Chapter 16
Networks and AI

Valeria Bodishtianu, Yanli Lin, Oliver Kiss, and Girish Bahal

Abstract This chapter traces the development of network econometrics from its origins in spatial dependence models and social interaction frameworks through to modern machine learning methods. It asks how the econometric treatment of interdependence among agents has changed as network data became available, methods for formation and interference were developed, and computational tools expanded to handle graphs at scale. The chapter covers peer-effects estimation on observed networks, strategic and statistical models of link formation, causal inference under network interference, and the more recent contributions of graph neural networks and large language models, connecting each development to the economic questions that motivated it. The central argument is that network structure has become both a source of econometric variation and an economic object in its own right, but that each methodological advance has reintroduced classical problems of identification, measurement, and interpretation in new forms.

Valeria Bodishtianu ✉
University of Western Australia, Perth, Australia
e-mail: valeria.bodishtianu@uwa.edu.au

Yanli Lin
University of Western Australia, Perth, Australia
e-mail: yanli.lin@uwa.edu.au

Girish Bahal
University of Western Australia, Perth, Australia
e-mail: girish.bahal@uwa.edu.au

Oliver Kiss
Spreadmonitor, Budapest, Hungary
e-mail: oliver.kiss@spreadmonitor.com

16.1 Introduction

For much of its history, empirical econometrics approached modeling agents as if each made decisions in relative isolation, with interaction entering consideration mainly through aggregate equilibrium behavior and outcomes. This abstraction was both helpful for productivity, and, frankly, often necessary — tractability of the models and a workable framework for statistical inference took priority. However, it forced the field to leave some very important questions outside the model: who is connected to whom, exactly, and how does the specific architecture of those connections affect the outcomes that we observe?

The past three decades have seen a significant shift in how we approach these questions. Network structure data (e.g., who is friends with whom, which firms or countries trade with or lend to one another, how information gets transferred between people and how it is affected in the process) started to become available on a scale that was impossible to collect or process before. That shift was not continuous: for long stretches of time researchers had to work with a very limited number of specialized sources, having to painstakingly construct the datasets themselves. School surveys, village censuses, government and bank records provided some of the first systematic sources; the later spread of digital platforms and machine-readable documents certainly brought on a much larger wave of data.

Alongside this sudden expansion, artificial intelligence and, more specifically, machine learning (ML) developed rapidly and gradually became tied to economics research questions as well. These methods offered new ways of working with the data, both in terms of analysis, collection, and representation. Natural-language processing (NLP) specifically opened a huge new avenue for data sources, allowing researchers to construct variables and agent connections from textual sources that had never been designed as datasets. Development of Large Language Models (LLMs) has certainly sped up that process. While their econometric use remains (relatively) new and uneven, economists have already started using them to classify text, construct economic networks, and generate measurements that previously would've been quite costly to obtain by hand.

Overall, over those decades, the network has ceased to be only a conceptual or theoretical device in empirical applications and has become a very much measurable, and sometimes even experimentally manipulable, object. However, as is usually the case, that transformation was not without a cost — each step, be it including the adjacency matrix in a peer-effects regression, or building up a structural model of network formation, created its own econometric difficulties that researchers had to gradually learn to deal with. The network is often likely to be endogenous because agents choose their links; influence is hard to distinguish from selection. Interference between connected units violates the no-interference assumptions maintained in conventional causal models and forces researchers to think around them. Networks constructed from surveys, administrative records, or textual documents will, of course, contain measurement error. Flexible black-box models, that came to us from other fields have reopened the gap between predictive fit and causal interpretation, and therefore the debate over the relationship between the two. The history of network

econometrics, therefore, should not be treated as a linear story of uninterrupted triumphant progress; rather, it is a story of a recurring cycle, in which new data and methods open up new horizons and possibilities, and at the same time bring back old problems of identification, measurement, and model specification in new and occasionally convoluted forms.

This chapter traces this cycle from its origins to its current state. We begin with the main predecessors: sociometry, graph theory and spatial econometrics, which supplied network researchers with much of the language and the initial tools. Section 2 reviews the main predecessors: Manski's framework for social interactions, group-based models of peer effects, and the spatial autoregressive tradition that provided the first econometric tools for modelling cross-unit dependence. In Section 3 we describe the emergence of network econometrics in 1990s and 2000s, including the definitions used throughout the chapter, the meaning the economists attached to those terms, the first systematic data sources of link-level data, and the development of peer-effects models on observed networks and methods adapted for their estimation. Section 4 is dedicated to network formation models, moving from descriptive network measures to game-theoretic and statistical models of link formation. Section 5 examines causal inference under network interference, covering experimental designs, quasi-experimental approaches, and inference under network dependence. Section 6 turns to the introduction of big data and how it affected the scale and the form of networks data and analysis, including high-dimensional network data, embeddings, dynamically evolving networks and state-dependent links. Section 7 turns to Artificial Intelligence and machine learning, covering probabilistic graphical models, graph neural networks (GNNs), causal machine learning under interference, and LLM-constructed economic networks. Section 8 reconnects all the previously discussed methods to the central questions of the economics of networks: welfare, inequality, information diffusion, social learning, and systemic risk.

The main lessons from the chapter can be summarized as follows. The first is that network structure, rather than the mere existence of the links between agents, has become a source of both economic content and econometric variation. Under suitable assumptions, network features and characteristics become appropriate and useful tools for identification. The second lesson is that the scalability and predictive power introduced by machine learning and AI overall can complement, but not substitute for, structural economic reasoning. A model might predict which firm is likely to default, but that prediction alone cannot tell us which mechanism in the network is the cause of the default risk, or what would happen if an outside intervention changed the network structure. This tension between prediction and interpretation is familiar from the broader debate over machine learning in econometrics, but it becomes particularly sharp in econometrics of networks due to the network being both the source of the data, an object of economic interest, and a channel through which shocks and policy interventions propagate.

Part I: Foundations

We start from early matrix-based spatial tools and show how network economists generalised them to observed social/economic networks, and then extended to formation and causal designs under spillovers.

16.2 Spatial Models and Social Interactions (1970s–1990s)

The econometric study of spatial dependence and social interactions developed along two broadly parallel tracks during the 1970s and 1980s, converging only imperfectly before the reflection problem identified by Manski (1993) placed identification at the centre of the literature. One track inherited a body of statistical theory on spatial autocorrelation; the other emerged from economists' growing interest in neighbourhood effects, peer behaviour, and the aggregate consequences of individual interactions. Their intersection produced a set of tools that proved both influential and, in several respects, problematic.

The formal apparatus of spatial econometrics was consolidated primarily through the work of Cliff and Ord (1973), whose monograph on spatial autocorrelation gave statisticians and applied researchers a systematic treatment of dependence structures across geographic units. Building on earlier probabilistic foundations due to Whittle (1954), Cliff and Ord (1981) extended this programme and provided an empirical toolkit that would become standard in regional science and, eventually, in mainstream applied econometrics. The core objects of the resulting literature were two model classes. The spatial autoregressive model (SAR) specifies that each unit's outcome depends linearly on a weighted average of its neighbours' outcomes, scaled by a parameter ρ, alongside observed covariates and a disturbance. The spatial weights matrix W determines which neighbouring units enter this average and with what weight (Cliff & Ord, 1973; Anselin, 1988). The spatial error model (SEM) instead leaves the conditional mean free of spatial lags and places the dependence in the disturbances, which themselves follow a spatial autoregression governed by a parameter λ (Anselin, 1988). In the SAR model, dependence propagates through the cross-section in a way formally analogous to autoregression in time. In the SEM model, the systematic component of the conditional mean is not itself spatially lagged, but unobserved shocks are spatially correlated. Both specifications share a common ingredient: the spatial weights matrix W, an $n \times n$ array whose (i, j)-th entry encodes the strength of the presumed link between units i and j. In the canonical formulation that Anselin (1988) synthesised into the first comprehensive econometric treatment of the field, W is typically constructed from geographic contiguity or inverse distance, row-standardised so that each row sums to unity, and then treated as known and fixed throughout estimation and inference. Ord (1975) provided the maximum likelihood estimator for the SAR model and derived its properties under the assumption of a known W, enabling a wave of applications across regional income convergence, housing markets, agricultural yields, and local public expenditure.

The most durable contribution of this framework is the weights matrix as a way of representing interdependence. By collecting cross-sectional links in a single operator,

spatial econometrics gave researchers a tractable way to describe how outcomes or shocks in one unit may be related to those in other units. This idea later carried over naturally to network econometrics. Once researchers began to study social and economic networks directly, the network adjacency matrix played essentially the same role as W, except that its entries were based on observed links rather than geographic proximity. Much of the algebra developed in the spatial literature therefore remained useful: reduced-form equilibrium outcomes in linear interaction models, stability conditions restricting the strength of feedback, and the interpretation of indirect effects through powers of W. These ideas reappeared, in richer and more explicitly network-based form, in the network econometrics literature of the 2000s and 2010s.

A less durable feature of this literature was the tendency to treat W as known, exogenous, and correctly specified. In many applications, researchers constructed weights matrices from geographic proximity and then treated the resulting matrix as fixed. This practice combined two separate assumptions: first, that distance is a useful proxy for economic or social interaction; and second, that a particular rule, such as contiguity, inverse distance, k-nearest neighbours, or distance bands, captures the relevant channel well enough. These assumptions were often not examined as carefully as the coefficients estimated after W had been chosen. The broader survey discussion in Anselin and Bera (1998) already made clear how central specification choices are in applied spatial work, and influential applications such as Case (1991) illustrated both the usefulness and the limits of modelling local interaction through geographically defined neighbourhoods. The deeper issue is that geography, while correlated with many economic and social linkages of primary interest, such as trade, information diffusion, labour mobility, local public goods, or social contact, is not itself those linkages. Treating geography as a sufficient measure of interaction structure introduced a form of misspecification that standard robustness checks often struggled to diagnose.

This limitation was particularly important because early spatial methods were frequently applied in settings where the true mechanism of dependence was not purely geographic. Distance-based weights are natural in some applications: environmental spillovers, land use, agricultural diffusion, housing markets, or regional infrastructure all have a plausible spatial component. But for many economic and social outcomes, the relevant channels run through workplaces, schools, kinship networks, production chains, or information networks. Administrative neighbours may not be social neighbours, and social neighbours may not be geographically adjacent. Thus, one of the central lessons of this period is that a tractable dependence structure is not the same as a credible interaction structure. The weights-matrix framework aged well because it gave econometricians a portable language for interdependence; the unvalidated use of makeshift "neighbour" definitions aged much less well because it often substituted convenience for measurement.

The economic interest in peer effects and neighbourhood influences predates the formalisation of the reflection problem, but it was Manski (1993) who gave the literature both its canonical specification and its central challenge. Consider a population partitioned into reference groups. In stylised form, individual i's outcome is modelled as a function of i's own characteristics, the mean characteristics of i's

reference group, and the mean outcome of that group. The coefficient on the group's mean outcome is the endogenous social effect δ, the response of behaviour to peers' behaviour; the coefficient on the group's mean characteristics is the exogenous or contextual effect γ; and correlated effects arise when members of the same group behave similarly because of shared environments, common unobserved traits, or non-random sorting (Manski, 1993).

Manski (1993) showed, with considerable generality, why this interpretation is difficult. In a linear model with homogeneous parameters, the endogenous social effect and the contextual effect are not separately identified from group-level variation alone: a regression of individual outcomes on group means cannot distinguish between a world in which individuals respond to peers' outcomes and one in which they respond only to peers' predetermined characteristics. The reflection metaphor captures the intuition. When A affects B and B simultaneously affects A, one observes the equilibrium reflection of each in the other without being able to attribute the correlation to its structural source. Observed similarity among group members may reflect endogenous interaction, contextual composition, correlated unobservables, common shocks, or sorting. This impossibility result applies regardless of sample size; it is a population identification failure, not a finite-sample problem.

By isolating the precise source of non-identification, the framework immediately suggested what would be needed to recover it: variation in group composition that is plausibly exogenous to individual unobservables, restrictions that separate endogenous from contextual effects, nonlinearity in the social interaction function, or access to the network structure of interactions rather than group membership alone. The subsequent empirical literature pursued each of these avenues. Sacerdote (2001) exploited the random assignment of roommates in college dormitories to isolate exogenous variation in peer composition; other studies used institutional rules generating quasi-random variation in class, cohort, or neighbourhood composition. As network data became available, the graph structure of connections provided identifying variation that the group-level model could not exploit, because two individuals can share the same group but have different sets of direct and indirect links. Such network intransitivity can create differential exposure to peer outcomes that helps break the reflection symmetry under appropriate rank and exclusion-type conditions (Bramoullé, Djebbari & Fortin, 2009).

The reflection problem is therefore one of the more instructive cases in the history of applied econometrics: an identification issue that had always been present, but whose systematic statement transformed the field's self-understanding. Estimates of neighbourhood and peer effects had been reported throughout the 1970s and 1980s without adequate treatment of the simultaneity at their core. Manski's contribution was to show that this was not merely a practical difficulty that could be remedied by a larger sample or a more efficient estimator, but a structural feature requiring richer data, a richer model, or a more credible source of identifying variation. This placed identification considerations, rather than estimation technique alone, at the centre of subsequent work, a reorientation that resonated well beyond social interactions and contributed to the broader identification turn in applied microeconometrics.

What did not survive from this period is the easy conflation of group membership with interaction structure. The linear-in-means model often assumes, explicitly or implicitly, that every pair of individuals within a reference group interacts with equal intensity. This is a strong and typically implausible restriction. Real social interactions are sparse and heterogeneous: most individuals interact intensively with a small subset of their peers and negligibly with the rest. Imposing a dense, symmetric group structure on what is in reality a sparse network discards identifying information and may bias parameter estimates in ways that depend on the correlation between network position and individual characteristics. The subsequent network econometrics literature can in part be read as a response to this misspecification. Once the analyst has access to the adjacency matrix of interactions rather than only group boundaries, the reflection problem can often be relaxed or resolved under additional rank and exclusion-type conditions, and richer heterogeneity in social influence becomes estimable. But the deeper lesson is that credible estimation of social effects requires the interaction structure itself to be measured, justified, and connected to a plausible source of identifying variation.

The 1980s and 1990s also saw a number of early quasi-network applications that anticipated later developments without yet having the data infrastructure or theoretical language of modern network econometrics. Their distinctive contribution was not that they solved the identification problem, but that they revealed the data gap that would motivate the network turn. Studies of interlocking directorates, information transmission in agricultural communities, neighbourhood effects, local crime interactions all treated social or economic proximity as a source of dependence. Researchers increasingly wanted to study interactions among classmates, co-workers, firms, or criminals, but the relevant links were often observed only indirectly. Group membership, geographic location, classroom assignment, or administrative region therefore became proxies for networks.

This proxy strategy produced useful empirical questions but fragile answers. Random classroom assignment may provide useful peer variation, while neighbourhood sorting can badly confound neighbourhood averages and geographic clustering of firms may reflect common cost conditions rather than knowledge spillovers. Early quasi-network applications were therefore most convincing when the institutional setting generated plausibly exogenous variation in group composition or exposure, and least convincing when the reference group was defined mainly because it was available in the data. The quasi-network period helped clarify the two requirements that would organise later network econometrics: measurement and identification. Measurement asks whether the researcher observes the relevant links, or only proxies for them. Identification asks whether variation in those links or exposures can distinguish interaction from sorting, common shocks, and contextual composition.

Modern network econometrics can be viewed as a synthesis of the spatial and social-interactions traditions. It inherits from spatial econometrics the idea that dependence can be represented through a matrix, but it asks more carefully whether that matrix actually represents the relevant economic links. It inherits from the social-interactions literature the insistence that the identifying variation must have an

economically meaningful interpretation, but it moves beyond group averages toward more detailed link structures.

In retrospect, the legacy of the spatial and social-interaction literature of the 1970s–1990s is double-edged. It established the conceptual architecture of equilibrium interaction models and the algebraic tools for handling cross-sectional dependence. It also demonstrated, through its own limitations, that geography is often a poor substitute for observed economic links, and that group averages are often a poor substitute for network structure. The weights-matrix framework endured because it provided the matrix-based language later generalised to networks. The over-reliance on makeshift neighbour definitions proved less durable because it left too much of the interaction structure unvalidated. Manski's reflection problem has proved especially consequential because it reframed the central question from how to estimate peer effects to whether they are identified at all. Modern network econometrics grew out of precisely this combination: a useful mathematical representation of interdependence, a persistent concern about causal interpretation, and a growing recognition that the structure of connections is itself an economic object rather than a background detail.

16.3 The Emergence of Network Econometrics (1990s–2000s)

The language that economists use when they speak of networks is borrowed, with varying degrees of closeness to the source, from graph theory and social network analysis. Graph theory traces its commonly accepted origin to Euler (1741), who replaced the physical layout of Königsberg and its bridges by an abstract collection of points and connections, turning a question about walking routes into a question about the structure of connections among points and its features. Social network analysis has a separate history, beginning with the sociometric diagrams of Moreno (1934), who drew maps of people and their relationships in order to study the internal structure of small groups. Some of the graph-theoretic language later used in this literature was systematised by Harary (1969), while Wasserman and Faust (1994) gave the social-network tradition its comprehensive methodological reference, with definitions of concepts such as centrality, distance, and structural equivalence.

