Synthetic Supply Networks

Galvin Ng[1,2], Damien Bertrand[3], Luca Mungo[4,5], and François Lafond[5,6]

[1]Complexity Science Hub, Vienna, Austria
[2]Vienna University of Economics and Business
[3]École Polytechnique Fédérale de Lausanne, Switzerland
[4]Macrocosm Inc, New York, USA
[5]Smith School of Enterprise and the Environment, University of Oxford
[6]Institute for New Economic Thinking, University of Oxford

Keywords: Synthetic population, supply networks, input-output analysis, firm-level data, gradient-based optimization

Motivation. There is increasing recognition that modelling economic dynamics at a highly detailed level can improve our ability to predict and monitor economic systems [1, 2, 3, 4]. Modelling economic systems at a near 1:1 scale requires representations of individuals (workers, consumers, and household members) and firms (buyers, suppliers, and employers). While detailed synthetic populations of households now exist [5], there is a lack of synthetic populations of firms and the supply networks linking them. Modelling economic systems requires not only realistic firms but also realistic firm-to-firm transaction networks. Recent datasets mapping complete national business-to-business transactions reveal that supply networks have a highly specific structure (compared to other complex networks) and are fairly universal [6].

However, confidentiality concerns limit access to administrative micro-data, and re-identification risks make it delicate for researchers to release machine learning models trained directly on raw firm-level data [7, 8].

Compared to other firm-level network reconstruction approaches [9, 10, 11, 12], as reviewed in [13], we are the first to rely exclusively on publicly available information.

Approach and Methodology. We generate synthetic supply networks in two stages: topology construction and weight assignment.

Firstly, we generate in- and out-degree sequences by sampling from Burr distributions calibrated to match the empirical mean, tail exponent, and variance of the log-transformed variable. We then construct a sparse directed binary network A_{ij} using the configuration model, which preserves the specified degree sequences.

Secondly, to assign weights, we draw firm-level "fitnesses" f^{in} and f^{out}, which serve as latent variables. We draw these from a joint lognormal distribution whose mean vector and covariance matrix are inferred from those of the in- and out-strengths reported in [6]. Conditional on the network topology, we initialize edge weights as

$$\mathbf{W}_{ij}^{\text{init}} = (f_i^{\text{out}})^{\delta}(f_j^{\text{in}})^{\gamma}(k_i^{\text{out}})^{\beta}(k_j^{\text{in}})^{\alpha}\mathbf{A}_{ij},$$

and rescale them to match the industry-level Input-Output Table (IOT):

$$\mathbf{W}_{ij} = \frac{\text{IOT}_{g_i g_j}}{\sum\limits_{f\in g_i}\sum\limits_{h\in g_j}\mathbf{W}_{fh}^{\text{init}}}\mathbf{W}_{ij}^{\text{init}}.$$

The parameters $\alpha, \beta, \gamma, \delta$ are optimized using the *Adam* algorithm to minimize a weighted loss function targeting multiple empirical network properties, including tail exponents of strength distributions, strength-degree correlations, and strength-degree elasticities. This differentiable, gradient-based calibration allows the synthetic network to simultaneously capture the micro-level features while remaining exactly consistent with the macro-level Input-Output Table. This implementation is fast, scales very well, and will be released as an easy-to-use Python repository.

Results. Table 1 and Figure 1 present the network and distributional properties of a synthetic supply network of $100,000$ firms that matches the 2015 Belgium Input-Output Table upon aggregation. The synthetic network reproduces the firm-level properties documented in [6] very closely, with the exception of clustering and reciprocity, which cannot be matched using a configuration model (we are exploring fine-tuning algorithms that can fix this). Notably, the synthetic network reproduces a key non-trivial feature emphasized in [6]: the out-degree distribution has much fatter tails than the in-degree distribution.

Conclusions and Outlook. Our results show that it is possible to construct plausible synthetic networks consistent with macroeconomic aggregates without direct access to administrative micro-data. This framework can be extended to other economic networks recorded in national accounts, including consumer–firm networks based on scanner and survey data, as well as employer–employee networks, using publicly released statistics. As distributed micro-data initiatives make higher-order network moments available, our pipeline can incorporate inter-country trade links, geographic structure, and richer industry-level constraints in a straightforward and modular way. We release the algorithm as open source, enabling researchers to generate synthetic networks for OECD ICIO countries by specifying a target country, number of firms, and selected loss components.

