Guanyi Wang
Assistant Professor of Economics
Department of Economics, Washington University in St. Louis
guanyi@wustl.edu

Dear Professors Breza, Chandrasekhar, and Viviano,

I am writing to apply for the NBER Workshop on Development Economics and Networks for Early-Career Researchers, to be held in Cambridge, MA on December 5, 2026. I am an Assistant Professor of Economics at Washington University in St. Louis and received my PhD from University College London in 2025. My research lies at the intersection of network econometrics and policy learning.

I would like to present my working paper "Causal Inference with Endogenous Network Formation," coauthored with Mengsi Gao and Toru Kitagawa. The paper develops identification and estimation for direct and indirect treatment effects in settings where the network responds to the treatment. This is motivated by the microfinance and network change setting studied in Banerjee, Breza, Chandrasekhar, Duflo, Jackson, and Kinnan (2024), and I believe it speaks directly to the workshop's focus on the frontier of empirical network research in development.

If I am shortlisted, Professor Toru Kitagawa (Brown University, toru_kitagawa@brown.edu) will provide a recommendation on my behalf.

Enclosed with this letter are my curriculum vitae, including a list of my PhD coursework; a one-page statement of my motivation for attending and current research interests; and the working paper. Thank you for considering my application.

Sincerely,

Guanyi Wang

Guanyi Wang

Personal info

Department of Economics
Washington University in St. Louis
One Brookings Drive
St. Louis, MO 63130

Email: guanyi@email.wustl.edu
Web: www.guanyi-wang.com
Citizenship: Chinese
Last Update: 28th of June 2026

Employment

Assistant Professor, Washington University in St. Louis 2025-

Education

PhD in Economics, University College London 2025
Supervisors: Toru Kitagawa, Aureo de Paula, Timothy Christensen, Andrei Zeleneev

MRes in Economics, University College London 2019

MSc in Economics, University College London 2018

BSc in Economics, University of Liverpool 2017

Research Field: Econometric theory, applied microeconomics, and social network

Grant

Weidenbaum Center Small Grants - $11,080 2026

Weidenbaum Center Small Grants - $15,000 2025

Working Paper

wp1 Robust Network Targeting with Multiple Nash Equilibria

Published (including forthcoming) Articles

1. Individualized Treatment Allocation in Sequential Network Games, with *Toru Kitagawa* (2026). *Quantitative Economics*, conditionally accepted.
2. Who Should Get Vaccinated? Individualized Allocation of Vaccines Over SIR Network, with *Toru Kitagawa* (2023). *Journal of Econometrics*, Vol. 232 (1), pp 109-131.

Scholarship

Center for Microdata Methods and Practice (CeMMAP) 2022-2025
PhD scholarship

Conferences and Seminars (including scheduled)

2026 Southern Methodist University, Indiana University, CES NA Annual Conference, University College London, Africa Econometric Society Meeting, INFORMS Annual Meeting, SEA conference

2025 Chamberlain Seminar, Georgetown University, Microeconometrics Class of 2025 Conference, Econometric Society World Congress, SETA 2025, Janeway Institute PhD Workshop on Economic Networks, Washington University St. Louis, University of Toronto, Xiamen University, Jinan University IESR, DUFE, Liaoning University

2024 Tsinghua University, Peking University, $(EC)^2$ Conference, Midwest Econometrics Group Conference, UCL Economics, Oxford Economics, UCL-CeMMAP Ph.D. Econometrics Research Day, Bristol Econometric Study Group, IAAE Annual Conference, The 8th IAER Econometrics Workshop, North American Summer Meeting of the Econometric Society

2023 $(EC)^2$ Conference, UCL-CeMMAP Ph.D. Econometrics Research Day, Using Data to Make Decisions Workshop, Bristol Econometric Study Group, Asian Meeting of the Econometric Society, North American Summer Meeting of the Econometric Society, The Seventh CCER Summer Institute, UCL Brownbag Seminar, LSE Work-in-progress Seminar

2021 Bristol Econometric Study Group, IAAE Annual Conference, Asian Meeting of the Econometric Society, North American Summer Meeting of the Econometric Society

Referee Service for Academic Journals

Journal of Applied Econometrics, Economic Journal, Journal of Econometrics, Journal of Business & Economic Statistics, Journal of the Royal Statistical Society Series A, Journal of Economic theory

Teaching

Washington University in St. Louis:

2025-2026	**ECON8150: Advanced Topics in Econometric Theory** **ECON3150: Introduction to Econometrics**