By the time these ideas entered economics, the vocabulary was therefore already old; economists, as one does with useful mathematics, kept the notation and gave it their own interpretation. Vertices turned into decision-making agents or institutions; edges into channels through which information, goods, or risk could pass. A network therefore, in the sense now standard in the economics literature, consists of a finite set of agents (nodes or vertices), and a collection of pairwise relationships among them (links, ties, or edges). The mathematical object that defines the structure of these relationships is the adjacency matrix D, an $n \times n$ matrix whose entries D_{ij} record

whether a link exists between agents i and j, or, in more general formulations, the intensity or weight of that link.[1]

Some useful distinctions and definitions on describing these networks also followed from a graph-theoretical vocabulary. For instance, in an undirected network, a link between i and j is reciprocal by definition: if $D_{ij} = 1$ then $D_{ji} = 1$, and the adjacency matrix is symmetric. Many economic relationships have this feature, at least on the surface, including marriages, co-authorships, and some bilateral trade agreements. In a directed network, by contrast, the link from i to j need not imply a link from j to i, and the adjacency matrix can be asymmetric. Citation networks, where paper i cites paper j without any obligation of reciprocity, provide a standard example; so do directed supply-chain links and friendship nominations in surveys.

A separate distinction concerns link intensity. In an unweighted (binary) network, $D_{ij} \in \{0, 1\}$ agents are either connected or they are not, and the adjacency matrix is a matrix of zeros and ones. In a weighted network, D_{ij} can take continuous values that represent the strength, frequency, or volume of the interaction: following our previous examples, it could be the amount of bilateral trade between two countries, the number of co-authored papers linking two researchers, or the strength of a friendship, whether self-reported or observed through interaction frequency. Early empirical work in network econometrics typically operated with binary networks, in part because the available data recorded only the presence or absence of a link, and in part because the theoretical models were formulated for unweighted graphs. As richer data became available, the weighted formulation gained importance and prominence, since the economic content of a connection often depends on its intensity in addition to the fact of it existing at all. The similarity with the spatial-econometric framework is immediate: both use a matrix to represent dependence among agents or units. What distinguishes network econometrics is the status given to that pattern of interaction itself. In spatial econometrics, the weights matrix W usually summarises a maintained pattern of dependence among units; it is important for estimation, but its entries are not usually interpreted as choices or relationships, whose own structure calls for economic explanation. In network econometrics, by contrast, the topology of D itself is part of the economic object of the study. This distinction is what makes the move to observed peer networks important: once the entries of D are interpreted as relationships, the relevant peer group is narrowed down from the whole neighborhood (or classroom, or cohort) to a particular set of neighbours to whom an individual is connected. The move from group-based social interaction models to peer-effects models on observed networks changed the empirical object itself. In the earlier literature reviewed in the previous section, individuals were assigned to reference groups, such as classrooms, neighbourhoods, or cohorts, and

[1] The notation varies across literatures. In spatial econometrics the weights matrix is typically denoted W; in the social interactions and peer-effects literature the adjacency matrix is often denoted G (as in the preceding section and in the discussion of peer-effects models below). We use D for the raw adjacency structure and G for the row-normalised version that appears in peer-effects models: if $d_i = \sum_j D_{ij} > 0$, then $G_{ij} = D_{ij}/d_i$, so that G_{ij} gives the weight that agent i places on each of its d_i neighbours. Isolated nodes, for which $d_i = 0$, require a separate convention, typically a zero row. See Jackson (2008) and Goyal (2007) for related discussion.

the relevant peer measure was a group average. The adjacency structure within the group was left unspecified: every pair of group members was treated as equally connected. The availability of detailed network data, most notably the Add Health survey of friendship nominations among American adolescents,[2] made it possible to replace the group average with a weighted average over each individual's observed or nominated peers. This shift preserved the linear-in-means structure while giving it a more explicit economic interpretation. Writing G for the row-normalised adjacency matrix, so that G_{ij} is the weight placed by i on peer j, the network peer-effects model relates each individual's outcome to the average outcome of their network neighbours, Gy; their own characteristics; and the average characteristics of their neighbours, GX. The coefficient on Gy is the endogenous social effect ρ, the own-characteristic coefficients capture the association with individual covariates, and the coefficient on GX captures contextual or exogenous peer effects. The structural similarity to the SAR model of spatial econometrics is immediate: the network model replaces the geographic weights matrix W with an observed, potentially asymmetric adjacency matrix G, whose entries are determined by actual social or economic ties rather than assumed proximity.

The key advance over Manski's group framework was that the network structure itself provided identifying variation. Bramoullé et al. (2009) showed that when the network contains intransitive triads, configurations in which i and k are both connected to j but not to each other, the characteristics of k can affect i's outcome through j's behaviour without entering i's outcome equation directly. This creates a natural exclusion logic: characteristics of peers of peers who are not direct peers may be correlated with the endogenous social effect but excluded from the direct contextual effect. Under appropriate rank conditions involving G, G^2, and the covariate matrix, endogenous and contextual effects can be separately identified. The intuition is elegant. Network intransitivity breaks the perfect collinearity between peer outcomes and peer characteristics that makes the reflection problem irresolvable in the group model. Calvó-Armengol, Patacchini and Zenou (2009) applied related network ideas to Add Health data, estimating peer effects in educational achievement and delinquency using the position of each student in the friendship network as a structural predictor. Their results illustrated both the empirical promise of the approach and its sensitivity to the quality of the underlying network data.

The substantive gain from attaching explicit network structure was interpretability. Knowing not just that peers matter, but which peers, through which connections, and with what intensity, gave the peer-effects model a precision that the group-average specification could not achieve. A student whose friends are central in the network faces a different peer-effect exposure from a peripheral student even if both belong to the same classroom. A firm whose suppliers are densely connected to other large buyers occupies a structurally different network position from a firm at the periphery

[2] Add Health refers to the National Longitudinal Study of Adolescent to Adult Health, a longitudinal study of a nationally representative sample of U.S. adolescents who were in grades 7–12 in the 1994–95 school year. The study includes friendship nomination information that has been widely used to construct adolescent school friendship networks; see the official Add Health documentation at https://addhealth.cpc.unc.edu/.

of a supply chain. The network model made these distinctions estimable rather than absorbing them into a homogeneous group effect. More broadly, the entries of G could correspond to observed friendships, transactions, exposures, or information channels rather than researcher-imposed distance bands.

Yet the early literature also revealed a persistent vulnerability. The observed network, represented by the adjacency matrix G, was typically treated as fixed and exogenous in estimation. This is the network analogue of the spatial econometrics assumption that W is given. As in the spatial case, the assumption is often difficult to defend. Friendship nominations may reflect unobserved characteristics, mutual selection, or common shocks that are correlated with the outcomes being studied. A student who is nominated as a friend by high-achieving peers may be high-achieving for the same unobserved reasons that attracted those friendships, rather than because of peer influence. If G is endogenous, standard estimators for ρ can conflate influence with homophily or selection into links. The recognition that network endogeneity is a first-order problem, not a second-order refinement, became central to later work and motivated the formation models reviewed in the next section.

The estimation of network autoregressive models required adapting and extending the econometric tools developed for spatial models. In the peer-effects model of the previous subsection, the term Gy is endogenous even when G is treated as fixed, because each agent's equilibrium outcome depends on the outcomes of others through network feedback. The reduced form involves the inverse matrix $(I-\rho G)^{-1}$, which makes every element of Gy a function of the disturbance vector u. Ordinary least squares is therefore generally inconsistent for ρ, even when the network is taken as given.

The identification result of Bramoullé et al. (2009) translates naturally into an instrumental-variables strategy. Peers-of-peers characteristics, represented by G^2X, help predict Gy through the network equilibrium but, under an appropriate exclusion restriction, need not enter the structural equation directly. Higher-order network neighbours, such as G^3X, may provide additional instruments when the network is sufficiently sparse and when indirect connections remain plausibly excluded from direct effects. The resulting overidentifying restrictions can also be examined empirically, although their credibility ultimately depends on the economic plausibility of the exclusion restrictions. Liu and Lee (2010) formalised a GMM framework for network autoregressive models, deriving moment conditions that combine linear IV moments with quadratic moments from the model's network structure, in a way related to the approach developed for spatial autoregressions by Kelejian and Prucha (1999).

Quasi-maximum likelihood estimation offered an alternative that does not require explicit instrument selection. Lee (2004) established asymptotic properties of the QMLE for spatial autoregressive models under a Gaussian quasi-likelihood, showing that the estimator can be consistent and asymptotically normal under conditions that do not require the true disturbance distribution to be Gaussian. The network analogue uses the reduced form generated by $(I-\rho G)^{-1}$ and imposes stability restrictions on ρ, so that the feedback induced by the network does not become explosive. Such restrictions are usually expressed in relation to the eigenvalues or spectral radius of

G. Economically, they require local interaction effects not to amplify without bound as they circulate through the network. Large-sample arguments also depend on the structure of the graph. Sparsity conditions are often used to ensure that no individual node's influence dominates the cross-sectional average, but such conditions become restrictive in highly centralised networks, where a small number of nodes account for a large share of connections.

The GMM and QMLE approaches each carried implicit assumptions about the network's role in identification. GMM relied on the exclusion of higher-order peer characteristics from the structural equation, an assumption that is economically interpretable when agents have limited information about their extended network, but may fail when information travels rapidly or when agents strategically account for indirect connections. QMLE relied on stability and dependence restrictions that can be difficult to verify in practice, especially in financial networks, interbank markets, or trade networks with highly connected hubs. Neither approach, in its early form, fully addressed the possibility that G itself might be endogenous. Both therefore inherited the same vulnerability as the peer-effects models they were designed to estimate. The econometric advances reviewed above would have remained largely theoretical if it had not been for a parallel development in the availability of data that recorded bilateral links directly or at least made them directly observable. The Add Health survey of adolescent friendship nominations became one of the more influential datasets for the econometric study of peer effects, but it was not the only empirical setting in which network data started to matter. From the late 1990s and through the early 2010s, researchers in several areas of economics began gathering up network data of their own, often by hand and at considerable cost, and the resulting empirical studies made it quite easy to see both the promise and the limitations of the rapidly developing network econometric toolkit.

In financial economics, the question of how the failure or distress of one institution can spread to others had long been discussed, but often in less quantitative terms — in part because the network structure of these exposures was difficult to observe. Allen and Gale (2000) provided an influential theoretical framework in which the pattern of interbank claims, whether diversified across many counterparties or concentrated in a sparser structure, determines the extent and severity of financial contagion. Their analysis showed that the same aggregate shock can produce very different systemic outcomes depending on the architecture of the interbank network, giving network structure a direct role in the theory of financial stability. On the empirical side, Furfine (2003) proposed a method for inferring bilateral federal-funds exposures from payments data, and Boss, Elsinger, Summer and Thurner (2004) used supervisory data from the Austrian central bank to reconstruct the interbank network and study its topological properties. These early empirical studies illustrated what could be learned from network data, and revealed how much of the network remained unobserved or had to be estimated rather than measured.

In international trade, Rauch (1999) argued that bilateral flows reflect not only factor endowments and trade costs but also network channels of information and trust, especially for differentiated products where prices are less informative and matching buyers with sellers is harder. Rauch and Trindade (2002) extended this argument to

ethnic Chinese networks (one of the first papers to treat diasporas as proper economic networks rather than demographic categories), showing that the presence of ethnic Chinese populations in both the origin and destination country was associated with higher bilateral trade volumes, an effect they interpreted as evidence that social and business networks reduce informational barriers to trade. These studies did not estimate peer-effects or network autoregressive models of the kind reviewed above; their contribution was to show that network structure, understood broadly as the pattern of pre-existing social and institutional connections, could explain variation in economic outcomes that standard gravity-type regressions left unaccounted for. A more explicitly graph-theoretic treatment of trade data followed with the work of Serrano and Boguñá (2003), who represented countries as nodes and bilateral trade flows as directed links and studied the topological properties of the resulting network, finding scale-free degree distributions, high clustering, and short average path lengths.

In development economics, a growing number of studies collected link-level network data in rural settings, often through labour-intensive household surveys. Fafchamps and Gubert (2007) recorded bilateral transfers and mutual insurance links between households in rural Philippines and showed that geographic proximity was a major determinant of risk-sharing links; while kinship and differences in age and wealth also mattered, occupational differences and the correlation of incomes, by contrast, had little explanatory power. Conley and Udry (2010) collected data on information-sharing networks among pineapple farmers in Ghana by directly asking each farmer whom they talked to about farming practices, and used this measured communication structure to show that farmers adjust their inputs toward those of information neighbours who had been surprisingly successful in previous seasons. One of the most extensive data-collection efforts of this kind was the village network census conducted by Banerjee, Chandrasekhar, Duflo and Jackson (2013), who collected detailed social-network data across 75 villages in Karnataka, India, and used the resulting networks to study how the diffusion of microfinance participation depends on the network centrality of the individuals who are informed first. Their finding that diffusion centrality of initial seeds predicted village-level adoption rates gave network topology a directly measurable role in the analysis of diffusion, and the dataset itself became one of the more widely used resources in the empirical networks literature.

The common limitation of these early network datasets was not only their scale and coverage, but the incomplete and uneven ways in which the economically relevant network was observed. Friendship networks were limited to the students in a school or a set of schools, interbank networks to a single national banking system, and risk-sharing networks to the households in a particular village. In each case, the boundary of the observed network was a function of the data collection design rather than of the economic process being studied: students who were friends with individuals outside the sampled schools or households that participated in arrangements beyond the village were either excluded or represented as missing links. These boundary and censoring problems would later become central to the literature on network measurement and inference, but in the early period they were mostly noted and

set aside. What mattered at the time was that the data existed at all, and that the econometric tools developed for spatial and peer-effects models could be applied to link-level data rather than to undifferentiated group averages or proximity-based dependence structures.

The lessons from this period are instructive in the same way that the spatial econometrics experience was instructive a generation earlier. Extending spatial tools to arbitrary adjacency matrices was a productive and durable contribution. It brought a rich estimation toolkit to bear on a much wider class of economic interaction problems, and it clarified conditions under which endogenous peer effects can be separated from contextual and correlated effects. Network intransitivity, in particular, showed that the topology of the network itself, rather than only external institutional variation, could help generate exclusion restrictions under suitable assumptions. What proved less robust was the maintained assumption that the observed network is fixed, correctly measured, and exogenous. This assumption was a natural first step, but it introduced into network econometrics the same specification risk that came with the spatial literature's uncritical use of geographic weights matrices. Selection and formation issues accumulated as a deferred problem that would require richer models of endogenous network formation, causal designs under interference, and methods accounting for uncertainty in the measured network.

16.4 Modelling Network Formation

Before structural econometric models of network formation became available, researchers put the network itself to use in a more reduced-form way, by computing summary statistics of the adjacency matrix and entering them into regressions. The metrics employed for this purpose had long histories outside of economics.

The most elementary measure of a node's position is its degree, the number of links incident to it. In a directed network, one distinguishes between out-degree (the number of links a node sends) and in-degree (the number it receives); in an undirected network the two coincide. Degree is easy to compute and interpret, but it is entirely local: it says nothing about where in the network i's neighbours are situated or how they are connected to one another.

The broader family of centrality measures was developed to capture richer aspects of a node's position.[3] Bavelas (1950) was among the first to formalise the idea that a person's position in a communication structure affects their ability to coordinate and influence others, drawing on small-group research in which communication was restricted to imposed structures. One of the earliest and amongst the more influential centrality indices with a recursive structure was proposed by Katz (1953), who introduced what is now called Katz centrality: it counts the walks emanating from

[3] For directed networks, we use the convention that D_{ij} records a link from i to j. Unless otherwise stated, the centrality measures below are written in their outgoing/reach interpretation, counting paths or walks emanating from node i. Incoming versions are obtained by replacing D with $D^\top$. For undirected networks $D = D^\top$ and the distinction disappears.

each node while giving progressively less weight to longer walks. Unlike degree, it therefore captures both direct and indirect connections.

Bonacich (1972) proposed eigenvector centrality as a related recursive measure[4]. A node's centrality is proportional to the sum of the centralities of the nodes to which it is linked; the recursive logic is that a node is central when it is connected to other central nodes. Bonacich (1987) generalised this into a parametric family of power centrality measures $c(\alpha,\beta)$ governed by a parameter β that controls how connections to well-connected nodes are valued. When $\beta > 0$, connections to well-connected others are advantageous, and the measure is closely related to Katz centrality; when $\beta < 0$, connections to poorly connected nodes are advantageous instead, capturing the idea that bargaining power is greater when one's partners have fewer outside options.