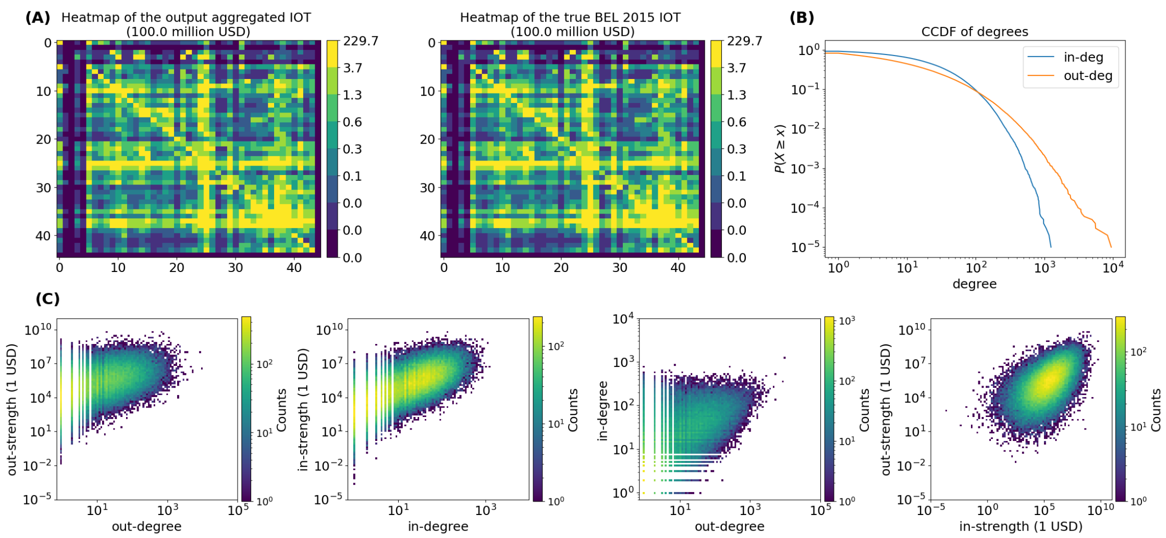

Figure 1: Distributional properties of the synthetic firm-level supply network for $100,000$ firms, optimized to match the 2015 Belgium Input-Output Table (IOT). (A) Industry-level aggregation of the synthetic network compared with the 2015 Belgium IOT (values in units of 100 million USD). The root mean square error is $66,000$ USD, which is negligible. (B) Complementary cumulative distribution functions (CCDFs) of the in- and out-degrees distribution. (C) Joint distributional relationships across strengths and degrees. The synthetic network closely replicates the empirical supply network.

Metric	Value	Empirical	Metric	Value	Empirical
Number of nodes	100,000	100,000	Hill α, k^{in}	2.5	2.5
Number of edges	3,796,120	4,000,000	Hill α, k^{out}	1.6	1.5
LWCC	94,639		Hill α, s^{in}	1.1	1
Share $k^{\text{out}} = 0$	13		Hill α, s^{out}	1.1	1
Share $k^{\text{in}} = 0$	2.5		Hill α, influence	1.5	[1.2, 1.3]
Mean degree	37.96	40	Hill α, weights	1.1	[1.1, 1.2]
Max k^{in}	1,251		Corr: $k^{\text{in}} \sim k^{\text{out}}$	0.54	0.55
Max k^{out}	9,439		Corr: $s^{\text{in}} \sim s^{\text{out}}$	0.54	0.5
Var log k^{in}	1.5	2	Corr: $s^{\text{out}} \sim k^{\text{out}}$	0.48	0.5
Var log k^{out}	2.7	3	Corr: $s^{\text{in}} \sim k^{\text{in}}$	0.63	0.75
Var log s^{in}	8.8	9	OLS: $s^{\text{out}} \sim k^{\text{out}}$	0.88	0.76
Var log s^{out}	9	8	OLS: $k^{\text{out}} \sim s^{\text{out}}$	0.26	0.33
Avg path length	3	3	R^2(out)	0.23	0.25
Degree assortativity	-0.014	$[-0.2, -0.01] < 0$	OLS: $s^{\text{in}} \sim k^{\text{in}}$	1.5	1.4
Reciprocity	0.002	[0.03, 0.05]	OLS: $k^{\text{in}} \sim s^{\text{in}}$	0.26	0.4
Avg clustering	0.014	[0.19, 0.28]	R^2(in)	0.39	0.56
Global clustering	0.01	low	TLS: $s^{\text{out}} \sim s^{\text{in}}$	1.02	1
			TLS: $k^{\text{in}} \sim k^{\text{out}}$	0.58	0.7

