, activity locations, and activity sequences derived from surveys, one can also derive a person-person contact graph. In Barrett et al. [1], a coupled, network system involving a contact network, a transportation network, an electrical power network, and a communication network are constructed for analysis of the National Planning Scenario no. 1.

2 Network Construction Framework

Here, we briefly describe a multi-pathway spread model, of which the seasonal commodity flow network is a critical component. Spatial spread processes such as invasive species spread occur at multiple scales through different pathways that can be broadly categorized as natural spread (self-mediated, wind, water, etc.) and human-mediated spread (trade of host crops, transportation, etc.) [4,15]. Our multi-pathway model consists of two component graphs defined at different spatial scales. The first component is the self-mediated dispersal pathway $G_S(V, E)$ defined at a grid level. A grid is overlayed on the study region, and let V denote the set of grid cells. Each cell v is associated with an attribute vector $a(v, t)$, where t corresponds to a time stamp. Cell attributes can correspond to administrative levels to which the cell belongs to, quantity of host crop production and consumption, population, climatic variables, number and types of operations, etc. The spread occurs from one grid cell to its adjacent cells. The edge set E is defined accordingly using some distance-based neighborhood criterion such as Moore or Euclidean-distance-based neighborhood. A schematic for the construction of this graph is provided in Fig. 1 (gray blocks). Specific construction details are provided in the description of the SN and NP networks.

Let $\{L_1, L_2, \ldots\}$ denote a collection of n_L subsets of V that are mutually disjoint. Each L_i corresponds to a *locality* $\underline{v_i}$. Localities represent areas of high human activity that are relevant to human-mediated spread of the invasive

species. These include production and consumption areas. Each locality typically consists of spatially contiguous nodes representing a city or a district for example. The second component graph corresponds to the *inter-locality graph* that captures the locality to locality trade of host crops and its influence on spread. This graph is denoted by G_{LD} and is defined on the localities. Let V_{LD} denote the set of all localities. The graph F_{LD} is defined on the locality set $\mathcal{L}$ with each edge directed representing link between two localities. In our case, the edge weight is directly proportional to the amount of (estimated) flow of the host crop. A schematic for the construction of this graph is provided in Fig. 1 (blue blocks). Specific construction details are provided in the description of the SN and NP networks.

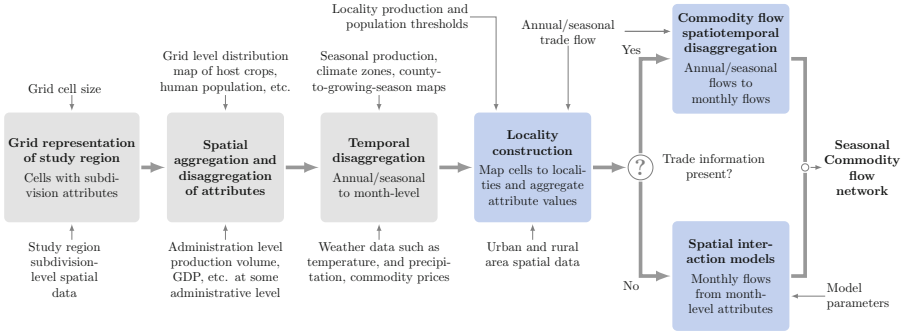

Fig. 1. The pipeline for constructing multi-pathway networks. (Color figure online)

2.1 The Senegal Network (SN)

For each cell, the production was assigned as follows. Spatial Production Allocation Model (MAPSPAM) [28] provides estimates of vegetable production at a finer grid resolution. This was mapped to the grid cells in our model. For cell v, let m_v denote this value. Even though, this quantity is not representative of tomato production in the cell, it is indicative of how suitable the cell is for tomato production. In particular, cells not suitable for any production can be identified using this quantity. For Senegal, through extensive surveys seasonal production data at province-level and trade at city and market level were collected in 2017 [17]. There are three major seasons: cold dry (November to February), hot dry (March to June), and rainy. First, the production was disaggregated to monthly production by uniformly distributing it to corresponding months. Let total tomato production in each department (administrative unit 2) S be denoted by $P(S,t)$ for month t. For each cell $v \in S$, the estimated tomato production is given by $P(v,t) = P(S,t) \times m_v / \sum_{v' \in S} m_{v'}$. Here, we distribute the production to cells according to weights obtained from MAPSPAM data. In some states where the production is very low, it is possible

that $\forall v' \in S$, $m_{v'} = 0$, but $P(S,t) \neq 0$. To avoid such scenarios, initially we add a small positive constant ϵ to each $m_{v'}$.