University College London:

2021-2025	ECON0019: Quantitative Economics and Econometrics
2020-2021	ECON0064: Econometrics ECON0019: Quantitative Economics and Econometrics
2019-2020	ECON0058: Time Series Econometrics ECON0060: Advanced Microeconometrics ECON0019: Quantitative Economics and Econometrics

Research Assistant

2019-2020	Public Policy: Machine Learning for Tax Audit Targeting Work For: Toru Kitagawa

Skills

Language: English (fluent), Mandarin (Native)

Software: Matlab, R, and Stata

Statement of Motivation and Research Interests

My research lies at the intersection of econometrics, causal inference, and the economics of networks. I am particularly interested in settings where economic interventions affect not only individuals directly but also the social interactions through which information, behavior, and treatment effects spill over. These settings create challenges for both identification and policy design because networks may themselves respond endogenously to interventions.

I am applying to the NBER Workshop on Development Economics and Networks because its focus is closely connected to my current research agenda. In the paper I am submitting, *Causal Inference with Endogenous Network Formation*, joint with Mengsi Gao and Toru Kitagawa, we study randomized interventions when treatment can change the network itself. Although treatment assignment is randomized, exposure to treated peers can remain endogenous when unobserved heterogeneity affects both network formation and outcomes. We develop a control function approach to identify causal effects in this setting and study estimation methods that can accommodate sparse networks, including bounded degree settings. I would especially value feedback from researchers working with empirical network data on the assumptions needed to make these methods useful in development applications.

More broadly, my research studies causal inference and policy learning in settings with strategic and network interactions. One part of my work considers individualized treatment allocation when agents interact through networks, including settings in which choices are determined by a network game and multiple equilibria may arise. Another part studies causal inference when treatments generate broader general equilibrium effects in decentralized markets, so that an individual's outcome may be affected not only by the direct treatment assignments but also through equilibrium effects. This work aims to develop nonparametric identification strategies for causal effects in decentralized market environments. I am also interested in social learning and in how information transmitted through networks affects technology adoption and other economic decisions.

Many of the applications motivating this work come naturally from development economics, where interventions frequently operate through social relationships. Programs involving financial access, technology adoption, information provision, or other household-level treatments can change both individual behavior and patterns of interaction. Understanding these endogenous network responses is therefore important for separating direct and indirect treatment effects and for designing policies that account for equilibrium responses.

The workshop would provide an especially valuable opportunity to discuss these questions with researchers combining network methods with substantive development applications. I hope to receive feedback on how to connect the econometric framework more closely to empirically relevant network settings, learn from recent work using network data in development economics, and meet other early career researchers working on related questions. These interactions would be particularly useful as I develop a broader research agenda on causal inference and policy design in a network setting.

Causal Inference with Endogenous Network Formation

Mengsi Gao Toru Kitagawa Guanyi Wang

University of Southern California Brown University Washington University in St. Louis

Extended Abstract

Treatment interventions can affect outcomes not only by changing individuals' treatment status, but also by changing whom they interact with. For example, access to formal finance may directly affect household wealth while also altering informal lending relationships (Banerjee et al., 2024). In such settings, the network is itself part of the behavioral response to treatment, rather than a fixed channel through which treatment effects spill over. Treatment induced network formation therefore creates additional causal channels that have received limited attention in the existing literature. Moreover, even when treatment is randomly assigned, the resulting network exposure may remain endogenous because unobserved heterogeneity can jointly determine network interactions and outcomes. Motivated by these challenges, this paper develops a framework for disentangling the causal effects operating through endogenous network formation. We provide nonparametric identification and estimation results using a control-function approach based on the non-backtracking matrix Abbe et al. (2023), and show that our method continues to apply in sparse networks with bounded expected degree.