Banerjee et al. (2013) introduced diffusion centrality to measure the expected reach of information originating from a node over a finite number of periods. Suppose that an informed node transmits each piece of information to each neighbour with probability q in each period, and that diffusion continues for T periods. Diffusion centrality counts the expected number of times information originating from each node reaches other nodes through walks of length up to T, allowing the same node to be reached through multiple paths.

It also connects several familiar centrality measures. When $T = 1$, diffusion centrality is proportional to degree centrality. As T tends to infinity, it converges to Katz–Bonacich centrality when $q < 1/\lambda_1(D)$; when $q > 1/\lambda_1(D)$, its normalised ranking approaches eigenvector centrality as the diffusion horizon grows. Diffusion centrality therefore provides a bridge between local degree, finite-horizon diffusion, discounted walk-based centrality, and the long-run eigenvector ranking.

Freeman (1977) introduced betweenness centrality, which captures a different aspect of positional importance: the extent to which a node lies on the shortest paths between other pairs of nodes. In the standard undirected version, a node has high betweenness when many geodesic paths between other nodes pass through it. A node with high betweenness can be interpreted as a gatekeeper or broker whose removal would lengthen the paths between other agents. In a follow-up paper, Freeman (1979) unified degree, closeness, and betweenness under a single conceptual framework for thinking about what centrality means in a network. For an overall recent taxonomy of centrality measures, see Bloch, Jackson and Tebaldi (2023).

A second class of descriptive statistics concerned the aggregate structure of the network. The tendency for two neighbours of a given node to be themselves connected, a property known as transitivity, had been studied in social network analysis since the work of Holland and Leinhardt (1971) on triad counts and reciprocity in directed networks. Watts and Strogatz (1998) gave this idea renewed importance by defining the clustering coefficient of a node as the fraction of its neighbour pairs that are themselves connected, and showing that many real networks combine high average

[4] The idea behind eigenvector centrality is considerably older though: Landau (1895) proposed measuring the quality of tournament chess players by the strength of the opponents they beat, so that the quality vector satisfies an eigenvalue equation; precisely the eigenvector problem, though Landau did not use the term. In a follow-up Landau, 1914, he established the existence and positivity of the solution by invoking the then-recent Perron–Frobenius theorem.

clustering with short average path lengths, a combination they called the small-world property.

The question of what generates the degree distributions observed in real networks also attracted attention from outside the social sciences. Price (1965), studying citation networks among scientific papers, observed that the distribution of citations follows a heavy-tailed pattern and proposed a cumulative-advantage mechanism: papers that have already been cited many times are more likely to attract further citations. Price (1976) formalised this mechanism mathematically. Over three decades later, Barabási and Albert (1999) independently proposed a closely related preferential-attachment model for growing networks and showed that it produces power-law degree distributions, a finding that attracted wide attention across physics, computer science, and eventually economics. These findings entered economics somewhat indirectly, but they shaped the way economists thought about which features of a network might matter for outcomes and which models of network formation a reasonable specification would need to reproduce.

A particularly influential economic use of centrality appears in Ballester, Calvó-Armengol and Zenou (2006), which showed that in a class of network games with linear-quadratic payoffs and local complementarities, the Nash equilibrium action of each player is proportional to her Katz-Bonacich centrality. Building on the walk-based measures developed by Katz (1953) and Bonacich (1987), Katz-Bonacich centrality captures both direct and indirect connections while progressively reducing the weight placed on the more distant connections; the weighting is determined by the structure of the game rather than freely chosen by the researcher. The result gave centrality an equilibrium microfoundation rather than treating it as a descriptive index: a player's network position determined her equilibrium behaviour through the structure of strategic interaction, not only through a correlation with unobservables. From a policy perspective, the result implied that the key player whose removal leads to the largest reduction in aggregate activity is the player with the highest intercentrality measure, a weighted variant of Katz-Bonacich centrality that accounts for the indirect equilibrium effects of removal. The paper established a template for giving centrality measures equilibrium content rather than treating them as purely descriptive statistics. An applied example appears in Banerjee et al. (2013), discussed in the preceding section, who showed that the diffusion centrality of the individuals first informed about microfinance predicted village-level adoption rates, giving it a concrete empirical interpretation.

In practice, however, the use of network metrics in applied work often preceded this kind of microfoundation. Researchers used such network metrics in standard regression specifications, either to explain economic outcomes or as outcomes to be explained themselves (Hochberg, Ljungqvist & Lu, 2007; Calvó-Armengol et al., 2009; Hidalgo & Hausmann, 2009). The appeal was that these statistics were easy to compute from the adjacency matrix and could be given intuitive economic interpretations: a more central firm might have better access to information, or a more clustered neighbourhood might support more effective monitoring. The risk was that the same network position could be correlated with many unobserved characteristics, so that the coefficient on a centrality measure in a cross-sectional regression was

difficult to interpret causally without a model of how agents came to occupy those positions in the first place. This tension between the convenience of descriptive network metrics and the need for a structural account of why agents have the positions they have was one of the forces that pushed the literature toward the formation models discussed in what follows.

A natural next step after spatial and peer-effects models was to ask where the links themselves come from. In the earlier spatial literature, the matrix W was usually treated as fixed, known, and externally supplied. In many social and economic settings, however, this assumption is difficult to defend. Friendships, trading relationships, lending ties, co-authorship, inter-firm partnerships, and information links are often chosen by agents. They reflect preferences, constraints, opportunities, and strategic incentives. Treating such links as exogenous can therefore confound the effect of the network with the process that generated it. The econometrics of network formation emerged from this recognition: links are not merely a background structure through which outcomes propagate, but economic objects to be explained.

The move from describing networks to modelling their formation required a statistical framework capable of capturing the mechanisms by which links arise. The simplest approach treats each potential link as a binary outcome and regresses an indicator of link formation on dyad-level covariates, such as geographic proximity, similarity in observable characteristics, or pre-existing institutional connections, using probit or logit specifications. This dyadic regression approach, developed and applied in empirical studies of risk-sharing networks and social ties (Fafchamps & Gubert, 2007), provided a tractable entry point. It made homophily empirically operational: if agents with similar education, income, ethnicity, location, occupation, or exposure to institutions are more likely to form links, this can be represented through covariates measuring pairwise similarity. Such models were attractive because they were transparent, easy to estimate, and closely connected to applied questions about who interacts with whom.

The limitation is that dyadic regression treats links as conditionally independent once observed dyadic covariates are controlled for. This is often too restrictive. If i is linked to j, and j is linked to k, the probability that i links to k is often higher than a model of independent dyads would predict. This phenomenon, usually called triadic closure, is one of the most basic features of real networks. Similar dependence arises from reciprocity, popularity, common group membership, and the presence of shared intermediaries. A friendship between two people may depend on whether they have common friends; a trade relationship may depend on existing supply chains; a collaboration may be more likely when two researchers are connected through a third party. These forms of dependence are not dyadic in isolation. They are features of the network as a whole.

Exponential random graph models (ERGMs), sometimes called p^* models, were developed to address this limitation by specifying the probability of an entire network configuration as a member of the exponential family, with sufficient statistics capturing the local structural features of interest. The foundational work of Holland and Leinhardt (1981) introduced the p_1 model for directed graphs, parameterising link probabilities as a function of sender and receiver effects. Frank and Strauss

(1986) established the class of Markov random graphs, in which the presence of a tie between two nodes depends on the rest of the network only through the ties incident to those nodes. This framework accommodates statistics such as the number of triangles and k-stars, giving the model the capacity to represent triadic closure and degree heterogeneity simultaneously. Wasserman and Pattison (1996) extended and popularised the p^* framework, which became a dominant specification in the social networks literature through the late 1990s and 2000s.

In a generic ERGM, the probability of a network belongs to the exponential family: its log-probability is linear in a vector of network statistics $q(d)$, such as the number of links, reciprocated links, triangles, stars, or homophilous ties, weighted by parameters θ, and normalised by a constant that sums over all possible networks (Frank & Strauss, 1986; Wasserman & Pattison, 1996). For an undirected network on n nodes, that normalising sum runs over $2^{n(n-1)/2}$ configurations, which makes direct computation infeasible at any realistic size. The appeal of ERGMs is nevertheless genuine. By allowing the probability of a network to depend on structural features through sufficient statistics, the framework captures endogenous clustering, reciprocity, popularity, and degree concentration. Homophily, the tendency of agents to form links with others similar to themselves in income, education, ethnicity, occupation, or other characteristics, can be incorporated through dyadic covariates, while triadic closure and degree heterogeneity can be represented through network statistics. This separation matters economically. A network that looks clustered may owe its clustering to homophily alone, to endogenous triadic closure alone, or to both, and the distinction has different implications for the diffusion of information, the persistence of inequality, and the welfare effects of network-based interventions.

More fundamentally, highly parameterised ERGMs are prone to a degeneracy problem, where the model's probability mass can concentrate on unrealistic networks, such as nearly empty or nearly complete graphs, leaving little probability on the sparse, locally clustered configurations often observed in economic and social data (Handcock, 2003). Small perturbations in parameter values can shift the model toward these degenerate regions, making estimation surfaces difficult to navigate. These pathologies are not merely incidental computational difficulties. They reveal a deeper tension between the statistical flexibility of the ERGM class and the information content of a single observed network.

The lesson from this experience is not that ERGMs are unimportant. They provided a powerful way to represent network dependence statistically and made concepts such as homophily, reciprocity, and triadic closure estimable within a unified framework. What proved less convincing is the temptation to treat increasingly flexible ERGMs as self-validating because they fit a collection of observed network statistics. A model may reproduce the degree distribution and triangle count of a single observed network while still leaving the underlying parameters weakly identified or difficult to interpret structurally. The ERGM experience therefore clarified an important lesson for later network econometrics: fitting the topology of a network is not the same as explaining why the network formed.

The same distinction between fitting a network's topology and explaining its formation also applies to the descriptive metrics discussed above: a centrality measure

may be informative, but it does not explain why an agent occupies that position. A firm may be central because centrality creates informational or bargaining advantages, or because already profitable firms are better able to form and maintain links. The analogy to omitted-variable bias is direct, but harder to resolve, because the structure of the network links every agent's position to the positions and characteristics of everyone else in the graph.

The simplest benchmark for thinking about network formation is the class of random graph models introduced independently by Erdős and Rényi (1959), who studied graphs with a fixed number of edges placed uniformly at random, and by Gilbert (1959), who proposed the model in which each pair of nodes is connected independently with the same probability p. These models are mathematically tractable but they miss several features commonly observed in social and economic networks: real networks exhibit far more clustering, degree heterogeneity, and assortative mixing than a uniform random graph would predict. The departure from Erdős–Rényi and Gilbert random graph models therefore defined the agenda for formation models: further specifications would need to explain why observed networks have the features they have.

The response to this came from two research directions that had, until that point, developed somewhat independently. One was rooted in statistical network analysis, evolving from the p_1 model of Holland and Leinhardt (1981) through the ERGMs and latent-space models. The other was rooted in economic theory, where strategic models grounded link formation in payoffs and equilibrium concepts, as discussed below. These two traditions differ in what they take as primitive: the statistical models specify a probability distribution over graphs, often with local dependence assumptions, while the game-theoretic models specify payoff functions and equilibrium concepts, and the distribution over graphs is derived from the equilibrium analysis. From an econometric perspective, both traditions supplied ways to avoid treating the observed network as a fixed background object, and both were relevant to the recognition that conditioning on network structure without modelling its formation inherits a selection problem of unknown sign and magnitude. Goldsmith-Pinkham and Imbens (2013) made this point directly, showing that when network formation is modelled jointly with peer effects using Bayesian methods, the resulting estimates of peer influence can differ substantially from those obtained by treating the network as exogenous. The following discussion traces the game-theoretic and statistical lines of development in turn.

The economic theory of network formation begins with the idea that links are the result of decisions by agents who anticipate the consequences of their connections. One of the earliest contributions was Myerson (1977), who incorporated graph structure into the analysis of cooperative games, showing that when communication between players is restricted by a network, the feasible coalitions and therefore the payoff allocation depend on which players can reach one another through the graph. His allocation rule, later called the Myerson value, generalised the Shapley value to settings in which the cooperation structure is not the complete graph, and it established the principle that the structure of connections has payoff consequences that can be analysed formally. Aumann and Myerson (1988) took the next step by

making the formation of links endogenous: in their extensive-form game, players sequentially propose links, and each pair decides whether to form a connection, anticipating how the resulting cooperation structure will affect the division of surplus. This was a canonical early model in which the network emerged as an equilibrium object rather than being imposed as an exogenous constraint, and it provided the conceptual template for the non-cooperative literature.

The foundational paper in the non-cooperative area is Jackson and Wolinsky (1996), who introduced pairwise stability as a central solution concept for network formation. In their framework, a link between agents i and j forms if both agents consent and is severed if either agent prefers to remove it. A network is pairwise stable if no existing link is such that one of its endpoints would prefer to delete it, and no absent link is such that both endpoints would prefer to add it. Pairwise stability differs from Nash equilibrium in that it considers only single-link deviations and requires bilateral consent for link addition but allows unilateral deletion; this makes it a natural solution concept for settings where relationships require mutual agreement but can be dissolved by either party. Jackson and Wolinsky (1996) also introduced the connections model, in which the value of an indirect connection decays with the geodesic distance in the network. They showed that efficient networks need not be pairwise stable, resulting in a tension between individual incentives and social welfare.

Bala and Goyal (2000) proposed an alternative framework in which link formation is one-sided: each agent chooses a set of directed links to initiate, and the formation game is a standard simultaneous-move game in which Nash equilibrium is the natural solution concept. Depending on whether benefits flow only in the direction of the initiated link or along the link regardless of who initiated it, Bala and Goyal (2000) derived tractable equilibrium networks under one-sided link formation. The contrast with Jackson and Wolinsky (1996) is clear: unilateral formation is more tractable, while bilateral formation is more natural to many economic relationships that require mutual consent.

A further question is how a network evolves over time when agents have the opportunity to adjust their links. Jackson and Watts (2002) studied the dynamics of network formation by defining improving paths, sequences of networks in which each transition involves the addition or deletion of a single link that is beneficial for the agents involved. Their analysis showed that pairwise stable networks and efficient networks do not need to coincide even in the dynamic setting, and that the selection among multiple stable networks depends on the structure of the transition dynamics.

The main econometric challenge posed by game-theoretic formation models is multiplicity of equilibria. A formation game with n agents has $2^{n(n-1)/2}$ possible undirected networks in its outcome space, and the number of stable or equilibrium networks can be very large. If the data consists of a single observed network, one faces a problem that is familiar from the broader literature on estimation of games: the model predicts a set of networks consistent with the estimated parameters, not a unique network, and without further assumptions it may not be possible to recover the parameters from the data. Mele (2017) showed that if agents meet sequentially at random and myopically update their links, the network formation process can be cast

as a potential game whose stationary distribution takes the form of an exponential random graph model. This result bridges the game-theoretic and statistical traditions: the ERGM, originally a purely statistical object, receives a microfoundation from the strategic formation game, and the parameters of the ERGM can be interpreted in terms of the payoff primitives. Mele (2017) also noted a limitation related to a result established by Chatterjee and Diaconis (2013) in the probability literature: for dense networks with nonnegative link externalities, the ERGM stationary distribution becomes asymptotically close to a mixture of Erdős–Rényi random graphs, so that the structural parameters of the formation game are difficult or impossible to identify from the observed network.

An alternative approach to the multiplicity problem was more recently taken by de Paula, Richards-Shubik and Tamer (2018), who used partial identification methods to bound the parameters of a network formation game without selecting among equilibria. In their framework, each pair of agents has a surplus from forming a link that depends on observable characteristics and unobserved heterogeneity, and links form when the surplus exceeds a threshold, subject to a constraint on the maximum number of connections each agent can maintain. The bounded-degree assumption is economically motivated, since many real networks are sparse, and it provides the restriction needed to derive informative bounds on the preference parameters. Sheng (2020) took a similar approach, using the frequencies of small subnetworks (triads, tetrads) as the basis for inference on the parameters of pairwise-stable formation games, and showed how partial identification bounds can be computed even when the game allows many equilibria. Separately from the strategic-formation tradition, Graham (2017) proposed an econometric model of dyadic link formation that allows for unrestricted agent-level heterogeneity in link surplus, analogous to fixed effects in panel data, but does not model strategic interdependence between links. His tetrad logit estimator conditions on a sufficient statistic for the degree heterogeneity and recovers the homophily parameter, the tendency for agents with similar characteristics to form links, without imposing distributional assumptions on the unobserved node-level effects. The analogy to the conditional logit of Chamberlain (1980) is direct: the incidental-parameters problem that arises when each node has its own fixed effect is solved by conditioning on a sufficient statistic, just as in the panel-data case. Graham (2020) provides a comprehensive econometric treatment of network data, covering dyadic regression, network parameters, and strategic formation. A further strand of the game-theoretic literature addressed the question of how strategic considerations shape the composition of observed networks. For example, Currarini, Jackson and Pin (2009) provided a theoretical account of the patterns of racial and ethnic homophily documented in the Add Health friendship data, and explained how homophily can arise as an equilibrium outcome of strategic choice. What this line of work showed was that the composition of the network is itself an equilibrium object; not including the strategic incentives involved in the link creation might lead to the misattribution of the resulting features of the network to exogenous forces.