Table 1: Summary statistics of the synthetic firm-level supply network for 100,000 firms, optimized to match the 2015 Belgium Input-Output Table (IOT). In-degree is the number of suppliers, out-degree is the number of customers. The "Value" column report values computed from the synthetic network, compared against empirical values taken from [6]. LWCC stands for Largest Weakly Connected Component. OLS and TLS stand for Ordinary and Total Least Squares, and α stands for a tail exponent computed using a Hill estimator.

References

[1] Robert L Axtell and J Doyne Farmer. Agent-based modeling in economics and finance: Past, present, and future. *Journal of Economic Literature*, pages 1–101, 2022.

[2] Sebastian Poledna, Michael Gregor Miess, Cars Hommes, and Katrin Rabitsch. Economic forecasting with an agent-based model. *European Economic Review*, 151: 104306, 2023.

[3] Samuel Wiese, Jagoda Kaszowska-Mojsa, Joel Dyer, Jose Moran, Marco Pangallo, Francois Lafond, John Muellbauer, Anisoara Calinescu, and J Doyne Farmer. Forecasting macroeconomic dynamics using a calibrated data-driven agent-based model. *arXiv preprint arXiv:2409.18760*, 2024.

[4] Marco Pangallo, Alberto Aleta, R Maria del Rio-Chanona, Anton Pichler, David Martín-Corral, Matteo Chinazzi, François Lafond, Marco Ajelli, Esteban Moro, Yamir Moreno, et al. The unequal effects of the health–economy trade-off during the covid-19 pandemic. *Nature Human Behaviour*, 8(2):264–275, 2024.

[5] Marijn J Ton, Michiel W Ingels, Jens A de Bruijn, Hans de Moel, Lena Reimann,

Wouter JW Botzen, and Jeroen CJH Aerts. A global dataset of 7 billion individuals with socio-economic characteristics. *Scientific Data*, 11(1):1096, 2024.

[6] Andrea Bacilieri, András Borsos, Pablo Astudillo-Estévez, Mads Hoefer, and François Lafond. Firm-level production networks: what do we (really) know? Technical report, Institute for New Economic Thinking at the Oxford Martin School, University of Oxford, 2025.

[7] Luc Rocher, Julien M Hendrickx, and Yves-Alexandre De Montjoye. Estimating the success of re-identifications in incomplete datasets using generative models. *Nature communications*, 10(1):1–9, 2019.

[8] James Jordon, Lukasz Szpruch, Florimond Houssiau, Mirko Bottarelli, Giovanni Cherubin, Carsten Maple, Samuel N Cohen, and Adrian Weller. Synthetic data–what, why and how? *arXiv preprint arXiv:2205.03257*, 2022.

[9] Leonardo Niccolò Ialongo, Camille de Valk, Emiliano Marchese, Fabian Jansen, Hicham Zmarrou, Tiziano Squartini, and Diego Garlaschelli. Reconstructing firm-level interactions in the dutch input–output network from production constraints. *Scientific reports*, 12(1):11847, 2022.

[10] Luca Mungo, François Lafond, Pablo Astudillo-Estévez, and J Doyne Farmer. Reconstructing production networks using machine learning. *Journal of Economic Dynamics and Control*, 148:104607, 2023.

[11] Massimiliano Fessina, Giulio Cimini, Tiziano Squartini, Pablo Astudillo-Estévez, Stefan Thurner, and Diego Garlaschelli. Inferring firm-level supply chain networks with realistic systemic risk from industry sector-level data. *arXiv preprint arXiv:2408.02467*, 2024.

[12] Leonardo Niccolò Ialongo, Sylvain Bangma, Fabian Jansen, and Diego Garlaschelli. Multi-scale reconstruction of large supply networks. *arXiv preprint arXiv:2412.16122*, 2024.

[13] Luca Mungo, Alexandra Brintrup, Diego Garlaschelli, and François Lafond. Reconstructing supply networks. *Journal of Physics: Complexity*, 5(1):012001, 2024.