For locality construction, we used trade data obtained from survey as the reference. For most cities we found that mapping each city to the second administrative unit (referred to as department) was most suitable. However, there are some cities that are very small compared to the departments they belong to. In such cases, a circular region with the center corresponding to the coordinates of the city was chosen as the locality so that the population covered was comparable to the known population of the city. Like production, commodity flow for each season from locality $\underline{u}$ to $\underline{v}$ was divided uniformly over the months corresponding to that season. In Fig. 2, we have visualized the Senegal networks for the months of January and March in the backdrop of a heatmap of cell-level tomato production.

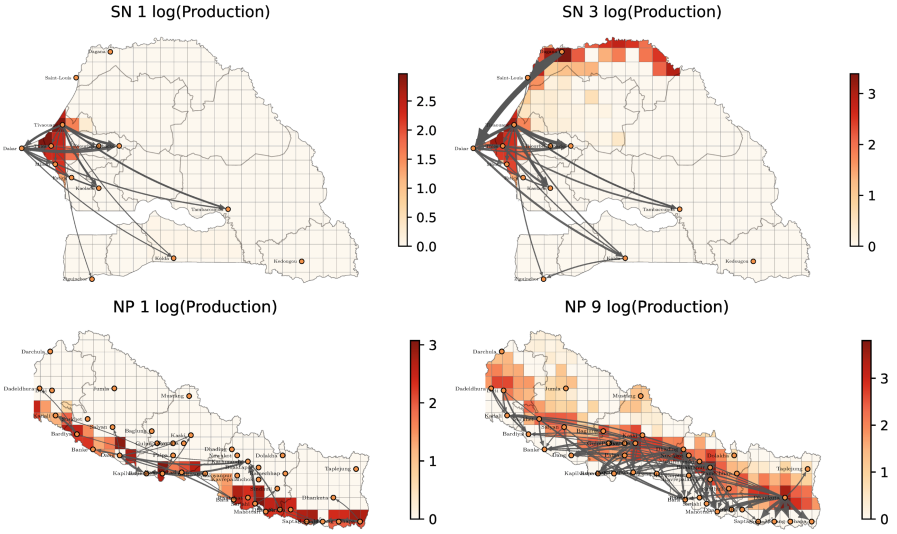

Fig. 2. Seasonal tomato flow networks along with production. The edge weights are proportional to logarithm of the flow volume. In the title, we have the network name followed by month. In each case, to limit the number of edges displayed, we have applied a threshold on the edge weight.

2.2 The Nepal Network (NP)

The production and trade data for this network was obtained from Venkatramanan et al. [26]. Here, grid and node attributes are created the same way as in the SN network. Annual tomato production data is provided at the administrative level 3 (district). Using MAPSPAM and production data, spatial disaggregation of the production was performed in the case of Senegal network. For

temporal disaggregation, we used information about tomato growing seasons. Unlike Senegal, the period of cultivation of vegetables depends on the region. Nepal has a unique geography; in a span of 100 kms from south to north, the elevation increases from near sea level to several thousand meters above sea level (masl). Accordingly, the country is divided into three regions: Terai (67–300 masl), Mid Hills (700–3000 masl), and High Hills (>3000 masl). Tomato production season widely varies across these regions [26]. We used this information to uniformly disaggregate the annual production for each cell to monthly production.

