Rearranging Attributes in Networked Economies

Janelle Schlossberger*

July 2021

Abstract

This work studies an important departure from the classical networked economy. In the benchmark case, an external decision-making observer has full information about the networked economy. In this work, the observer does not have full information, yet must make decisions impacting the economy's agents. Specifically, the observer does not know how the attributes of the economy sit on the network's nodes. This work develops a complete, closed-form statistical approach that enables the observer to overcome this lack of information and still execute a decision. The observer must consider all possible arrangements of attributes on the network nodes. By exhaustively rearranging attributes, the observer can construct probability distributions that accurately characterize the economy. In this work, we first develop the necessary theoretical tools and we then show how the observer can employ these tools in the following settings: (1) education with peer effects, (2) consumption with network externalities, and (3) crime.

Key Words: networked economy, rearranging attributes, uncertainty modeling, policy selection

1 Introduction

This work studies networked economies with an external decision-making observer. The external observer lacks full information about the networked economy, yet must still execute a decision that impacts all of the economy's agents. This work

*First draft: May 2021. Email: jschloss@post.harvard.edu. The author would like to thank Krishna Dasaratha, Mira Frick, Kevin He, and Jonathan Libgober for helpful comments.

develops a set of theoretical tools that the observer can use to overcome this lack of information. The tools help the observer learn as much as possible about the networked economic environment. The observer then uses all available gathered information to make a decision.

Specifically, the theoretical setting considered in this work is as follows. There is an economy with a network and a collection of real-valued attributes. The attributes individually sit on the nodes of the network. The outside observer knows the topology of the network as well as the value of each attribute in the collection, as depicted in the left panel of Figure 1. However, the outside observer does not know how the attributes precisely sit on the nodes of the network. The particular mapping of attributes to nodes matters because it impacts the actions of the economy's agents. It yields a crucial data point that the observer would ordinarily use, in the full-information setting, for decision-making.

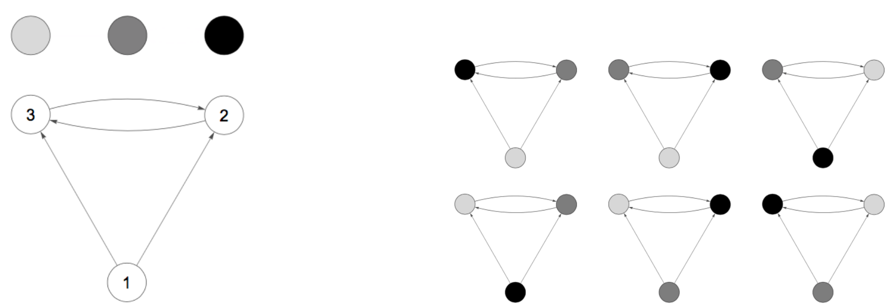

Figure 1: (Left) The knowledge of the outside observer: the collection of attributes and the topology of the network on which these attributes sit. (Right) Exhaustively rearranging attributes on the network.

Unfortunately, the outside observer does not know this exact data point. What the observer can do is consider all of the different ways that the attributes can possibly sit on the nodes of the network. The outside observer starts with a particular mapping of attributes to nodes. The observer then rearranges the attributes on the network nodes until exhaustion. Figure 1 illustrates this approach. The outside observer has knowledge of a networked economy with three nodes and three attributes, as depicted in the left panel of Figure 1. There are $3! = 6$ possible arrangements of these attributes on the network nodes. The right panel of Figure 1 presents all six possible arrangements. Now, associated with each arrangement of attributes on network nodes is a data point. By considering all possible arrangements and assuming that each arrangement is equally likely, we end up with a probability distribution of data points that the observer can then use to make a decision.

It is therefore important that we be able to construct this probability distribution and characterize its features through closed-form expressions, as the properties of this probability distribution impact: (1) the observer's understanding of the networked economy and (2) the observer's ultimate decision. In this work, we develop a comprehensive set of theoretical tools that thoroughly characterize the probability distribution of data points. First, we establish closed-form expressions for the first four moments of the probability distribution. These moments provide meaningful information to the observer. The first moment, for example, tells the observer the value of the average data point after considering all possible arrangements. The second moment, meanwhile, identifies the scope for variation in the value of the data point. Second, we provide the necessary and sufficient conditions that make the probability distribution degenerate; in this setting, every arrangement of attributes yields the same data point, so it is as if the observer actually knows how the attributes sit on the network. Third, we show how to compute the lower and upper bounds on the support of the probability distribution. The lower bound tells the observer the value of the smallest possible data point, while the upper bound tells the observer the value of the largest possible data point. Fourth, we establish a central limit theorem-type result that characterizes the asymptotic behavior of the probability distribution for a large number of network nodes. Fifth, we construct an asymptotic expansion to approximate the CDF of the probability distribution. As we will later see, the asymptotic expansion is incredibly useful to the observer. The asymptotic expansion allows the observer to make a decision ensuring the desired outcome for a certain percentage (say, 95 percent) of all feasible arrangements. All of these theoretical tools are completely general: the observer can study the probability distribution of data points through closed-form expressions for a network of any size and/or topology and for any collection of real-valued attributes.

After developing these theoretical tools, we show how the decision-making observer can use these tools to better understand the economic environment in three different settings: (1) education with peer effects, (2) consumption with network externalities, and (3) crime. We proceed to provide an overview of each application, with Table 1 breaking down the items of interest. For the first application, the observer is a school principal. It is the beginning of the school year, and the principal would like to know which students are going to fail and the overall failure rate in the student body in the absence of any educational intervention. Now, student academic performance and student effort depend on individual demographic characteristics and the behavior of one's peers. The principal knows the topology of the peer influence network from the observations of faculty and staff. The principal also has access to anonymized student demographic information through the administration of a school-wide survey; such anonymous surveys are fairly common, as will be discussed later on in this work. However, since the demographic information is anonymous, the principal does not know how to map the demographic information to particular network nodes. The principal can use the theoretical tools developed in this work to instead compute each student's probability of failure and the overall expected failure rate in the student body, and then design an appropriate intervention.