Framework and causal parameters. We consider a randomized experiment in which researchers observe individual outcomes $\boldsymbol{Y} \coloneqq (Y_1, ..., Y_n) \in \mathbb{R}^n$, treatment assignments $\boldsymbol{D} \coloneqq (D_1, ..., D_n) \in \{0,1\}^n$, and covariates $\boldsymbol{X} \coloneqq (X_1, ..., X_n) \in \mathcal{X} = \mathbb{R}^{n\times k}$. We consider an endogenous network setting in which network formation can depend on the assigned treatments and unobserved heterogeneity that can be correlated with the potential outcomes. Accordingly, we distinguish the pre-intervention network $G^0 = (G^0_{ij} : i, j \in \mathcal{N})$ from the post-intervention network $G^1 = (G^1_{ij} : i, j \in \mathcal{N})$. In G^0, agents i and j form a link if the *surplus function* (Leung and Moon, 2025) from linking prior to treatment implementation is nonnegative:

$$G^0_{ij} = \mathbb{1}\{g_0(X_i, X_j, a_i, a_j) - u^0_{ij} \geq 0\}. \tag{1}$$

where $\boldsymbol{a} \coloneqq \{a_i\}_{i\in\mathcal{N}} \subset \mathbb{R}^{d_a}$ denotes unit-specific time-invariant unobserved heterogeneity and $\boldsymbol{u}^0 \coloneqq \{u^0_{ij}\}_{i,j\in\mathcal{N}}$ denotes a pre-intervention idiosyncratic perference shock to each potential link. In G^1, agents i and j form a link if the total surplus from linking after treatment implementation is nonnegative:

$$G^1_{ij} = \mathbb{1}\{g_1(X_i, X_j, D_i, D_j, a_i, a_j) - u^1_{ij} \geq 0\}, \tag{2}$$

where $\boldsymbol{u}^1 \coloneqq \{u^1_{ij}\}_{i,j\in\mathcal{N}}$ is a post-intervention idiosyncratic perference shock to each potential link. Network formation models in (1) and (2) correspond to the well-known pairwise-stability solution concept under transferable utility (Jackson, 2010).

In our endogenous network setting, treatment may affect M_i both directly—through its dependence on D—and indirectly, through treatment-induced changes in the post-intervention

network structure G^1. To be explicit about such causal dependence of M_i onto the entire treatment assignment vector D, we introduce the potential outcome notation (Rubin, 1978; Rubin et al., 1990) for the exposure mapping: for each $d = (d_i, d_{-i}) \in \{0,1\}^n$,

$$M_i(d_i, d_{-i}) := h_i(d, G^1(d)), \;\; G^1(d) = (\mathbb{1}\{g_1(X_i, X_j, d_i, d_j, a_i, a_j) - u_{ij}^1 \geq 0\} : i \neq j \in \mathcal{N}). \quad (3)$$

To define the causal effects of the treatments D on the outcomes, through the direct effect of own treatment and the indirect spillover effect of others' treatments through the exposure mapping, we introduce the potential outcomes of unit i that we would observe if we would exogenously set the treatment and exposure mapping to $D_i = d_i$ and $M_i = m_i$,

$$\{Y_i(d_i, m_i) := f(X_i, d_i, m_i, a_i, \varepsilon_i) : d_i \in \{0,1\}, m_i \in \mathcal{M}\}, \quad (4)$$

where $\mathcal{M}$ is the set of possible mediator values. As the notation represents, the potential outcomes $Y_i(d_i, m_i)$ are indexed by unit i's own treatment and the exposure mapping, where the dependence of potential outcomes through others' assigned treatments are only through only through M_i. Because M_i itself depends on the treatment assignment vector, we also define the nested potential outcomes

$$\{Y_i(d_i, M_i(d_i', \boldsymbol{d}_{-i})) : d_i \in \{0,1\}, (d_i', \boldsymbol{d}_{-i}) \in \{0,1\}^n\}, \quad (5)$$

We interpret $M_i(d_i, d_{-i})$ as the value of exposure mapping of unit i that would realize if the treatment assignment vector were exogenously set to $d = (d_i, d_{-i}) \in \{0,1\}^n$. such as the number or share of treated friends. The network affects outcomes only through this exposure, while the outcome function otherwise permits nonlinearities, treatment–exposure interactions, and unobserved heterogeneity.

We distinguish controlled effects, which hold exposure fixed or intervene on its value, from natural effects, which allow exposure to respond endogenously to treatment. The latter include the natural direct effect of own treatment, the natural indirect own effect operating through treatment-induced changes in exposure, and the natural indirect spillover effect generated by other individuals' treatments. These estimands separate the effect of receiving treatment from the consequences of the social interactions that the intervention induces. They also distinguish policies that directly manipulate exposure from policies that change treatment assignments while allowing the network to adjust.