To construct localities, we used major vegetable wholesale markets data. Since markets belonging to the same district are very close to each other, we used the districts as localities. For Nepal, trade data is available for only one market. Therefore, we applied a doubly-constrained gravity model [13,15,26]. The main assumption here is that the trade volume from locality $\underline{u}$ to $\underline{v}$ is driven by production at $\underline{u}$, consumption at $\underline{v}$, and the distance between them. For each locality i, let $O_i(t)$ and $I_i(t)$ denote total outflow and total inflow for month t. The total outflow accounts for amount of production in the locality, imports and exports, i.e., $O_i(t) = \text{production}(i,t) + \text{import}(i,t) - \text{export}(i,t)$. Inflow $I_i(t)$ corresponds to estimated consumption, which is modeled as a function of population and GDP. The flow F_{ij} from locality $\underline{v}_i$ to city $\underline{v}_j$ is given by $F_{ij}(t) = a_i(t)b_j(t)O_i^{\alpha_1}(t)I_j(t)^{\alpha_2}f(d_{ij})$, where, d_{ij} is the time to travel from $\underline{v}_i$ to $\underline{v}_j$, and $f(\cdot)$ is the *distance deterrence function*: $d_{ij}^{-\beta}\exp(-d_{ij}/\kappa)$, where α_1, α_2, β and κ are model parameters. The coefficients a_i and b_j are computed through an iterative process such that the total outflow and total inflow at each node agree with the input values [13]. For distance, we used the Google API [9] to compute travel time between pairs of localities. In Fig. 2, we have visualized the NP network for the months of January and September.

3 Network Analysis

Dynamics of Production and Trade in the SN Network. The SN network spans 331 cells and comprises of 14 localities. We recall that for SN network, the trade data was available, while for NP, it was estimated using gravity model. Using decision trees [3], we analyzed the relationship between the flow and node attributes (production and population) and edge attribute (distance) between source and target. The results are in Fig. 3. Our analysis indicates that production at source is a primary driver of outflows (plots in the first two rows). There is no evidence of a clear inverse relationship with distance. We notice some very long distance flows from a high production area to the city capital. This could also be attributed to the country's population distribution, which is concentrated on the west coast. Another important thing to notice is that target locality production and target population are highly correlated (last row in the figure) indicating that much of tomato production occurs close to localities where there is demand.