Estimation of statistical network formation models developed along several complementary paths. Pseudo-likelihood methods offered one practical response to the intractability of the full ERGM likelihood. Introduced by Besag (1975) and

applied to social network models by Strauss and Ikeda (1990), pseudo-likelihood approximates the full likelihood by the product of conditional probabilities for individual links, given the rest of the network. The resulting estimator resembles a logistic regression on dyad-level outcomes with network statistics as covariates, and it greatly reduces computational burden. However, the convenience comes at a cost. Because links are interdependent, pseudo-likelihood can perform poorly when dependence is strong, especially when the conditioning network is itself an endogenous object. It is therefore best understood as a useful approximation whose reliability degrades as dependence among links strengthens.

Simulation-based maximum likelihood provided a more direct response to the intractable normalising constant. The approach developed by Geyer and Thompson (1992) and adapted to ERGMs by Hunter and Handcock (2006) approximates likelihood ratios between candidate parameter values using Markov chain Monte Carlo draws from the model. Under appropriate conditions, this allows researchers to approximate the maximum likelihood estimator without enumerating all possible networks. In practice, however, simulation-based MLE does not by itself remove the underlying specification problem. It can make degeneracy and poor mixing visible, but it cannot guarantee that the chosen model places probability mass on realistic networks or that the estimated parameters have a stable interpretation across nearby specifications.

Method-of-moments and simulation-based approaches provided another route, especially as researchers moved toward richer models of formation. Rather than writing down the full probability of the network, researchers could match selected features of the observed network, such as average degree, clustering, homophily, or the distribution of common neighbours, to their model-implied counterparts. This approach is attractive because it focuses attention on substantively meaningful features of the network and fits naturally with economic models in which the likelihood is unavailable but simulation is possible. In models allowing strategic interaction or equilibrium constraints, researchers often simulate networks under candidate parameters and compare simulated features with the observed network, following the logic of simulated method of moments or indirect inference. These methods expanded the feasible set of models, but they also made computation a central part of econometric credibility. The cost is that identification and efficiency depend heavily on the chosen moments, the reliability of the simulations, and diagnostics showing that the simulated networks resemble the observed network in relevant dimensions.

The lesson from this period is that link endogeneity is a substantive feature of the data-generating process. A researcher who estimates peer effects or diffusion parameters while treating the observed network as exogenous assumes implicitly that the reasons agents form links are unrelated to the outcomes being studied. This assumption is rarely innocuous in economic applications, where agents select into relationships partly on the basis of characteristics, expectations, and opportunities that also determine their behaviour. Similarly, if policy changes the incentives to form links, then holding the network fixed may miss an important part of the policy effect. Link formation is often part of the economic mechanism rather than a preliminary object to be set aside before studying outcomes.

These developments suggest that credible network econometrics requires more than a tractable representation of links. Dyadic regressions made link formation easy to estimate but often ignored dependence among links. ERGMs represented dependence more flexibly but could become computationally fragile and difficult to diagnose. Moment-based and simulation-based methods allowed richer formation models but required careful attention to identification, simulation reliability, and computational diagnostics. Strategic models of network formation, which ground link formation in preferences, economic constraints, and equilibrium concepts, have proved especially durable because they connect estimable parameters to economic primitives. This is the principle that the ERGM experience illustrates in the negative: micro-founded and interpretable formation models have endured precisely because they explain why the network looks the way it does, rather than merely fitting its topology.

16.5 Interference, Experiments, and Causal Effects

The workhorse framework for causal inference in economics and statistics rests on the potential outcomes model, in which each unit i has a potential outcome $Y_i(d_i)$ that depends only on its own treatment status d_i. This assumption, that the treatment of one unit does not affect the outcomes of others, was formalised by Cox (1958) and later codified by Rubin (1980) as part of the Stable Unit Treatment Value Assumption, or SUTVA. SUTVA requires both that there are no hidden versions of treatment and that there is no interference between units. In many economic applications, particularly those involving policy interventions, public goods, or information, the no-interference requirement is a strong and often implausible restriction. When units are connected through a network, the treatment of one agent can propagate to others through social influence, information transmission, resource reallocation, or price adjustment. Vaccinating one person reduces the infection risk of their contacts; subsidising one farmer's adoption of a new technology may induce imitation among neighbours; providing employment opportunities to one worker may change the job-search behaviour of their peers. In each case, the outcome of an untreated unit depends on the treatment of connected units, and SUTVA is violated.

Ignoring interference on networks changes the estimand, not only the precision of the estimate. When interference is present, the observed outcome of an untreated unit is not the same as its potential outcome under universal non-treatment, because connected treated units have already changed the environment. A naive comparison of treated and untreated units in a network experiment therefore conflates the direct effect of treatment with the indirect effects that propagate through the network. If treated and control units are geographically or socially proximate, as they typically are when assignment is not carefully designed, control units may be partially treated through spillovers, and the estimated treatment effect may be attenuated or otherwise reinterpreted. Conversely, if the estimand of interest is the total effect of a policy that would apply to all connected agents simultaneously, the direct-only estimate

will misstate the policy-relevant effect: understating it when spillovers are positive, and overstating it when spillovers are negative, as the displacement result of Crépon, Duflo, Gurgand, Rathelot and Zamora (2013) discussed below illustrates. Sobel (2006) drew attention to this problem in the context of neighbourhood experiments, showing that standard estimators of neighbourhood effects may be fundamentally misidentified when residents of comparison neighbourhoods are indirectly affected by the intervention. Miguel and Kremer (2004) provided an influential empirical illustration in the context of a deworming programme in Kenya, where the health benefits of treatment spilled over to untreated children in the same school and community, making the conventional within-school comparison a poor estimate of either the direct or the total effect.

The recognition that interference is endemic in network settings did not immediately produce a unified identification framework, but it did clarify what such a framework would need to accomplish. At a minimum, it would need to define potential outcomes that remain stable under a specified pattern of interference, distinguish direct effects from spillover effects in a way that is both estimable and economically interpretable, and propose experimental or quasi-experimental designs capable of generating the variation needed to separate these components. Each of these requirements posed non-trivial challenges that the literature addressed progressively through the 2000s and 2010s.

The three requirements identified in the preceding discussion - stable potential outcomes under a specified interference pattern, a decomposition of direct and spillover effects, and a design that generates the variation needed to estimate them - gave rise to a set of experimental and quasi-experimental methods that developed progressively over the 2000s and 2010s.

The oldest response to interference is to randomise treatment at the level of groups rather than individuals. The formal analysis of this design in biostatistics begins with Cornfield (1978), who showed that randomising by group rather than by individual inflates the variance of the treatment-effect estimator and derived the conditions under which the efficiency loss is offset by the ability to capture group-level effects (see Moberg & Kramer, 2015 for a historical survey of this literature). These early contributions were motivated primarily by logistical and administrative considerations, since many public-health interventions are naturally delivered at the group level. The conceptual reframing came when Halloran and Struchiner (1995) recognised, in the context of vaccine trials, that cluster randomisation is a necessity when one individual's treatment affects another's outcome: individual-level randomisation confounds direct and indirect effects, while group-level and two-stage designs preserve a clean comparison. Hudgens and Halloran (2008) formalised this into a general framework, defining direct, indirect, total, and overall causal effects under what they called partial interference, the assumption that interference operates within pre-defined groups but not between them, and proposing a two-stage randomisation design in which clusters are first assigned to treatment regimes and individuals are then randomised within clusters, creating the variation needed to identify each component separately. In economics, the empirical importance of these ideas was established by Miguel and Kremer (2004) (as described above)

and Angelucci and De Giorgi (2009), who documented consumption spillovers in the Mexican Progresa programme. The method's limitation is that it requires the researcher to define the relevant clusters before randomisation, and when interference crosses cluster boundaries, the partial-interference assumption fails.

A refinement of cluster randomisation that allows the researcher to estimate how treatment effects vary with the intensity of treatment in the local environment is the saturation design. The conceptual idea that varying treatment intensity across groups could help identify social interaction effects was articulated by Moffitt (2001), who argued that standard programme evaluations miss the spillover channel when every eligible individual within a group is either treated or untreated. Hudgens and Halloran (2008) formalised this idea into a concrete experimental protocol, which they called the randomised saturation design: in the first stage, clusters are randomly assigned different treatment probabilities, or saturation levels; in the second stage, individuals within each cluster are randomly assigned to treatment or control at the cluster's assigned probability. This two-stage structure creates two independent sources of exogenous variation, an individual's own treatment status and the fraction of treated peers in their cluster, and allows the researcher to identify direct effects (the effect of one's own treatment holding the cluster saturation fixed) and indirect effects (the effect of increasing the cluster saturation holding one's own treatment fixed) as separate estimands. Crépon et al. (2013) provided the first large-scale implementation of this design in economics, applying it to a job-placement programme in France where different labour markets were randomly assigned different treatment fractions, and detecting displacement effects: treated job-seekers gained at the expense of the untreated in the same market, so that the net market-level effect was close to zero. Baird, Bohren, McIntosh and Özler (2018) subsequently derived the optimal allocation of saturation levels across clusters for maximising statistical power to detect both direct and spillover effects, providing a design-based framework that can be applied before the data are collected.

When the interference structure is a known social network rather than a set of pre-existing administrative or geographic clusters, the researcher can use the network itself to construct the experimental clusters. Graph partitioning, the problem of dividing a graph into groups of nodes with few edges crossing group boundaries, is a classical problem in computer science with a long algorithmic history, but its application to experimental design is more recent. Backstrom and Kleinberg (2011) were the first to connect graph partitioning to the problem of A/B testing on networks, introducing the concept of network bucket testing: for features whose value depends on whether a user's friends also have access to the feature, standard individual-level randomisation fails because treated users surrounded by untreated friends do not experience the feature as intended. Backstrom and Kleinberg proposed sampling well-connected subgraphs as test units and showed how unbiased estimators of mean user satisfaction can be constructed from such samples. Ugander, Karrer, Backstrom and Kleinberg (2013) built on this idea and formalised it into a general experimental design that they called graph cluster randomisation. Their procedure partitions the network into clusters, randomises treatment at the cluster level, and uses Horvitz–Thompson weighting based on the resulting exposure probabilities.

The key formal contribution was the definition of a network exposure condition under which a subject's effective treatment environment is fully determined by their cluster's assignment, and the derivation of conditions on the partition under which this exposure condition holds approximately. Eckles, Karrer and Ugander (2017) studied the bias-variance trade-off that governs the choice of partition resolution: coarser partitions capture more interference within clusters and reduce bias, but they also reduce the number of independent clusters and increase variance. Their analysis showed that graph cluster randomisation can substantially reduce mean squared error relative to individual-level randomisation, and they provided practical guidance on how to choose the partition granularity. This line of work was developed at and motivated by experiments at Facebook and LinkedIn, where the social network connecting users is observed at scale and individual-level A/B tests are known to be contaminated by interference Saveski et al., 2017.

Alongside these experimental methods, a quasi-experimental strategy exploited institutional sources of random variation in peer assignment to identify network and peer effects without requiring the researcher to design the randomisation around interference. The key methodological insight, first implemented by Sacerdote (2001) using the random assignment of first-year students to dormitory rooms at Dartmouth College, was that administrative randomisation of group membership provides exogenous variation in peer composition that directly solves the selection problem identified by Manski (1993). Zimmerman (2003) replicated the approach at Williams College and found that peer effects on academic performance were concentrated among students at the lower end of the ability distribution. Carrell, Sacerdote and West (2013) pushed the method further by exploiting the random assignment of entering cadets to squadrons at the US Air Force Academy. Their main finding (for the purposes of this chapter) was methodological: when the Academy attempted to engineer optimal peer compositions on the basis of the linear peer-effects estimates from the observational data, the intervention produced worse outcomes than random assignment, because the true peer effects turned out to be nonlinear and the endogenous formation of social subgroups within the assigned squadrons undermined the intended composition. Angrist (2014) provided a cautionary assessment of the broader quasi-experimental peer-effects literature, arguing that many studies suffer from weak instruments, mechanical correlations, and unclear economic mechanisms. Sacerdote (2014) surveys this literature comprehensively.

A more recent development uses network structure not only to design the randomisation but to choose the targets of intervention. Banerjee et al. (2013) provided the empirical foundation by showing that diffusion centrality predicted village-level microfinance adoption in Karnataka. The initial contacts were chosen by the microfinance institution, not by the researchers; the contribution was the ex post demonstration that seed centrality explains cross-village variation in take-up. The step from observational finding to experimental method was taken by Kim et al. (2015), who conducted an early randomised trial in which seeds were selected on the basis of their network position: in 32 villages in Honduras, they compared targeting individuals nominated as friends by random villagers (a technique that exploits the friendship paradox to reach high-centrality nodes without mapping the full network)

against targeting random individuals, and found significantly greater adoption of health behaviours under network-based targeting. Banerjee, Chandrasekhar, Duflo and Jackson (2019) compared alternative targeting strategies across 213 Indian villages and found that individuals nominated by other villagers as good information spreaders generated substantially more diffusion than formally designated leaders. Beaman, BenYishay, Magruder and Mobarak (2021) applied complex-contagion diffusion models to social network data from 200 villages in Malawi to identify optimal seed farmers for a new agricultural technology, and found that theory-based network targeting outperformed extension workers' selections. The conceptual resolution of the interference problem in network settings required a reformulation of the potential outcomes framework itself. Rather than indexing potential outcomes by own treatment alone, a unit's outcome must be allowed to depend on the entire treatment assignment vector $\mathbf{d} = (d_1, \ldots, d_n)$. This creates an immediate dimensionality problem: if all 2^n treatment vectors must be considered, the number of potential outcomes grows exponentially with the population, and standard estimands such as the average treatment effect are no longer well-defined without further restrictions.

The concept of an exposure mapping, formalised and developed by Aronow and Samii (2017), provides a tractable solution. An exposure mapping $f_i(\mathbf{d}, G)$ summarises the information in the full treatment vector $\mathbf{d}$ and the network G that is relevant for unit i's outcome, reducing the high-dimensional vector to a lower-dimensional exposure level. If the potential outcome $Y_i(\mathbf{d})$ depends on $\mathbf{d}$ only through $f_i(\mathbf{d}, G)$, a condition sometimes called stratified interference, then potential outcomes can be indexed by exposure levels rather than full treatment vectors, and standard causal estimands can be defined within each exposure stratum. For example, a common exposure summary records each unit's own treatment status together with the number or share of its treated neighbours, $\sum_j G_{ij} d_j$. This distinguishes units by own treatment and by neighbourhood treatment intensity, generating exposure conditions such as directly treated with high neighbour exposure, directly treated with low neighbour exposure, untreated with high neighbour exposure, and untreated with low neighbour exposure. Each contrast between exposure conditions defines a distinct causal estimand.

Within this framework, the literature distinguishes several effects that are well-defined even under interference. The direct effect compares the outcomes of treated and untreated units holding the exposure of their neighbours fixed, isolating the impact of own treatment net of spillovers. The spillover or indirect effect compares units with the same own treatment status but different levels of neighbourhood treatment, capturing the effect of peers' treatment on one's own outcome. The total effect combines both channels, comparing a unit that is treated and surrounded by treated peers with one that is untreated and surrounded by untreated peers. The overall average treatment effect averages these components over the population distribution of treatment and network position. Separating these estimands requires variation in both own treatment and neighbour treatment, which is why experimental design matters so greatly in network settings. Without independent variation in each component, the direct and spillover effects are not separately identified, and the researcher may

observe a compound effect whose interpretation depends on unstated assumptions about the network mechanism.

The exposure mapping approach imposes structure on the interference pattern rather than eliminating it, and this structure must be justified by the application. In a vaccination study, it may be plausible that only the share of directly connected contacts who are vaccinated matters for one's infection risk, making a simple neighbourhood saturation a credible exposure mapping. In a financial contagion study, higher-order network connections, such as the solvency of creditors of creditors, may be relevant, requiring a richer mapping that accounts for network distance and link weights. In a labour market study, information about job openings may travel through weak ties rather than strong ties, so the exposure mapping should reflect the character of the relevant connections rather than their raw count. When the exposure mapping is misspecified, because the researcher uses a simpler summary than the true interference pattern requires, the resulting estimands conflate distinct causal channels, and the analysis recovers neither the direct nor the spillover effect cleanly. Manski (2013) showed more generally that when the social interaction function is not fully known, the effects of treatment are only partially identified, and the identification region may be wide unless the researcher is willing to impose strong parametric or network-structural restrictions.

When considering causal effects under network interference, researchers need procedures that remain valid when outcomes of connected individuals are dependent. Two main approaches have been developed: randomisation-based inference (mainly coming from the experimental design), and asymptotic inference under network dependence, a more theory-driven approach.