Identification. The key feature of the peer effect model, equations (1), (2), and (3), is that individual i's unobserved characteristic a_i confounds the link formations and i's outcome. As a result, even though D_i's are randomly assigned, we cannot assume mediator $M_i = h_i(D, G^1)$ to be randomly assigned conditional on any pre-treatment observable characteristics. This makes identification of the causal estimands defined in the previous section challenging. Moreover, if the intervention can reshape the network, it becomes difficult to identify the direct effect of a unit's own treatment separately from indirect effects induced by the network changes.

Given a is unobserved, our approach exploits an eigenfunction of the graphon function as a control function. We show that, once we control for unit i's eigenfunctions, M_i and $Y_i(d_i, m)$

are conditionally independent. This delivers inverse probability weighted and regression representations of controlled mean potential outcomes without imposing a parametric outcome equation. Identifying natural effects additionally requires counterfactual mediator distributions. We use pre-intervention network information to connect latent network characteristics across treatment states: under an injectivity condition on the pre-treatment eigenfunction vector, together with the relevant conditional-independence and overlap restrictions, this information identifies the counterfactual distributions needed to distinguish natural direct and indirect own effects.

Estimation in bounded-degree networks. The estimation problem is distinct from the control-function identification argument. When expected degrees remain bounded, exposure can retain nondegenerate variation, but consistent recovery of each individual's latent network position cannot simply be assumed. Our estimation strategy therefore targets population moments involving the latent controls rather than inserting estimated unit-level eigenfunctions into an outcome regression or an inverse-probability-weighted estimator.

The construction uses nonbacktracking spectral scores and weighted-star statistics from an edge-split pre-intervention network, building on Abbe et al. (2023). We extend the moment construction by attaching an observed mark to each star's center. For a treatment–exposure cell, let $I_i(d,m) = \mathbf{1}\{D_i = d, M_i = m\}$ and let ψ_i^0 denote the pre-treatment eigenfunction control. Using marks 1, $I_i(d,m)$, and $Y_i I_i(d,m)$ produces estimates of

$$\mathbb{E}[(\psi_i^0)^\alpha], \qquad \mathbb{E}[I_i(d,m)(\psi_i^0)^\alpha], \qquad \mathbb{E}[Y_i I_i(d,m)(\psi_i^0)^\alpha],$$

for fixed multi-indices α. Observed covariates are included in the conditioning sets and marked moments whenever they are needed for the control-function argument. Thus, the network statistics retain the association of latent characteristics with exposure and outcomes, not merely the marginal distribution of network characteristics.

We establish consistency of each fixed collection of marked moments under homogeneity, regularity, and spectral-signal conditions on the bounded-degree pre-treatment graphon, bounded marks, conditional exogeneity of the held-out pre-treatment edges, and conditional mean and covariance restrictions on the marks. These restrictions allow dependence across individuals and dependence of the marks on latent characteristics. The retained pre-treatment eigenfunctions must also be sufficiently informative to support the causal control-function argument.

To recover causal averages, we combine the ordinary, exposure-marked, and outcome-marked moments through a regularized polynomial-sieve estimator. Under regularity conditions, we establish consistency for controlled mean potential outcomes along suitable sequences of increasing sieve dimensions and vanishing regularization. Contrasts of these means recover controlled direct, mediator, and total effects. Importantly, this result concerns an estimator built from aggregate marked moments; it does not require consistent estimation of every individual's control value or propensity score.

The paper connects causal mediation analysis with endogenous network formation and sparse-network estimation. Its central contribution is to distinguish treatment effects operating through changing social connections while exploiting network information to

address unobserved confounding. The estimation result shows how causal averages can be recovered from joint network–outcome moments even when individual latent network positions are not consistently recovered.

References

ABBE, E., S. LI, AND A. SLY (2023): "Learning sparse graphons and the generalized Kesten–Stigum threshold," *The Annals of Statistics*, 51, 599–623.

BANERJEE, A., E. BREZA, A. G. CHANDRASEKHAR, E. DUFLO, M. O. JACKSON, AND C. KINNAN (2024): "Changes in social network structure in response to exposure to formal credit markets," *Review of Economic Studies*, 91, 1331–1372.

JACKSON, M. O. (2010): *Social and economic networks*, Princeton university press.

LEUNG, M. P. AND H. R. MOON (2025): "Normal approximation in large network models," *arXiv preprint arXiv:1904.11060*.

RUBIN, D. B. (1978): "Bayesian inference for causal effects: The role of randomization," *The Annals of statistics*, 34–58.

RUBIN, D. B. ET AL. (1990): "Comment: Neyman (1923) and causal inference in experiments and observational studies," *Statistical Science*, 5, 472–480.