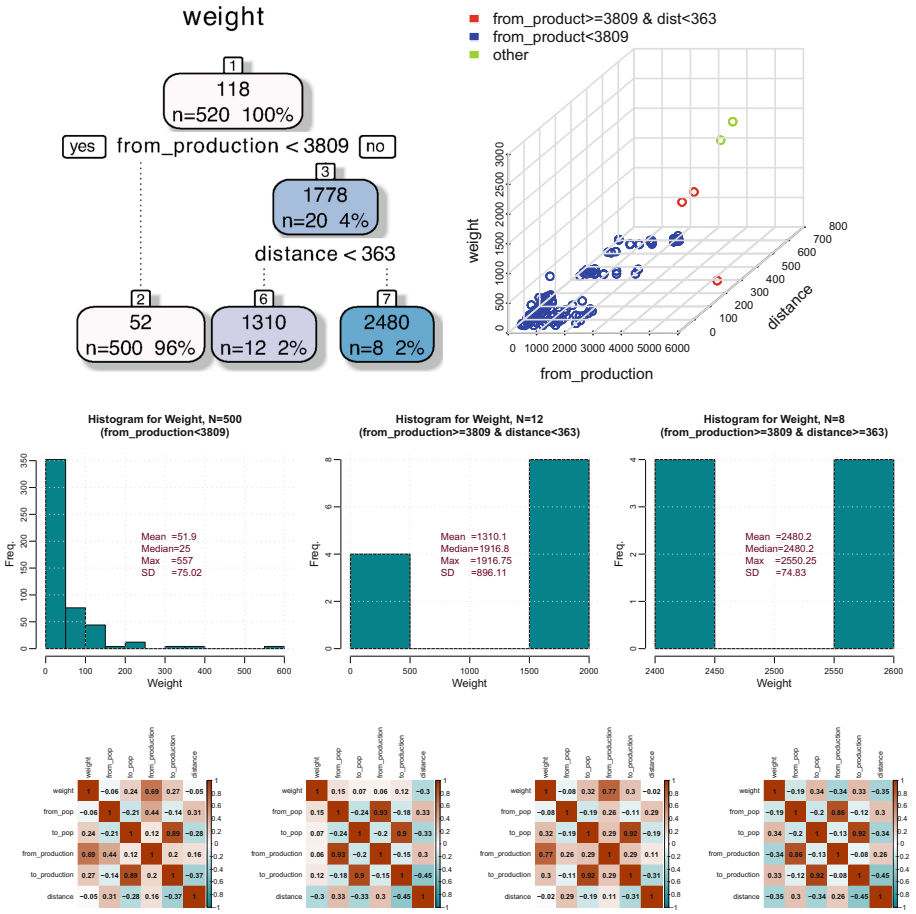

Fig. 3. SN network analysis. Here, "from" means source locality, "to" means target locality, "prod" is production, "pop" is population, "distance" is the distance between source and target, and "weight" is the commodity flow volume. In the last row, the heatmaps are for "All" seasons, "cold dry", "hot dry", and "rainy".

Dynamics of Production and Trade in the NP Network. The NP network spans 274 cells and comprises of 42 localities, three times that of SN network. This suggests that the role played by commodity trade in the spread process is much higher in NP than in SN. For the NP network, even though trade flow information is unavailable, there is some information on commodity flow from various districts to one of the most important wholesale markets in the capital city Kathmandu. We used this information to compare our model outputs for various parameter values $(\alpha_1, \alpha_2, \beta)$. The wholesale market data consists of inflows. To compare with this dataset, we computed the inflows to the Kathmandu locality from each locality as well as aggregated inflows from each zone,

which is a collection of districts). We used Pearson correlation coefficient for comparison of the two sets of inflows, and analyzed the correlation with respect to model parameters using decision tree analysis. The results are in Fig. 4. We note that at the district level, the correlation is very weak, while at the zone level, it is very strong. This shows that at a coarser spatial resolution the gravity model is matching the reference data. At the district level, it is possible that the distribution of production to neighboring localities does not match ground-truth. However, from a spread perspective, this would be adequate enough to make useful inferences.

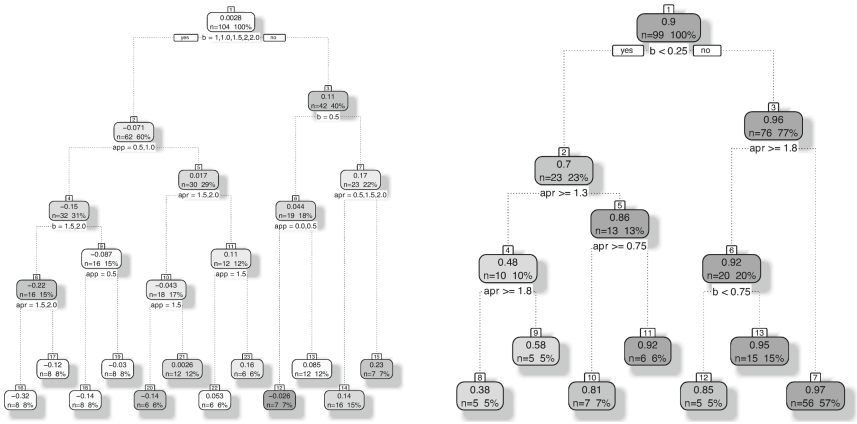

Fig. 4. NP network comparison with respect to an independent dataset on wholesale market data. The first plot corresponds to district level inflows compared with reference dataset. The second plot corresponds to zone (coarser) level inflows.

Cell-Level Production and Locality Flows. We recall that localities were chosen based on the level of human activity such as high population and high production of the crop. Therefore, the choice of localities is not completely dependent on cell-level production. Here, our goal is to analyze if there is any relationship between spatial distribution of cells with high production and characteristics of localities. The results are in Fig. 5. The first plot endorses the observation that production is highly concentrated; very few cells cover a majority of the production in each country. In SN, around 10% of the cells cover 90% of the production as Senegal is a semi-arid region with hardly any production in majority of the areas. In NP, around 30% of the cells cover 90% of the production. For each cell, we identified the nearest locality. In the second plot, we note that most of the high-production cells are very close to localities. This is part by design as high production areas are deliberately chosen as localities. However, as observed earlier, in many areas of high population, there is also high production. In the third plot, we note that for many top producing cells, their nearest localities have significant outflows. This is quite significant in the case of NP, and could

be attributed to seasonal shift in production. This shows that *many cells are particularly vulnerable to invasions due to imports from other localities.*