Randomisation inference, which goes back to Fisher (1935), derives the distribution of a test statistic from the experimental design itself, by considering all assignments the randomisation could have produced. Under SUTVA, the null hypothesis of no effect fixes every potential outcome, and the only remaining randomness is the known assignment; when interference is present, a unit's outcome can change with others' assignments even if that unit's own treatment has no effect. Rosenbaum (2007) showed how to restore the approach by testing a narrower hypothesis, which he called no primary effects, stating that the treatment assigned to i does not affect i's own outcome regardless of what others receive. Under this null, distribution-free rank statistics still have a known randomisation distribution, so valid tests can be carried out without modelling the spillovers. The limitation is that this tests only own-treatment effects and says nothing about the size of spillovers. Athey, Eckles and Imbens (2018) went further by constructing what they called artificial experiments that convert broader hypotheses, such as the hypothesis that only immediate neighbours' treatments matter, into sharp nulls for which exact p-values can be computed from the design. Basse, Feller and Toulis (2019) extended this line of work through conditional randomisation tests that handle a wider range of interference structures.

On the estimation side, the challenge was to define and estimate separate direct and spillover effects. As mentioned previously, Hudgens and Halloran (2008) provided the foundational definitions, distinguishing direct, indirect, total, and overall effects under partial interference (meaning interference operates within pre-defined groups

but not between them), and Aronow and Samii (2017) generalised this by introducing exposure mappings. Sävje, Aronow and Hudgens (2021) later showed that consistent estimation of a generalised average treatment effect is possible even when the form of interference is unknown, as long as the interference is sufficiently local; their result covers several common experimental designs and provides reassurance that strong structural assumptions, while helpful for efficiency, are not always necessary for consistency.

A separate problem arises when the researcher needs standard errors and confidence intervals for network-dependent data, whether from an experiment or from observational work. The time-series HAC estimator of Newey and West (1987) and the spatial HAC estimator of Conley (1999), which assumes dependence decays with geographic distance, provided the starting points. Extending these to networks is harder because graph distance is discrete and irregular, and the dependence structure can vary significantly across nodes. Drawing on an older probability literature on dependency-graph central limit theorems, Kojevnikov, Marmer and Song (2021) proved limit theorems for network-dependent random variables and proposed a network HAC estimator that is consistent when dependence decays fast enough with graph distance. Leung (2022) connected this to causal inference by deriving the asymptotic properties of inverse-probability-weighted estimators of direct and spillover effects under what he called approximate neighbourhood interference, providing the inferential foundation for valid confidence intervals when interference propagates through a known network.

To conclude, the old practice of assuming no interference, and estimating average treatment effects as if potential outcomes depended only on own assignment, fared poorly in network contexts: it produced biased or ambiguous estimates of direct effects when spillovers were present, and it could also obscure the distinction between a policy's direct effect and the part operating through spillovers or network adjustment. The exposure mapping framework proved useful because it provided a principled language for distinguishing what the researcher wants to estimate from what a given design can identify. Direct and spillover effects reflect qualitatively different mechanisms with different implications for policy targeting, general equilibrium responses, and welfare analysis. Making this distinction explicit, and building it into the design of the study rather than treating it as an afterthought, became characteristic of credible causal inference on networks.

Part II: Big-Data and AI Era

We next turn to the big-data shift: as networks become larger and more richly measured, high-dimensionality issues and computational constraints come to the forefront; ML methods enter the toolkits for network studies.

16.6 Network Econometrics in the Big-Data Era

The methodological developments of the 1990s and 2000s were constrained, to a significant degree, by data availability. The network models reviewed in preceding sections were largely developed with specific data sets in mind: school friendship surveys such as Add Health, hand-collected firm-level transaction data, or stylised examples calibrated to known network structures. The constraint was not theoretical but empirical. Constructing a network requires observing bilateral links, and until the mid-2000s such data were rare, expensive to collect, and typically limited to a few hundred or thousand nodes. The emergence of large-scale digital data sources relaxed this constraint dramatically, expanding the set of economic questions that could be approached through a network lens and simultaneously exposing the limitations of methods designed for small, hand-collected graphs.

Administrative registers provided one of the most economically consequential new sources of network data. Matched employer–employee data from tax and social security records, increasingly available in many countries, enabled researchers to construct large-scale labour market networks in which nodes are workers or firms and links are employment relationships, co-worker exposures, or inter-firm mobility flows. Such data made it possible to study knowledge spillovers, wage determination, worker mobility, and the propagation of firm-level shocks across the labour market at a scale and precision that survey data could not support. Input–output tables and firm-level transaction data, increasingly available through administrative records, credit registries, and tax authorities, allowed the construction of production networks whose properties bore on questions of aggregate fluctuations and systemic risk. Acemoglu, Carvalho, Ozdaglar and Tahbaz-Salehi (2012) demonstrated theoretically and empirically that idiosyncratic sectoral shocks can aggregate into macroeconomic fluctuations when the input–output network is sufficiently asymmetric, and subsequent work exploited administrative production-network data to quantify these propagation effects more directly.

Online platforms generated network data of an entirely different character: massive in scale, continuously updated, and often covering social rather than purely economic links. Data from social-media platforms, professional networks, online marketplaces, and digital communication systems provided researchers with network structures involving millions, and sometimes hundreds of millions, of nodes, far exceeding the computational limits of classical network estimators. Chetty et al. (2022) used anonymised Facebook friendship data linked to administrative information to construct measures of social capital and document their association with economic mobility, illustrating both the analytic power of platform data and the methodological challenges of working with networks at this scale. Financial networks constituted a third source: interbank exposure matrices, payment-system records, derivatives counterparty data, and common asset-holding data gave researchers more direct access to the structure of financial interconnection that had long been theorised but was difficult to observe before the global financial crisis.

The expansion of data sources shifted the binding constraint in network econometrics from observation to computation, measurement, and inference. With networks of

millions of nodes, methods that require storing, inverting, or repeatedly transforming the full adjacency matrix, such as likelihood-based SAR estimators or standard network IV estimators, may become computationally infeasible. With links observed at high frequency across many time periods, static models of network structure may miss the dynamic adjustment of links to shocks. And with data generated by platforms whose coverage, algorithms, and user composition may change over time, measurement issues that were minor nuisances in small hand-collected networks become first-order concerns. The big-data era therefore did not simply give existing methods more observations to work with; it required new methods, new asymptotics, and new ways of thinking about what it means to estimate a network model at scale.

The lesson is that richer data did not remove the need for structure. A larger network is not automatically a better measured network. Online links may record attention rather than influence; observed transactions may omit informal relationships; platform data may reflect recommendation algorithms as much as user preferences; and financial exposures may be measured at a frequency or level of aggregation that misses relevant channels of contagion. The researcher still has to explain what a link means, why it represents the relevant economic channel, and how missing or mismeasured links affect inference.

The curse of dimensionality is a familiar problem in econometrics, but it takes a particularly acute form in network settings. An undirected network with n nodes has up to $n(n-1)/2$ potential links, so the number of unrestricted link-level objects grows quadratically in the number of nodes. Even for moderate n, this renders many classical estimation procedures infeasible: likelihood-based methods may require evaluating functions of the full adjacency matrix, moment-based methods may require estimating equations whose dimension scales rapidly with n, and nonparametric methods face a curse of dimensionality that worsens with network size. The response of the literature was to impose structure, most commonly through sparsity, low-dimensional latent representations, or restrictions on the relevant neighbourhood of each node.

Sparsity methods drawn from high-dimensional statistics provided a natural toolkit. In the context of Gaussian graphical models, the graphical LASSO of Friedman, Hastie and Tibshirani (2008) estimates the precision matrix of a multivariate distribution by maximising the Gaussian likelihood subject to an ℓ_1 penalty on off-diagonal entries, inducing sparsity in the estimated inverse covariance matrix and thereby recovering a sparse network of conditional dependencies. This approach has been attractive in applications involving financial return co-movements, macroeconomic interdependence, and systems of variables where the conditional-dependence network is itself an object of interest rather than merely a nuisance. The ℓ_1 penalty imposes a bias–variance trade-off that can be calibrated through cross-validation or information criteria, and the resulting sparse estimator can recover the underlying network structure under conditions on signal strength, degree growth, and dependence.

For network regression and peer-effects models, high dimensionality creates a different set of problems. If G is an adjacency matrix and X is a matrix of node characteristics, then GX, G^2X, G^3X, centrality measures, clustering measures, and local exposure variables are all functions of the same underlying graph. They may be highly collinear, and their interpretation may change with the density and

topology of the network. Higher-order instruments such as G^2X and G^3X, useful in principle for identifying peer effects, may become unstable in finite samples when the network is dense or when indirect exposures are strongly correlated with direct exposures. Similarly, models with many node-level fixed effects can absorb unobserved heterogeneity but may introduce incidental-parameter concerns when the number of links per node is limited.

Regularisation and dimension-reduction methods therefore became attractive for both practical and interpretive reasons. Penalised regression, sparse matrix methods, matrix completion, spectral methods, and community detection algorithms all reduce the dimensionality of the network object in different ways. In some applications, sparsity is imposed on the set of relevant neighbours: only a small subset of possible links is assumed to matter for a given outcome. In others, sparsity is imposed on parameters: among many possible network statistics, only a few are allowed to enter the model. Low-rank methods instead assume that much of the network structure can be summarised by a smaller number of latent factors, positions, or communities. These methods make large-scale estimation feasible, but they also require economic discipline. A sparse statistical approximation is useful only if the selected structure corresponds to a plausible economic mechanism.

The sparsity paradigm has endured because it formalised a restriction that is often plausible in economic networks: most agents are not directly connected to most other agents, and much of the relevant interaction structure can often be summarised without modelling the full n^2 matrix. What proved more fragile was the application of sparsity methods without verifying that the estimated network structure was stable across subsamples, robust to the choice of penalty parameter, or interpretable as the object of economic interest rather than a statistical artefact of the regularisation procedure. Regularisation methods produce sparse estimates by construction; the question is whether the estimated zeros correspond to genuinely absent links, or to links that are present but weak, noisy, or confounded. External validation, sensitivity analysis, and economic interpretation therefore remain indispensable.

The descriptive network metrics discussed previously compress a node's position into one or two scalars and discard most of the structural information in the adjacency matrix. Entering the full $n \times n$ matrix into a regression is infeasible when n is large. The embedding approach presents a middle path: represent each node as a vector in a low-dimensional space so that geometric relationships between vectors approximate structural relationships in the graph. Historically this idea can be traced back to Fiedler (1973), who showed that the second-smallest eigenvalue of the graph Laplacian, later called algebraic connectivity, carries information about graph connectivity; the associated eigenvector, often called the Fiedler vector, provides a natural basis for partitioning the graph. Shi and Malik (2000) and Ng, Jordan and Weiss (2001) turned this into practical spectral clustering algorithms. The step from spectral clustering to explicit node embedding was made by Belkin and Niyogi (2003), who proposed Laplacian eigenmaps: each node is mapped to a point in $\mathbb{R}^d$ by taking its entries in the d eigenvectors corresponding to the smallest nonzero eigenvalues of the graph Laplacian, chosen so that nodes connected by an edge are placed close together in the embedding space. This made the representation explicit and continuous, and enabled

the overall structure that would follow: define a function that would preserve some version of the concept of graph proximity, then optimise over a low-dimensional representation.

The machine-learning stage introduced a different class of methods that learn node representations by optimising a predictive objective rather than solving an eigenvalue problem. The key idea, proposed by Perozzi, Al-Rfou and Skiena (2014) under the name DeepWalk, is to generate many short random walks on the graph, treating each walk as a sentence in which the nodes are words, and then apply the skip-gram word-embedding technique of Mikolov, Sutskever, Chen, Corrado and Dean (2013), originally developed for natural-language text, to learn vector representations from these synthetic sentences. The skip-gram objective trains the vectors so that a node's embedding can predict which other nodes appear nearby in the random walks; as a result, nodes that tend to appear in similar walk contexts, because they share neighbours or occupy similar positions in the graph, receive similar vectors. Grover and Leskovec (2016) generalised this idea in node2vec by replacing the uniform random walk with a parameterised walk governed by two parameters, p and q, that control whether the walk tends to stay close to its starting node or venture further away. By varying these parameters, the researcher can control whether the embedding emphasises community proximity or role similarity, adapting the representation to the structure that matters for the application at hand. These methods produce dense, low-dimensional vectors that can serve as features in prediction, classification, or, from an econometric perspective, as covariates in regression models.

For econometrics, the appeal of learned embeddings is their ability to compress the adjacency matrix into a form that standard estimation methods can handle, capturing nonlinear aspects of network structure. The cost is, as is often the case, interpretability: the dimensions of a learned embedding do not correspond to named quantities, and it is difficult to say what a coefficient on a particular embedding dimension means. There is also a deeper concern about endogeneity, since an embedding computed from the same network that determines outcomes will generally be correlated with unobserved confounders, and using such embeddings as controls in a causal regression requires the same care as using any other endogenous covariate.

The econometric models discussed in the preceding sections treat the network as a single cross-sectional object. In many economic applications, however, links form and dissolve over time, and the network observed at any given moment is a snapshot of a continuing process. If the formation and dissolution of links depends on the current state of the network, for instance because existing connections introduce agents to new partners or because redundant ties are dropped, then the cross-sectional approach loses information and can introduce bias.

For longitudinal social-network data, one of the most influential modelling frameworks came from sociology. Snijders (2001) proposed the stochastic actor-oriented model (SAOM), in which the network evolves in continuous time: each actor receives random opportunities to add or drop a single outgoing tie, choosing the change that maximises an objective function that depends on the actor's current network position, characteristics of potential partners, and exogenous covariates. The model is estimated from network observations at two or more discrete time

points, and the objective function can incorporate reciprocity, transitivity, homophily, and degree-related effects. Snijders, van de Bunt and Steglich (2010) extended the framework to the co-evolution of the network and individual behaviour, allowing actors' outcomes to influence link formation and vice versa, which addresses, at least in principle, the simultaneous determination of influence and selection that follows the peer-effects literature (see above).

A parallel development adapted the exponential random graph framework to temporal settings. Hanneke, Fu and Xing (2010) proposed the temporal ERGM, modelling the network at time t as a function of lagged network statistics computed from time $t-1$. Krivitsky and Handcock (2014) introduced the separable temporal ERGM (STERGM), which decomposes the transition between consecutive snapshots into a formation process for new links and a dissolution process for existing links, each specified as a separate ERGM. This decomposition is economically natural: the costs and benefits of initiating a new relationship typically differ from those of maintaining an ongoing one. For time-stamped interaction data, such as sequences of emails, transactions, or communications, Butts (2008) proposed the relational event model, which models the rate at which events occur between pairs of actors as a function of the history of past events, avoiding the need to aggregate interactions into discrete snapshots.

From an econometric perspective, the dynamic setting makes the identification challenges already present in the static case even worse. The state dependence of the network, the fact that links at time t depend on links at time $t-1$, creates an initial-conditions problem similar to the one that arises in dynamic panel data models with individual effects, as studied in the classic contributions of Heckman (1981) and Wooldridge (2005). If the network at the first observed period is itself the outcome of an unobserved earlier process, conditioning on it may not remove the correlation between network dynamics and unobserved heterogeneity.

The practical importance of allowing for evolving networks has grown with the availability of high-frequency digital data (social media, mobile communication, financial transactions), where timestamped interactions are observed and precise time-varying links can be constructed. This creates new econometric challenges: the relevant time scale for link dynamics may differ from the time scale at which outcomes are measured, the network may have dynamics related to platform growth or policy changes, and the sheer scale requires estimation methods that can handle networks with millions of nodes.

A third dimension of the big-data challenge concerns inference rather than estimation. Classical econometric inference rests on laws of large numbers and central limit theorems that require observations to be sufficiently independent or weakly dependent. In network data, this requirement is rarely automatic. Two observations that share a network neighbour may not be independent, and in dense networks with short average path lengths, many pairs of observations can be connected within a few steps. The standard response of clustering standard errors at a pre-specified group level often fails to resolve the problem, because the appropriate dependence structure may be the network itself rather than a school, firm, region, or market.

The asymptotic theory appropriate for network-dependent data depends critically on whether the graph is sparse or dense as n grows. In a sparse network, where the average degree remains bounded or grows slowly with n, each node's neighbourhood is small relative to the full network, and observations at sufficient network distance may be approximately independent. Under suitable dependence restrictions, laws of large numbers and central limit theorems can then be established for network-dependent arrays. Kojevnikov et al. (2021) formalised this framework by developing limit theorems for network-dependent random variables and proposing variance estimators that account for network dependence rather than assuming independence or simple group clustering. In a dense network, where the average degree grows quickly with n, the conditions required for standard central limit approximations may fail, and the limiting behaviour of estimators can differ qualitatively from the sparse-graph case.

This sparse-versus-dense distinction matters for both estimation and identification. In sparse graphs, local variation in neighbours and neighbours of neighbours may provide useful identifying information, as in peer-effects models based on network intransitivity. In dense graphs, the distinction between direct and indirect exposure can blur, and aggregate shocks or common factors may be difficult to separate from network effects. Highly centralised graphs create another difficulty: if a few nodes account for a large share of connections, asymptotic arguments based on many small contributions may be inappropriate. Financial networks, platform networks, and production networks often have such hub structures, making scale both an opportunity and a source of inferential fragility.

The failure of naive i.i.d. inference in network settings has practical implications that extend beyond standard error estimation. Cross-validation procedures, routinely used for model selection and hyperparameter tuning in high-dimensional settings, can produce misleading out-of-sample performance estimates when observations are network-dependent. A model that predicts well for held-out nodes that remain connected to the training set may perform poorly when applied to genuinely new nodes, new subgraphs, or new network configurations, because the prediction exploits network dependence rather than stable structural relationships. Dependence-aware validation, such as holding out entire subgraphs, communities, time periods, or graph partitions rather than individual nodes, is a first step toward addressing this problem, but it remains less systematically developed in econometrics than in machine learning applications on graphs.

Scalable estimators that remain computationally tractable in large networks while preserving the economic interpretability of their parameters represent one of the more durable contributions of this period. Stochastic optimisation, sparse-matrix computation, subgraph sampling, and distributed computing approaches have all been adapted from computer science and machine learning to economic network problems. Their durability reflects a principle that mirrors the earlier history of spatial econometrics: methods that trade some statistical efficiency for computational feasibility tend to persist when the efficiency loss is small relative to the gain in applicability. The challenge is to develop inference procedures that are simultaneously graph-robust, computationally scalable, and grounded in the economic structure of

the problem, rather than inheriting defaults from either classical econometrics or algorithmic machine learning.

The big-data era made network econometrics both computational and inferential. Rich new network data expanded the set of feasible economics questions, but they moved the constraints from data availability to measurement, scalability, and dependence-aware inference. Regularisation and scalable estimation proved durable because many classical approaches became infeasible at modern scale. What proved less satisfactory was the use of high-dimensional summaries or regularised estimates without sufficient validation of their economic meaning, stability, and inferential consequences. Successful approaches combine scalable algorithms with economic transparency: they use computation to handle the size of the graph, but they still require a clear account of what the links mean, how dependence enters, and what variation supports the inferential claim.

16.7 AI & Machine Learning for Network Data

Machine learning (ML) and artificial intelligence (AI) enter network econometrics at the point where the network becomes too complex to handle in the forms required by standard econometric models. Some methods compress the graph into usable representations, as embeddings and graph neural networks do; others use the graph to improve prediction, to estimate high-dimensional nuisance functions in causal designs, or to construct the network itself from text and other unstructured sources. The cost is that the familiar econometric problems return in a less transparent form: black-box predictions, unstable network measurements, and the gap between predictive fit and causal or structural interpretation.

Probabilistic graphical models developed largely independently of network econometrics, but they are nonetheless important for the history of machine learning on networks because they made the graph into a device for representing a probability model. A probabilistic graphical model is a compact representation of a high-dimensional joint distribution, in which the nodes are random variables and the edges encode how the distribution factorises. The graph tells the researcher which variables must be conditioned on to separate one part of the system from another. In this sense, graphical models brought together three problems that became central to machine learning: how to represent dependence among many variables, how to perform inference or prediction without summing over large joint distribution, and how to learn unknown parameters or even the graph itself from data.

The origins technically go back to path analysis, introduced by Wright (1921) in genetics, where directed graphs represented causal pathways between variables. The modern literature formed around two forms of graphical model. On the directed side, Pearl (1988) systematised the theory of Bayesian networks: directed acyclic graphs in which each node is a random variable, the directed edges encode a factorisation of the joint distribution into conditional distributions, and conditional independencies can be read from the graph using the criterion of d-separation. Heckerman (1995) later gave a

machine-learning overview of Bayesian networks, emphasising their use for learning from data, combining prior knowledge with evidence, and handling uncertainty in prediction. On the undirected side, Besag (1974) developed Markov random fields for spatial data, where a variable is conditionally independent of all others given its neighbours. Dempster (1972) introduced covariance selection, showing that in a Gaussian setting zeros in the inverse covariance, or precision, matrix correspond to absent edges in the graph. This last connection has proved especially important for econometrics and finance: sparse precision estimation gives a structured way to recover conditional-dependence relations among many variables.

The machine-learning contribution was in treating these graphical representations as computational objects. Lauritzen and Spiegelhalter (1988) showed that inference in graphical models can be organised through local computations on the graph, so that beliefs are updated by passing information between neighbouring parts of the model rather than by enumerating the full joint distribution. Buntine (1994) presented graphical models as a general framework for empirical learning, showing how algorithms such as expectation maximisation and Gibbs sampling could be understood through operations on a graphical specification. Jordan (1998) then placed graphical models at the centre of probabilistic machine learning, as a language connecting statistics, artificial intelligence, and computation; Koller and Friedman (2009) later gave the canonical textbook treatment of representation, inference, and learning in probabilistic graphical models. In the high-dimensional Gaussian case, as described above, Friedman et al. (2008) made covariance selection computationally practical through the graphical lasso, which estimates a sparse precision matrix by applying an ℓ_1 penalty to its off-diagonal entries, thereby inducing sparsity in the estimated conditional-dependence graph.

The link between these models and the economic networks discussed in this chapter operates at two levels. At the methodological level, the methods from the graphical-models literature provide computational tools for evaluating likelihoods and estimating dependence structures in models with many connected variables. When an econometric model specifies a full joint distribution in which each unit's outcome depends on neighbouring outcomes or shocks, these tools can apply directly. At the analogical level, many influential economic models use graph structure to represent channels of dependence without adopting the full PGM formalism. Eisenberg and Noe (2001) modelled the clearing of obligations in a financial network as a deterministic fixed-point problem on a directed liability graph, showing how the failure of one institution can cascade to others. Diebold and Yılmaz (2014) used variance decompositions from vector autoregressions to construct measures of directional connectedness among financial institutions, interpreting the resulting spillover structure as a time-varying weighted directed graph. Acemoglu, Ozdaglar and Tahbaz-Salehi (2015) studied how the architecture of the financial network determines whether small shocks remain localised or amplify into systemic crises. None of these is a probabilistic graphical model in the technical sense of Pearl or Lauritzen, but all share with the graphical-dependence tradition the idea that the graph organises joint outcomes, rather than merely supplying covariates for a peer-effects regression.

The random-walk embeddings discussed in the preceding subsection learn vector representations from network structure and are usually trained independently of the downstream econometric task. Graph neural networks (GNNs) go further: they learn to transform node features by aggregating information from neighbours in the graph, producing representations that are informed both by the node's own attributes and by the structure of its local network environment.

The neural-network foundations are the backpropagation algorithm of Rumelhart, Hinton and Williams (1986), which made gradient-based training of multi-layer networks practical, and the convolutional neural networks of LeCun, Bottou, Bengio and Haffner (1998), which showed that constraining a neural network to aggregate information locally, with shared weights across spatial positions, produces powerful representations for grid-structured data such as images. GNNs generalise this local-aggregation idea from regular grids to irregular graphs. Early formulations by Gori, Monfardini and Scarselli (2005) and Scarselli, Gori, Tsoi, Hagenbuchner and Monfardini (2009) defined graph neural networks as recurrent systems in which each node's hidden state is updated from the states of its neighbours and iterated until a stable representation is reached. These models made the graph itself part of the neural architecture, but they were computationally demanding and less directly compatible with the deep-learning toolkits that later became standard. Li, Tarlow, Brockschmidt and Zemel (2016) modernised this approach through gated graph sequence neural networks, replacing the older contraction-mapping formulation with gated recurrent updates trained by backpropagation.

A second line of development came from spectral graph convolution. Bruna, Zaremba, Szlam and LeCun (2014) defined convolution on graphs through the eigenvectors of the graph Laplacian, in analogy with the Fourier transform on regular grids. Defferrard, Bresson and Vandergheynst (2016) made this approach more practical by approximating spectral filters with Chebyshev polynomials, avoiding the need to compute the full eigendecomposition. The architecture that made graph convolution widely used was the graph convolutional network of Kipf and Welling (2017), which reduced the spectral construction to a simple layer-wise propagation rule. At each layer, the GCN combines a node's own features with those of its neighbours, normalises this information by node degree, and applies learned weights and a nonlinear transformation. In the linear spatial-lag model, to bring up a familiar by now example, the network matrix enters a simultaneous-equations system and the spatial autoregressive coefficient governs equilibrium feedback. In a GCN, by contrast, multiplication by the normalised adjacency matrix is a feed-forward smoothing operation applied layer by layer. Still, repeated aggregation through the normalised adjacency matrix plays a role similar to repeated spatial or network averaging, while the learned weight matrices and nonlinearities allow neighbouring information to affect different feature dimensions in different ways. In this limited sense, GNNs can be read as nonlinear, multi-layer successors to the linear network-smoothing operations used in spatial and peer-effects models.

Gilmer, Schoenholz, Riley, Vinyals and Dahl (2017) unified many of these architectures under the message-passing neural network framework. In this formulation, each node sends a message to its neighbours, aggregates the incoming messages, and

updates its own representation on the basis of the aggregated information. Different choices of message, aggregation, and update functions recover many of the main GNN variants. GraphSAGE, introduced by Hamilton, Ying and Leskovec (2017), was especially important for large economic and platform settings because it learned inductive aggregation rules that could be applied to nodes not observed during training. Graph attention networks Veličković et al., 2018 replaced uniform or degree-normalised averaging with learned attention weights, allowing the model to place more weight on some neighbours than on others. Wu et al. (2020) provide a comprehensive survey of GNN architectures and applications.

In economics and adjacent applied fields, GNNs are most useful where node features and network structure both carry predictive information. Link prediction and recommendation systems are natural cases, since both users and various providers often form bipartite or multi-sided graphs; Ying et al. (2018), for example, used graph convolutional methods for web-scale recommendation. Fraud detection in financial transaction networks is another common application, since suspicious behaviour may be visible not only in the features of an account or transaction but also in its connections to other accounts. Similar methods have begun to appear in credit-risk assessment, supply-chain disruption prediction, and matching platforms, though in many of these areas the econometric literature is still less settled than the predictive machine-learning literature. The econometric concern is familiar. A GNN can predict which firms will default or which transactions are suspicious, but the learned representation does not by itself identify the economic mechanism driving the prediction or the effect of changing the network - the aforementioned black-box problem.

The machine learning methods that entered empirical microeconomics in the 2010s were developed primarily for settings in which observations are independent or, at most, weakly dependent in ways that standard clustering or time-series corrections can address. Causal forests (Wager & Athey, 2018), double and debiased machine learning (Chernozhukov et al., 2018), and related doubly robust estimators adapted flexible prediction algorithms to the task of estimating heterogeneous treatment effects and average causal effects under the potential outcomes framework. Their common foundation is a combination of sample splitting, cross-fitting, and nonparametric nuisance estimation that allows causal quantities to be estimated while reducing the first-order impact of regularisation and model-selection bias. In settings where SUTVA holds and units are exchangeable, these procedures are well understood and their asymptotic properties are established.

Network settings violate both premises. When treatment propagates through connections, the potential outcome of each unit may depend on the treatment of its neighbours, its neighbours' neighbours, and, in principle, the entire assignment vector. This invalidates the standard definition of the average treatment effect and requires additional structure: either a restriction on the interference pattern, such as the stratified interference assumption underlying exposure mappings, or a partial-identification approach that characterises the range of effects consistent with the observed data. Causal machine learning methods must therefore be adapted before they can be applied credibly in network settings.

The first challenge is the definition of the estimand. Causal forests and double machine learning typically target the conditional average treatment effect or the average treatment effect under SUTVA. In the presence of network interference, these estimands are replaced by a family of exposure-conditional effects: the direct effect of treatment given a level of neighbourhood exposure, the spillover effect given own treatment status, and various total and overall effects that aggregate across the network distribution of treatment and position. Estimating these exposure-conditional effects requires not only flexible modelling of the outcome as a function of treatment and covariates, but also a credible exposure mapping that summarises the relevant neighbourhood treatment in a lower-dimensional form. The machine learning component can then be used to estimate heterogeneous effects across node characteristics, network positions, or exposure levels, extending the causal forest logic to the richer space of exposures defined by the network structure.

This adaptation requires estimating exposure-aware nuisance functions. Here, they describe expected outcomes and exposure probabilities conditional on observed pre-treatment individual and network characteristics. Machine learning allows these relationships to be estimated flexibly. Direct and spillover effects can then be represented as contrasts between predicted outcomes under different exposure states. This formulation makes clear that machine learning is useful for flexible nuisance estimation and heterogeneity analysis, but it does not replace the identifying assumptions that justify the exposure mapping or the treatment assignment mechanism.

The second challenge is sample splitting and cross-fitting under network dependence. The standard cross-fitting procedure in double machine learning randomly partitions the sample into folds and estimates nuisance functions on complementary folds, relying on the approximate independence of observations across folds to control the bias of the plug-in estimator. When units are connected through a network, randomly sampled folds may contain many pairs of observations that share neighbours, so that the nuisance function estimated on one fold is correlated with the outcome on the complementary fold through network links rather than through the structural parameters of interest. This correlation can reintroduce bias into the cross-fitted estimator and may invalidate the usual Neyman-orthogonality argument that underlies $\sqrt{n}$-consistent inference. Network-aware sample splitting, in which folds are defined by subgraphs, graph partitions, or sufficiently separated neighbourhoods rather than random node assignments, is one response to this problem. Athey et al. (2018) showed more broadly that inference under network interference requires careful treatment of the dependence structure, and that standard permutation-based or asymptotic tests need to be modified to account for the way treatment propagates across the graph.

The third challenge is inference on estimated causal effects. Even if the estimand is well defined and the nuisance functions are estimated consistently, the variance of the causal effect estimator under network interference is more complex than in the i.i.d. case. The influence function of a doubly robust estimator under exposure-mapping-based SUTVA contains terms that reflect the correlation of potential outcomes across connected units, and computing the asymptotic variance requires a network-robust variance estimator of the kind discussed in the previous section. Naive plug-in variance estimators that ignore network dependence can understate uncertainty, producing

confidence intervals that are too narrow and tests that reject too often. Leung (2022) developed a framework for causal inference under approximate neighbourhood interference, formalising conditions under which potential outcomes depend only on the treatment of nearby nodes and deriving limit theorems and inference procedures that are valid under this restriction. This framework connects the interference literature directly to the network-dependence literature, making clear that the two problems, what to estimate and how to make valid inferences, are linked through the structure of the network.

These adaptations represent genuine progress, but they also reveal a recurring tension in the application of machine learning to network data. The flexibility of causal machine learning is most valuable when the heterogeneity of treatment effects across units is substantial and when the functional form of the outcome model is unknown. Both conditions are plausible in network settings: the effect of a peer's treatment on one's own outcome may depend on the strength of the link, the centrality of the treated peer, the density of the local neighbourhood, and individual characteristics that interact with these features. But realising this potential requires specifying and validating the exposure mapping, which is where the machine learning component interacts most directly with economic theory. A causal forest that conditions on a misspecified exposure summary may estimate a stable predictive contrast, but not the causal object of economic interest. The interpretability problem that afflicts black-box machine learning methods more broadly is therefore compounded in network settings by the prior requirement that the interference structure be correctly specified.

Causal machine learning under network interference illustrates the same pattern: scalable algorithms and flexible function approximation help with the complexity of network data, but they do not replace structural thinking about what a link means and how treatment propagates through it. Good prediction of outcomes on a network, controlling for neighbourhood treatment, does not imply that the estimated quantities have a causal interpretation; that interpretation requires an exposure mapping that is credible, a source of variation that is plausibly exogenous, and inference procedures that are valid under the dependence structure of the observed graph. Making these requirements explicit, and building them into the design and analysis of causal machine learning on networks, is what separates the credibility from the predictive accuracy that machine learning optimises.

The network data used throughout this chapter, including friendship nominations, interbank exposures, trade flows, and village-survey links, were collected through surveys, administrative records, or direct observation. A more recent development constructs economic networks from text, using natural-language processing (NLP) and, increasingly, Large Language Models (LLMs) to extract relationships from documents that were not designed to record network data. Because LLMs are rapidly changing what is measurable from text, and because the resulting measurements increasingly serve as inputs to econometric models, this subsection traces the development from early text-based networks through the current state of LLM-enabled extraction and the econometric issues it raises.

The idea of building economic networks from text predates large language models. Hoberg and Phillips (2016) constructed time-varying product-market networks

among US firms by computing cosine similarities of the product-description sections of annual 10-K filings. The resulting text-based network industry classifications (TNIC) predicted managerial assessments of competition better than fixed SIC or NAICS codes and established the principle that text can generate economically meaningful, time-varying adjacency structures. Wichmann, Brintrup, Baker, Woodall and McFarlane (2020) took the next step by training deep neural networks on labelled sentence-level data to extract buyer-supplier relations from news articles, framing the problem as a supervised relation-extraction task. Both methods required either hand-crafted features (bag-of-words similarity) or substantial labelled training data, and their accuracy was limited by the representations available at the time. Gentzkow, Kelly and Taddy (2019) surveyed the broader text-as-data programme in economics, covering methods from word counts through topic models and word embeddings.