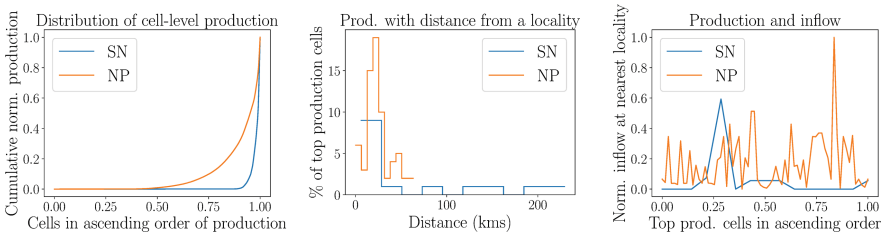

Fig. 5. Relationship between cell-level production and locality-level flows. In the last figure, we note that for SN, the plot corresponds to trade data while for NP, it corresponds to the outcome of the gravity model.

4 Conclusion

In this work, we developed a framework for constructing realistic agricultural commodity flow networks from disparate datasets. We applied it to construct two domestic tomato trade network in the context of invasive species spread. The resulting networks exhibit richer spatio-temporal characteristics than revealed by any individual dataset that was used to construct them. Future directions include, but are not limited to dynamical analysis of these networks using simulators, constructing networks for other regions of interest and other application domains (for e.g., economic impact), and novel network analytics to understand and validate these networks. In this regard, a potential candidate for applying the developed methods is the region of North America, which has been preparing for the impending invasion of the South American leafminer. Aggregate vegetable trade networks [12] and high-resolution gridded production data is available for this region, which present interesting opportunities as well as unique challenges for network inference. Our work emphasizes the need for developing representations that reflect the complexity of the phenomena being studied.

Acknowledgments. This work has been partially supported by USAID under the Cooperative Agreement No. AID-OAA-L-15-00001, USDA NIFA FACT 2019-67021-29933, DTRA (Contract HDTRA1-19-D-0007), University of Virginia Strategic Investment Fund award number SIF160, NSF grants IIS-1633028 (BIG DATA), CMMI-1745207 (EAGER), OAC-1916805 (CINES), CCF-1918656 (Expeditions), OAC-2027541 (RAPID) and IIS-1908530, The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

References

1. Barrett, C., et al.: Planning and response in the aftermath of a large crisis: an agent-based informatics framework. In: Proceedings of the 2013 Winter Simulation Conference, pp. 1515–1526 (2013). https://informs-sim.org/wsc13papers/includes/files/132.pdf
2. Biondi, A., Guedes, R.N.C., Wan, F.H., Desneux, N.: Ecology, worldwide spread, and management of the invasive South American tomato pinworm, *Tuta absoluta*: past, present, and future. Annu. Rev. Entomol. **63**, 239–258 (2018)
3. Breiman, L.: Classification and Regression Trees. Routledge, Milton Park (2017)
4. Carrasco, L., et al.: Unveiling human-assisted dispersal mechanisms in invasive alien insects: integration of spatial stochastic simulation and phenology models. Ecol. Model. **221**(17), 2068–2075 (2010)
5. ComTrade: Import and export (2021). http://comtrade.un.org/db/
6. Ercsey-Ravasz, M., Toroczkai, Z., Lakner, Z., Baranyi, J.: Complexity of the international agro-food trade network and its impact on food safety. PLoS ONE **7**(5), e37810 (2012)
7. Eubank, S., et al.: Modelling disease outbreaks in realistic urban social networks. Nature **429**(6988), 180–184 (2004)
8. FAO: Production and trade (2019). http://www.fao.org/faostat/en/#data
9. Google: Distance Matrix API (2017). https://developers.google.com/maps/documentation/distance-matrix/
10. Guimapi, R.Y., Mohamed, S.A., Okeyo, G.O., Ndjomatchoua, F.T., Ekesi, S., Tonnang, H.E.: Modeling the risk of invasion and spread of *Tuta absoluta* in Africa. Ecol. Complex. **28**, 77–93 (2016)
11. Hulme, P.E.: Trade, transport and trouble: managing invasive species pathways in an era of globalization. J. Appl. Ecol. **46**(1), 10–18 (2009)
12. Hwang, H.L., et al.: The freight analysis framework version 4 (FAF4)-building the FAF4 regional database: data sources and estimation methodologies. Technical report, Oak Ridge National Lab. (ORNL), Oak Ridge, TN (United States) (2016)
13. Kaluza, P., Kölzsch, A., Gastner, M.T., Blasius, B.: The complex network of global cargo ship movements. J. R. Soc. Interface **7**(48), 1093–1103 (2010)
14. Manning, L., Baines, R.N., Chadd, S.A.: Deliberate contamination of the food supply chain. Br. Food J. **107**(4), 225–245 (2005). https://doi.org/10.1108/00070700510589512
15. McNitt, J., et al.: Assessing the multi-pathway threat from an invasive agricultural pest: *Tuta absoluta* in Asia. Proc. R. Soc. B **286**(1913), 20191159 (2019)
16. Mistry, D., et al.: Inferring high-resolution human mixing patterns for disease modeling. Nat. Commun. **12**(1), 323 (2021). https://doi.org/10.1038/s41467-020-20544-y
17. Mohamed, I.A.: Cartographie des flux commerciaux de la tomate au Sénégal: reconstitution des filières, modélisation des flux et rôle dans la propagation du ravageur *Tuta absoluta* (2017)
18. Nath, M., et al.: Using network reliability to understand international food trade dynamics. In: Aiello, L.M., Cherifi, C., Cherifi, H., Lambiotte, R., Lió, P., Rocha, L.M. (eds.) COMPLEX NETWORKS 2018. SCI, vol. 812, pp. 524–535. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-05411-3_43
19. Nopsa, J.F.H., et al.: Ecological networks in stored grain: key postharvest nodes for emerging pests, pathogens, and mycotoxins. BioScience **65**(10), 985–1002 (2015)

20. Pimentel, D., Zuniga, R., Morrison, D.: Update on the environmental and economic costs associated with alien-invasive species in the united states. Ecol. Econ. **52**(3), 273–288 (2005)
21. Robinson, C., Shirazi, A., Liu, M., Dilkina, B.: Network optimization of food flows in the US. In: 2016 IEEE International Conference on Big Data (Big Data), pp. 2190–2198. IEEE (2016)
22. Rosenzweig, C., Iglesius, A., Yang, X.B., Epstein, P.R., Chivian, E.: Climate change and extreme weather events-implications for food production, plant diseases, and pests (2001)
23. Singh, S., Kumar, R., Panchal, R., Tiwari, M.K.: Impact of COVID-19 on logistics systems and disruptions in food supply chain. Int. J. Prod. Res. **59**(7), 1993–2008 (2021)
24. Suweis, S., Carr, J.A., Maritan, A., Rinaldo, A., D'Odorico, P.: Resilience and reactivity of global food security. Proc. Natl. Acad. Sci. **112**(22), 6902–6907 (2015)
25. USDA: Farms and Land in Farms 2017 Summary (2018). https://www.nass.usda.gov/Publications/Todays_Reports/reports/fnlo0218.pdf
26. Venkatramanan, S., et al.: Modeling commodity flow in the context of invasive species spread: study of *Tuta Absoluta* in Nepal. Crop Prot. **135**, 104736 (2020)
27. Voigt, A., et al.: Containing pandemics through targeted testing of households. BMC Infect. Dis. **21**, 548 (2021). https://bmcinfectdis.biomedcentral.com/articles/10.1186/s12879-021-06256-8
28. You, L., Wood-Sichra, U., Fritz, S., Guo, Z., See, L., Koo., J.: Spatial Production Allocation Model (SPAM) 2005 v3.2 (2017). http://mapspam.info