The computer-science developments that changed what is feasible came in rapid succession. The transformer architecture of Vaswani et al. (2017) introduced the self-attention mechanism that allows a model to weight the relevance of every word in a sequence to every other word, enabling the capture of long-range dependencies that earlier recurrent and convolutional architectures struggled with. Devlin, Chang, Lee and Toutanova (2019) showed that pre-training a transformer on large text corpora and then fine-tuning it on a specific task, an approach they called BERT, produces representations that substantially improve performance on named-entity recognition and relation extraction. For social science, an important empirical finding was that LLMs can, on some annotation tasks, match or outperform human coders and crowd workers, although performance varies. Gilardi, Alizadeh and Kubli (2023) showed that ChatGPT's zero-shot accuracy exceeded that of crowd workers on several text-annotation tasks including relevance detection and stance classification, at a fraction of the cost. Törnberg (2024) found that GPT-4 outperformed expert coders and supervised classifiers when classifying the political affiliation of US politicians from social-media messages. These results suggest that LLMs can serve as scalable, low-cost measurement instruments for constructing variables from text, including the relational variables needed to build network data, though reliability and reproducibility across tasks and model versions remain active concerns.

In economics, the use of machine-learning-generated measurements has expanded rapidly beyond network construction per se. Ludwig and Mullainathan (2024) argued that machine learning can support hypothesis generation by producing new measurements and patterns from high-dimensional data. Ludwig, Mullainathan and Rambachan (2025) proposed a formal econometric framework for integrating LLM-generated variables into causal and predictive analyses, treating the LLM as a fallible measurement device whose error structure must be characterised and accounted for in downstream inference. Their key point is that high average classification accuracy is not sufficient: if measurement errors are correlated with treatment status or covariates, downstream estimates can remain substantially biased, even when the annotation task appears accurate on average. For text-based network construction specifically, LLMs extend the earlier relation-extraction literature by allowing zero-shot and few-shot identification of entities and relationships in text without domain-specific labelled

data, though whether the resulting networks improve on supervised extraction or commercial databases remains an open empirical question.

The econometric issues accompanying LLM-constructed networks are significant and mostly unresolved. Text-extracted networks are measured with error, and the error structure is unlike classical measurement error in a covariate: the adjacency matrix determines the dependence structure of the entire model, so false positives and false negatives propagate through every regression that conditions on it. LLMs hallucinate, generating plausible-sounding relationships not present in the source text, at rates that vary widely with the model and the domain. The text sources themselves are not a random sample of economic relationships: frequently mentioned firms are overrepresented, and the model's training corpus may embed biases about which relationships are important. Temporal alignment between a textual mention and the economic relationship it describes is often unclear: a news article reporting that firm A supplies firm B may describe a relationship that began years earlier or has already ended. Reproducibility is fragile and very much debatable across all LLM-related research: the output of a proprietary LLM can change across model versions and API updates in ways the researcher cannot fully document. These concerns place LLM-constructed networks squarely within the classical econometric worry about data quality, and they connect to the broader measurement agenda that Ludwig et al. (2025) formalise. Validation against ground-truth network data remains essential and rare, and the development of econometric methods for inference with machine-generated network data is likely to become a substantial research agenda in the years ahead.

Part III: Economics & Interpretation

We close by showing how methodological developments affect the economics-of-networks research areas; what we can learn about welfare, diffusion, and systemic risk once econometrics handles formation and spillovers.

16.8 Connections to the Economics of Networks

The methods reviewed in this chapter have been developed or adapted from several disciplines to address economic questions about how networks affect welfare, inequality, diffusion, and systemic risk (see Jackson, Rogers & Zenou, 2017 for a comprehensive survey of these economic consequences). This section draws those connections explicitly. Many of the papers mentioned here have already appeared in earlier sections; the purpose of revisiting them is to show how the methods and the economics fit together.

The question of whether the networks that agents form are socially efficient, and if not, how a planner should intervene, has been central since Jackson and Wolinsky (1996) showed that pairwise-stable networks can diverge from efficient ones because agents do not internalise the externalities their links impose on others. Ballester et al. (2006) gave this a precise policy form: in linear-quadratic network games, the player

whose removal reduces aggregate activity the most is identified by an intercentrality measure derived from the game's payoff structure, giving the planner a network-based targeting rule. Galeotti, Golub and Goyal (2020) generalised the targeting problem to budget-constrained interventions, showing that optimal policy decomposes into the principal components of the adjacency matrix, with the direction of targeting depending on whether interactions are complementary or substitutable.

Network structure also shapes the distribution of economic opportunity. Granovetter (1973) argued that weak ties, acquaintances rather than close friends, are disproportionately important for accessing novel information because they bridge otherwise disconnected parts of the social structure. Calvó-Armengol and Jackson (2004) formalised this in a model where agents learn about job vacancies through social contacts, showing that small initial differences in employment status between groups with separate networks can produce persistent inequality through a referral mechanism. Currarini et al. (2009) showed that such segregation can itself arise endogenously from the interaction of preferences, group size, and meeting opportunities, with the resulting patterns of homophily differing systematically between majority and minority groups. At a much larger empirical scale, Chetty et al. (2022) used anonymised data from Facebook covering 21 billion friendships to construct measures of social capital across US communities and found that economic connectedness, the share of high-socioeconomic-status friends among low-socioeconomic-status individuals, is among the strongest predictors of upward income mobility in their analysis.

Diffusion of information, technology, behavioral beliefs and norms through networks has motivated some of the most productive interactions between economic theory and empirical research. The theoretical models range from threshold models of local interaction Morris, 2000 to the influence-maximisation framework of Kempe, Kleinberg and Tardos (2003). The empirical studies discussed earlier gave these ideas concrete content: Banerjee et al. (2013) showed that diffusion centrality of initial seeds predicts microfinance take-up; Conley and Udry (2010) showed that farmers adjust inputs toward those of successful information neighbours; and the seed-targeting experiments of Kim et al. (2015), Banerjee et al. (2019), and Beaman et al. (2021) demonstrated that network-informed targeting outperforms alternative selection methods.

The aggregation of information through networks raises a distinct set of questions. DeGroot (1974) proposed a model in which agents update beliefs by averaging over their neighbours' beliefs. Bala and Goyal (1998) studied a richer environment in which agents learn from their own outcomes and those observed among neighbours, showing how the structure of local observation affects whether behaviour converges to the superior action or settles on an inferior one. Golub and Jackson (2010) showed that such naïve updating aggregates information efficiently if and only if the influence of the most influential agent vanishes as the network grows, a condition that fails in networks with high-degree opinion leaders. This result connects network structure directly to collective decision-making, with implications for the design of communication structures in organisations and markets (see Golub & Sadler, 2016 for a survey).

Contagion and systemic risk in financial and production networks constitute another area where network architecture proves decisive. Allen and Gale (2000) showed that incomplete interbank structures can be more vulnerable to contagion than complete ones. Elliott, Golub and Jackson (2014) showed that diversification and integration have distinct and non-monotonic effects on systemic risk. Acemoglu et al. (2015) showed that dense connections promote risk-sharing under small shocks but amplify large ones. On the production side, Acemoglu et al. (2012) showed that idiosyncratic shocks to hub sectors can propagate through the network and generate aggregate volatility, and Barrot and Sauvagnat (2016) and Carvalho, Nirei, Saito and Tahbaz-Salehi (2021) provided firm-level empirical evidence of downstream propagation following natural disasters.

The common thread across these applications is that the network is a significant economic object whose structure determines welfare, opportunity distribution, information transmission, and risk propagation. The methods reviewed in this chapter have been used to estimate and test these mechanisms, while more recent graph-learning methods have expanded the range and scale of predictive applications. The extent to which this has succeeded, and the extent to which it remains limited by the measurement and interpretability challenges documented throughout the chapter, is the subject of the general lessons that follow.

16.9 General Lessons

Several principles emerge from the history traced in this chapter. The most durable contributions share three features: a clearly stated identification strategy, an explicit connection between the econometric specification and an economic mechanism, and a demonstrated capacity to scale beyond the original application. The linear-in-means model, despite its well-known identification limitations, proved durable precisely because it stated those limitations clearly and thereby organised the subsequent literature. The identification result of Bramoullé et al. (2009) survived because it gave the structure of the network a direct role in identification rather than treating it as a background object. Strategic formation models in the tradition of Jackson and Wolinsky (1996) have remained influential because they connected the network structure to the incentives that generated it, making the adjacency matrix an equilibrium object rather than a maintained assumption. Experimental work under interference, from the two-stage randomisation framework of Hudgens and Halloran (2008) to the seed-targeting trials of Banerjee et al. (2019) and Beaman et al. (2021), proved itself because it embedded the interference structure into the design itself rather than treating spillovers as a nuisance to be controlled away.

What aged less well falls into two broad categories. The first is the use of flexible statistical models without adequate diagnostics or structural interpretation. Complex ERGM specifications that reproduced observed network statistics but suffered from degeneracy, weak identification, or instability across nearby parameterisations are a key example of this pattern: fitting the topology of a network is not the same as

explaining why it formed. The second is the application of black-box machine learning methods to network data without addressing the gap between predictive accuracy and causal or structural content. A graph neural network can predict which firm will default or which transaction is fraudulent, but the learned representation does not by itself identify the mechanism or the counterfactual consequences of changing the network. This tension, again, is not unique to networks, but it is exacerbated here due to the all-encompassing role of networks in the related research questions.

Three main lessons can be taken from this. First, measurement and validation are crucial parts of any econometric analysis; whether the network is constructed from survey nominations, administrative records, or LLM-extracted text, the quality of the adjacency matrix affects every subsequent estimate. Second, the structure of the network is itself an economic object. Methods that exploit network features for identification have been most convincing when the features used correspond to a well-specified economic channel. Third, as echoed across most machine learning related advances, scalability without interpretability is insufficient. The computational tools that make large-network estimation feasible are valuable, but they do not substitute for a clear account of what the links mean or how dependence enters. The methods that have survived and will likely remain in active use are those that combine all three: credible measurement, economic structure, and the capacity to handle the data at hand.

16.10 Looking Ahead

Several developments are likely to shape network econometrics over the coming decade, each following directly from unresolved tensions documented in this chapter.

Privacy and data access present an immediate challenge. Many of the richest network datasets, including social-media graphs, are subject to growing legal and institutional restrictions. Differential privacy and synthetic-data methods can help, but they can distort the adjacency matrix or the statistics computed from it. Secure multi-party computation takes a different approach, allowing parties to compute jointly without sharing the underlying links, though at considerable institutional and computational cost. How to do valid econometric inference when the network has been filtered through such procedures is largely an open question.

A related frontier is the analysis of dynamic and high-frequency networks. Most of the models reviewed in this chapter treat the network as a single cross-sectional object, or at best observe it at a handful of discrete points. High-frequency digital data record timestamped interactions from which evolving link structures can be constructed, but exploiting this granularity requires methods that can handle time-varying adjacency at scale. The familiar problems of initial conditions, state dependence, and simultaneity from dynamic panel data all have natural analogues in the networks settings as well. Progress is likely to draw on longitudinal network models such as TERGMs and SAOMs alongside event-based approaches like relational-event and point-process models.

The relationship between graph-learning methods and econometric structure also remains incomplete. Graph neural networks and embeddings are effective at representing network structure, but they are typically trained to optimise predictive objectives that have no direct causal or welfare interpretation. Using these architectures to estimate nuisance functions within a causally grounded framework is a natural next step, and early work reviewed in Section 16.7 points in this direction, but the inferential theory is still under development.

Finally, the construction of economic networks from unstructured text using large language models is developing rapidly, but the econometric framework for using such data remains at early stages. LLM-extracted links come with hallucinations, systematic biases, and instability across model versions, all of which feed into every downstream estimate that conditions on the network. The framework of Ludwig et al. (2025), which treats LLM output as fallible measurement requiring validation data, offers a starting point, but its application to network data needs further development. Careful validation against ground-truth networks, where available, and sensitivity analysis across models and prompts will be necessary before LLM-constructed networks can be treated as reliable econometric inputs.

16.11 Conclusions

This chapter has traced the development of network econometrics from its origins in spatial dependence models and the social interactions framework of the 1970s–1990s, through the emergence of peer-effects models on observed networks and strategic formation models in the 2000s, to the big-data and machine-learning methods of the current period. Two broad lessons organise the narrative.

The first is that network structure has become a source of both economic content and econometric variation. The matrix representation of cross-unit dependence, used in spatial econometrics to encode geographic proximity through a weights matrix, has been reinterpreted in network econometrics as a record of economic and social relationships whose structure carries identifying information. The second is that each advance in data and methods has reintroduced, in new form, the classical econometric problems of identification, measurement, and model specification. The scalability and predictive power of machine learning expanded the feasible set of network applications, but the gap between predictive fit and causal interpretation widened as well. The construction of networks from text by large language models has created new data sources at unprecedented scale, but with measurement-error properties that are not yet well characterised.

New data and methods consistently open new questions, and new questions create new econometric difficulties. The methods that have proved most durable are those that confront this cycle directly; while network econometrics has matured considerably over the past three decades, the cycle is unlikely to end. As networks become larger, more dynamic, and constructed from increasingly diverse sources, the demand for

econometric methods that are simultaneously scalable and interpretable will only grow.

References

Acemoglu, D., Carvalho, V. M., Ozdaglar, A. & Tahbaz-Salehi, A. (2012). The network origins of aggregate fluctuations. *Econometrica*, *80*(5), 1977–2016.

Acemoglu, D., Ozdaglar, A. & Tahbaz-Salehi, A. (2015). Systemic risk and stability in financial networks. *American Economic Review*, *105*(2), 564–608.

Allen, F. & Gale, D. (2000). Financial contagion. *Journal of Political Economy*, *108*(1), 1–33.

Angelucci, M. & De Giorgi, G. (2009). Indirect effects of an aid program: How do cash transfers affect ineligibles' consumption? *American Economic Review*, *99*(1), 486–508.

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, *30*, 98–108.

Anselin, L. (1988). *Spatial econometrics: Methods and models*. Dordrecht: Kluwer Academic Publishers.

Anselin, L. & Bera, A. K. (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah & D. E. A. Giles (Eds.), *Handbook of applied economic statistics* (pp. 237–289). New York: Marcel Dekker.

Aronow, P. M. & Samii, C. (2017). Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*, *11*(4), 1912–1947.

Athey, S., Eckles, D. & Imbens, G. W. (2018). Exact p-values for network interference. *Journal of the American Statistical Association*, *113*(521), 230–240.

Aumann, R. J. & Myerson, R. B. (1988). Endogenous formation of links between players and of coalitions: An application of the Shapley value. In A. E. Roth (Ed.), *The shapley value: Essays in honor of lloyd s. shapley* (pp. 175–191). Cambridge: Cambridge University Press.

Backstrom, L. & Kleinberg, J. (2011). Network bucket testing. In *Proceedings of the 20th international conference on world wide web* (pp. 615–624). New York: ACM.

Baird, S., Bohren, J. A., McIntosh, C. & Özler, B. (2018). Optimal design of experiments in the presence of interference. *Review of Economics and Statistics*, *100*(5), 844–860.

Bala, V. & Goyal, S. (1998). Learning from neighbours. *Review of Economic Studies*, *65*(3), 595–621.

Bala, V. & Goyal, S. (2000). A noncooperative model of network formation. *Econometrica*, *68*(5), 1181–1229.

Ballester, C., Calvó-Armengol, A. & Zenou, Y. (2006). Who's who in networks. Wanted: The key player. *Econometrica*, *74*(5), 1403–1417.

Banerjee, A., Chandrasekhar, A. G., Duflo, E. & Jackson, M. O. (2013). The diffusion of microfinance. *Science*, *341*(6144), 1236498.

Banerjee, A., Chandrasekhar, A. G., Duflo, E. & Jackson, M. O. (2019). Using gossips to spread information: Theory and evidence from two randomized controlled trials. *Review of Economic Studies*, *86*(6), 2453–2490.

Barabási, A.-L. & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512.

Barrot, J.-N. & Sauvagnat, J. (2016). Input specificity and the propagation of idiosyncratic shocks in production networks. *Quarterly Journal of Economics*, *131*(3), 1543–1592.

Basse, G. W., Feller, A. & Toulis, P. (2019). Randomization tests of causal effects under interference. *Biometrika*, *106*(2), 487–494.

Bavelas, A. (1950). Communication patterns in task-oriented groups. *Journal of the Acoustical Society of America*, *22*(6), 725–730.

Beaman, L., BenYishay, A., Magruder, J. & Mobarak, A. M. (2021). Can network theory-based targeting increase technology adoption? *American Economic Review*, *111*(6), 1918–1943.

Belkin, M. & Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, *15*(6), 1373–1396.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B*, *36*(2), 192–236.

Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician*, *24*(3), 179–195.

Bloch, F., Jackson, M. O. & Tebaldi, P. (2023). Centrality measures in networks. *Social Choice and Welfare*, *61*(2), 413–453.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, *2*(1), 113–120.

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, *92*(5), 1170–1182.

Boss, M., Elsinger, H., Summer, M. & Thurner, S. (2004). Network topology of the interbank market. *Quantitative Finance*, *4*(6), 677–684.

Bramoullé, Y., Djebbari, H. & Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, *150*(1), 41–55.

Bruna, J., Zaremba, W., Szlam, A. & LeCun, Y. (2014). Spectral networks and locally connected networks on graphs. In *Proceedings of the 2nd international conference on learning representations.*

Buntine, W. L. (1994). Operations for learning with graphical models. *Journal of Artificial Intelligence Research*, *2*, 159–225.

Butts, C. T. (2008). A relational event framework for social action. *Sociological Methodology*, *38*(1), 155–200.

Calvó-Armengol, A. & Jackson, M. O. (2004). The effects of social networks on employment and inequality. *American Economic Review*, *94*(3), 426–454.

Calvó-Armengol, A., Patacchini, E. & Zenou, Y. (2009). Peer effects and social networks in education. *The Review of Economic Studies*, *76*(4), 1239–1267.

Carrell, S. E., Sacerdote, B. I. & West, J. E. (2013). From natural variation to optimal policy? The importance of endogenous peer group formation. *Econometrica*, *81*(3), 855–882.

Carvalho, V. M., Nirei, M., Saito, Y. U. & Tahbaz-Salehi, A. (2021). Supply chain disruptions: Evidence from the Great East Japan Earthquake. *Quarterly Journal of Economics*, *136*(2), 1255–1321.

Case, A. C. (1991). Spatial patterns in household demand. *Econometrica*, *59*(4), 953–965.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, *47*(1), 225–238.

Chatterjee, S. & Diaconis, P. (2013). Estimating and understanding exponential random graph models. *Annals of Statistics*, *41*(5), 2428–2461.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, *21*(1), C1–C68.

Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., ... Wernerfelt, N. (2022). Social capital I: Measurement and associations with economic mobility. *Nature*, *608*(7921), 108–121.

Cliff, A. D. & Ord, J. K. (1973). *Spatial autocorrelation*. London: Pion.

Cliff, A. D. & Ord, J. K. (1981). *Spatial processes: Models and applications*. London: Pion.

Conley, T. G. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics*, *92*(1), 1–45.

Conley, T. G. & Udry, C. R. (2010). Learning about a new technology: Pineapple in ghana. *American Economic Review*, *100*(1), 35–69.

Cornfield, J. (1978). Randomization by group: A formal analysis. *American Journal of Epidemiology*, *108*(2), 100–102.

Cox, D. R. (1958). *Planning of experiments*. New York: Wiley.

Crépon, B., Duflo, E., Gurgand, M., Rathelot, R. & Zamora, P. (2013). Do labor market policies have displacement effects? Evidence from a clustered randomized experiment. *Quarterly Journal of Economics*, *128*(2), 531–580.

Currarini, S., Jackson, M. O. & Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, *77*(4), 1003–1045.

Defferrard, M., Bresson, X. & Vandergheynst, P. (2016). Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in neural information processing systems* (Vol. 29).

DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, *69*(345), 118–121.

Dempster, A. P. (1972). Covariance selection. *Biometrics*, *28*(1), 157–175.

de Paula, A., Richards-Shubik, S. & Tamer, E. (2018). Identifying preferences in networks with bounded degrees. *Econometrica*, *86*(1), 263–288.

Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the north american chapter of the association for computational linguistics* (pp. 4171–4186).

Diebold, F. X. & Yılmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, *182*(1), 119–134.

Eckles, D., Karrer, B. & Ugander, J. (2017). Design and analysis of experiments in networks: Reducing bias from interference. *Journal of Causal Inference*, *5*(1), 20150021.

Eisenberg, L. & Noe, T. H. (2001). Systemic risk in financial systems. *Management Science*, *47*(2), 236–249.

Elliott, M., Golub, B. & Jackson, M. O. (2014). Financial networks and contagion. *American Economic Review*, *104*(10), 3115–3153.

Erdős, P. & Rényi, A. (1959). On random graphs I. *Publicationes Mathematicae Debrecen*, *6*, 290–297.

Euler, L. (1741). Solutio problematis ad geometriam situs pertinentis. *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, *8*, 128–140.

Fafchamps, M. & Gubert, F. (2007). The formation of risk sharing networks. *Journal of Development Economics*, *83*(2), 326–350.

Fiedler, M. (1973). Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, *23*(2), 298–305.

Fisher, R. A. (1935). *The design of experiments*. Edinburgh: Oliver and Boyd.

Frank, O. & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, *81*(395), 832–842.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, *40*(1), 35–41.

Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks*, *1*(3), 215–239.

Friedman, J., Hastie, T. & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*(3), 432–441.

Furfine, C. H. (2003). Interbank exposures: Quantifying the risk of contagion. *Journal of Money, Credit and Banking*, *35*(1), 111–128.

Galeotti, A., Golub, B. & Goyal, S. (2020). Targeting interventions in networks. *Econometrica*, *88*(6), 2445–2471.

Gentzkow, M., Kelly, B. & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, *57*(3), 535–574.

Geyer, C. J. & Thompson, E. A. (1992). Constrained monte carlo maximum likelihood for dependent data. *Journal of the Royal Statistical Society: Series B (Methodological)*, *54*(3), 657–699.

Gilardi, F., Alizadeh, M. & Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, *120*(30), e2305016120.

Gilbert, E. N. (1959). Random graphs. *Annals of Mathematical Statistics*, *30*(4), 1141–1144.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. (2017). Neural message passing for quantum chemistry. In *Proceedings of the 34th international conference on machine learning* (pp. 1263–1272).

Goldsmith-Pinkham, P. & Imbens, G. W. (2013). Social networks and the identification of peer effects. *Journal of Business & Economic Statistics*, *31*(3), 253–264.

Golub, B. & Jackson, M. O. (2010). Naïve learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, *2*(1), 112–149.

Golub, B. & Sadler, E. (2016). Learning in social networks. In Y. Bramoullé, A. Galeotti & B. W. Rogers (Eds.), *The oxford handbook of the economics of networks* (pp. 504–542). Oxford: Oxford University Press.

Gori, M., Monfardini, G. & Scarselli, F. (2005). A new model for learning in graph domains. In *Proceedings of the ieee international joint conference on neural networks* (pp. 729–734).

Goyal, S. (2007). *Connections: An introduction to the economics of networks*. Princeton: Princeton University Press.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica*, *85*(4), 1033–1063.

Graham, B. S. (2020). Chapter 2 - network data. In S. N. Durlauf, L. P. Hansen, J. J. Heckman & R. L. Matzkin (Eds.), *Handbook of econometrics, volume 7a* (Vol. 7, p. 111-218). Elsevier. Retrieved from https://www.sciencedirect.com/science/article/pii/S1573441220300015 doi: https://doi.org/10.1016/bs.hoe.2020.05.001

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, *78*(6), 1360–1380.

Grover, A. & Leskovec, J. (2016). node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 855–864). New York: ACM.

Halloran, M. E. & Struchiner, C. J. (1995). Causal inference in infectious diseases. *Epidemiology*, *6*(2), 142–151.

Hamilton, W. L., Ying, R. & Leskovec, J. (2017). Inductive representation learning on large graphs. In *Advances in neural information processing systems* (Vol. 30).

Handcock, M. S. (2003). *Assessing degeneracy in statistical models of social networks* (Working Paper No. 39). Center for Statistics and the Social Sciences, University of Washington.

Hanneke, S., Fu, W. & Xing, E. P. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics*, *4*, 585–605.

Harary, F. (1969). *Graph theory*. Reading, MA: Addison-Wesley.

Heckerman, D. (1995). *A tutorial on learning with bayesian networks* (Technical Report No. MSR-TR-95-06). Redmond, WA: Microsoft Research.

Heckman, J. J. (1981). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In C. F. Manski & D. McFadden (Eds.), *Structural analysis of discrete data with econometric applications* (pp. 179–195). Cambridge, MA: MIT Press.

Hidalgo, C. A. & Hausmann, R. (2009). The building blocks of economic complexity. *Proceedings of the National Academy of Sciences*, *106*(26), 10570–10575.

Hoberg, G. & Phillips, G. M. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, *124*(5), 1423–1465.

Hochberg, Y. V., Ljungqvist, A. & Lu, Y. (2007). Whom you know matters: Venture capital networks and investment performance. *Journal of Finance*, *62*(1), 251–301.

Holland, P. W. & Leinhardt, S. (1971). Transitivity in structural models of small groups. *Comparative Group Studies*, *2*(2), 107–124.

Holland, P. W. & Leinhardt, S. (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association*, *76*(373), 33–50.

Hudgens, M. G. & Halloran, M. E. (2008). Toward causal inference with interference. *Journal of the American Statistical Association*, *103*(482), 832–842.

Hunter, D. R. & Handcock, M. S. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics*, *15*(3), 565–583.

Jackson, M. O. (2008). *Social and economic networks*. Princeton: Princeton University Press.

Jackson, M. O., Rogers, B. W. & Zenou, Y. (2017). The economic consequences of social-network structure. *Journal of Economic Literature*, *55*(1), 49–95.

Jackson, M. O. & Watts, A. (2002). The evolution of social and economic networks. *Journal of Economic Theory*, *106*(2), 265–295.

Jackson, M. O. & Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of Economic Theory*, *71*(1), 44–74.

Jordan, M. I. (Ed.). (1998). *Learning in graphical models*. Cambridge, MA: MIT Press.

Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, *18*(1), 39–43.

Kelejian, H. H. & Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, *40*(2), 509–533.

Kempe, D., Kleinberg, J. & Tardos, E. (2003). Maximizing the spread of influence through a social network. In *Proceedings of the ninth acm sigkdd international conference on knowledge discovery and data mining* (pp. 137–146). New York: ACM.

Kim, D. A., Hwong, A. R., Stafford, D., Hughes, D. A., O'Malley, A. J., Fowler, J. H. & Christakis, N. A. (2015). Social network targeting to maximise population behaviour change: A cluster randomised controlled trial. *The Lancet*, *386*(9989), 145–153.

Kipf, T. N. & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *Proceedings of the 5th international conference on learning representations.*

Kojevnikov, D., Marmer, V. & Song, K. (2021). Limit theorems for network dependent random variables. *Journal of Econometrics*, *222*(2), 882–908.

Koller, D. & Friedman, N. (2009). *Probabilistic graphical models: Principles and techniques*. Cambridge, MA: MIT Press.

Krivitsky, P. N. & Handcock, M. S. (2014). A separable model for dynamic networks. *Journal of the Royal Statistical Society: Series B*, *76*(1), 29–46.

Landau, E. (1895). Zur relativen Wertbemessung der Turnierresultate. *Deutsches Wochenschach*, *11*(42), 366–369.

Landau, E. (1914). Über Preisverteilung bei Spielturnieren. *Zeitschrift für Mathematik und Physik*, *63*, 192–202.

Lauritzen, S. L. & Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society: Series B*, *50*(2), 157–224.

LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.

Lee, L.-f. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, *72*(6), 1899–1925.

Leung, M. P. (2022). Causal inference under approximate neighborhood interference. *Econometrica*, *90*(1), 267–293.

Li, Y., Tarlow, D., Brockschmidt, M. & Zemel, R. (2016). Gated graph sequence neural networks. In *International conference on learning representations.*

Liu, X. & Lee, L.-f. (2010). Gmm estimation of social interaction models with centrality. *Journal of Econometrics*, *159*(1), 99–115.

Ludwig, J. & Mullainathan, S. (2024). Machine learning as a tool for hypothesis generation. *Quarterly Journal of Economics*, *139*(2), 751–827.

Ludwig, J., Mullainathan, S. & Rambachan, A. (2025). *Large language models: An applied econometric framework.* (NBER Working Paper No. 33344)

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, *60*(3), 531–542.

Manski, C. F. (2013). Identification of treatment response with social interactions. *The Econometrics Journal*, *16*(1), S1–S23.

Mele, A. (2017). A structural model of dense network formation. *Econometrica*, *85*(3), 825–850.

Miguel, E. & Kremer, M. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, *72*(1), 159–217.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S. & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (Vol. 26). Curran Associates.

Moberg, J. & Kramer, M. (2015). A brief history of the cluster randomised trial design. *Journal of the Royal Society of Medicine*, *108*(5), 192–198.

Moffitt, R. A. (2001). Policy interventions, low-level equilibria, and social interactions. In S. N. Durlauf & H. P. Young (Eds.), *Social dynamics* (pp. 45–82). Cambridge, MA: MIT Press.

Moreno, J. L. (1934). *Who shall survive? a new approach to the problem of human interrelations*. Washington, DC: Nervous and Mental Disease Publishing Company.

Morris, S. (2000). Contagion. *Review of Economic Studies*, *67*(1), 57–78.

Myerson, R. B. (1977). Graphs and cooperation in games. *Mathematics of Operations Research*, *2*(3), 225–229.

Newey, W. K. & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3),

703–708.

Ng, A. Y., Jordan, M. I. & Weiss, Y. (2001). On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems* (Vol. 14). MIT Press.

Ord, J. K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, *70*(349), 120–126.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo, CA: Morgan Kaufmann.

Perozzi, B., Al-Rfou, R. & Skiena, S. (2014). DeepWalk: Online learning of social representations. In *Proceedings of the 20th acm sigkdd international conference on knowledge discovery and data mining* (pp. 701–710). New York: ACM.

Price, D. d. S. (1965). Networks of scientific papers. *Science*, *149*(3683), 510–515.

Price, D. d. S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, *27*(5), 292–306.

Rauch, J. E. (1999). Networks versus markets in international trade. *Journal of International Economics*, *48*(1), 7–35.

Rauch, J. E. & Trindade, V. (2002). Ethnic chinese networks in international trade. *The Review of Economics and Statistics*, *84*(1), 116–130.

Rosenbaum, P. R. (2007). Interference between units in randomized experiments. *Journal of the American Statistical Association*, *102*(477), 191–200.

Rubin, D. B. (1980). Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American Statistical Association*, *75*(371), 591–593.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536.

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly Journal of Economics*, *116*(2), 681–704.

Sacerdote, B. (2014). Experimental and quasi-experimental analysis of peer effects: Two steps forward? *Annual Review of Economics*, *6*, 253–272.

Saveski, M., Pouget-Abadie, J., Saint-Jacques, G., Duan, W., Ghosh, S., Xu, Y. & Airoldi, E. M. (2017). Detecting network effects: Randomizing over randomized experiments. In *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining* (pp. 1027–1035). New York: ACM.

Sävje, F., Aronow, P. M. & Hudgens, M. G. (2021). Average treatment effects in the presence of unknown interference. *Annals of Statistics*, *49*(2), 673–701.

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M. & Monfardini, G. (2009). The graph neural network model. *IEEE Transactions on Neural Networks*, *20*(1), 61–80.

Serrano, M. A. & Boguñá, M. (2003). Topology of the world trade web. *Physical Review E*, *68*(1), 015101.

Sheng, S. (2020). A structural econometric analysis of network formation games through subnetworks. *Econometrica*, *88*(5), 1829–1858.

Shi, J. & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(8), 888–905.

Snijders, T. A. B. (2001). The statistical evaluation of social network dynamics. *Sociological Methodology*, *31*(1), 361–395.

Snijders, T. A. B., van de Bunt, G. G. & Steglich, C. E. G. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks*, *32*(1), 44–60.

Sobel, M. E. (2006). What do randomized studies of housing mobility demonstrate? causal inference in the face of interference. *Journal of the American Statistical Association*, *101*(476), 1398–1407.

Strauss, D. & Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, *85*(409), 204–212.

Törnberg, P. (2024). Large language models outperform expert coders and supervised classifiers at annotating political social media messages. *Social Science Computer Review*. (Advance online publication)

Ugander, J., Karrer, B., Backstrom, L. & Kleinberg, J. (2013). Graph cluster randomization: Network exposure to multiple universes. In *Proceedings of the 19th acm sigkdd international conference on knowledge discovery and data mining* (pp. 329–337). New York: ACM.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (Vol. 30).

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P. & Bengio, Y. (2018). Graph attention networks. In *Proceedings of the 6th international conference on learning representations.*

Wager, S. & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242.

Wasserman, S. & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.

Wasserman, S. & Pattison, P. (1996). Logit models and logistic regressions for social networks: I. an introduction to markov graphs and p*. *Psychometrika*, *61*(3), 401–425.

Watts, D. J. & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, *41*(3/4), 434–449.

Wichmann, P., Brintrup, A., Baker, S., Woodall, P. & McFarlane, D. (2020). Extracting supply chain maps from news articles using deep neural networks. *International Journal of Production Research*, *58*(17), 5320–5336.

Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics*, *20*(1), 39–54.

Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, *20*, 557–585.

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C. & Yu, P. S. (2020). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(1), 4–24.

Ying, R., He, R., Chen, K., Eksombatchai, P., Hamilton, W. L. & Leskovec, J. (2018). Graph convolutional neural networks for web-scale recommender systems. In *Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining* (pp. 974–983).

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics*, *85*(1), 9–23.

Index