In the second application, the decision-making observer is a government. The government would like to levy a consumption tax on individuals to raise a certain amount of tax revenue. Each individual's consumption decision depends on preferences, the size of the tax, the price of the good, individual wealth, and, importantly, the consumption behavior of network neighbors. The government knows each individual's wealth, say, from collecting income taxes. The government also has a sense of the network topology, say, from publicly available anonymized data on social networks and/or communication networks.[1] However, the government does not know how these individuals and their wealth actually sit on the nodes of the network. If the government had full information, it could set a tax rate and there would be one possible amount of tax revenue raised. Since the government does not know how to map wealth attributes to network nodes, for any given tax rate, there is instead a distribution of possible levels of tax revenue raised. Operating in this uncertain environment, the government chooses a tax rate so that tax revenue raised exceeds a

[1]For example, the Stanford Network Analysis Project, available at http://snap.stanford.edu, offers a repository of anonymous network datasets.

	Application 1: Peer Networks and Education Outcomes	**Application 2: Tax Revenue in an Environment with Consumption Network Effects**	**Application 3: Managing Crime**
Identity of Observer	School principal	Government	Mayor and the police
Type of Network	Peer influence network	Social influence network	Criminal influence network
Attribute	A linear combination of a student's demographic characteristics	Consumer wealth	The amount of resources used to monitor each criminal
What the Observer Does Not Know	The network position of each anonymized student's demographic characteristics	The network position of each consumer	The network position of each criminal
The Observer's Decision	Designing an appropriate educational intervention given each student's probability of failure and the overall expected failure rate	Determining the size of the consumption tax to levy given that the government desires a certain amount of total revenue	Determining how many resources to allocate towards monitoring each criminal so that total crime is sufficiently low

Table 1: Items of interest for each of the three applications.

minimum desired threshold with a 95-percent probability.

In the third application, the decision-making observers are the mayor and the city's police. The mayor is trying to get re-elected to another term, but the level of crime in the city is too high. There is a population of individuals, organized on an interaction network, that engage in criminal activity. The police know the topology of the interaction network because they are targeting a specific type of crime that has a naturally occurring network structure. However, the police, at least initially, do not know where each criminal sits on the network. The mayor and the police must implement a monitoring policy to reduce crime, in which they decide how many monitoring resources to allocate to each individual, without full information. Now the police can expend resources to learn how the criminals sit on the network, perhaps by wiretapping or going undercover to infiltrate the gang. If it is too costly or dangerous to acquire this information, though, the police must take a different approach. The police must craft a monitoring policy without knowing where these individuals sit on the network. For any monitoring policy, the police can use the theoretical tools developed in this work to construct a probability distribution of possible levels of total crime. The police choose a monitoring policy to maximize, as much as is feasible, the probability of crime being below the threshold that the public is willing to tolerate.

This work has two main assumptions. First, in developing the theoretical tools, this work assumes that each possible arrangement of attributes on network nodes is equally likely. This is a reasonable benchmark, and therefore the one that we choose to employ. Second, this work assumes that the decision-making observer knows the network topology. One might view the acquisition of full network knowledge by the decision-making observer as being costly or difficult. However, for each of the applications studied in this work, it is indeed a very natural assumption. For example, depending on the particular application, the observer might be able to gather network knowledge from agents in the economy, the observer might have access to a publicly available anonymized network, and/or the observer might know the generative process by which the network forms. Now, the exact topology of the network does not always need to be known in order for the observer to make an appropriate decision. In certain settings, the observer can compute the necessary features of the network without full network knowledge. For instance, Dasaratha (2020) shows how network features, namely centralities, can be computed with high probability for classes of large random networks. Parise and Ozdaglar (2019) shows how agent equilibria can be computed for large network games sampled from a graphon. Breza et al. (2020) shows how to obtain the parameters of a network formation model in the absence of network knowledge by using aggregated relational data.

1.1 Related Literature

Within the past decade, a literature has emerged that involves the study of economic networks with an outside observer. Depending on the particular study, this outside observer can assume many different possible identities: a social planner, a policymaker, a principal, a monopolist, an interest group, or perhaps a monetary authority. Candogan et al. (2012), Bloch and Quérou (2013), and Fainmesser and Galeotti (2016) all study settings in which a monopolist must determine the optimal discriminatory pricing strategy for a good being sold to networked consumers. Battaglini and Patacchini (2016) studies how interest groups should allocate campaign contributions when legislators sit on a social influence network. La'O and Tahbaz-Salehi (2020) studies how a monetary authority should implement optimal monetary policy in an economy with a general production network.

There is also work that examines more general observers intervening in more general interaction networks. Valente (2012) provides an overview of the possible types of network interventions, whether that involves targeting individuals, groups, or the underlying network structure. There are several papers whose objective is to design a network intervention that either maximizes or minimizes the aggregate action in the population. These papers include the following: Ballester et al. (2006), Demange (2017), Belhaj and Deroïan (2018), Belhaj and Deroïan (2019), and Belhaj et al. (2020). Essentially all of this work assumes that the external observer has full information.

Only very recently has the literature begun to consider very specific departures from full information for the decision-making observer. Parise and Ozdaglar (2019, 2020) consider large networks generated from a stochastic network formation model represented by a graphon, rather than an explicit underlying network topology. Brown and Patange (2020) similarly studies optimal interventions on networks when the adjacency matrix is unknown and drawn from a distribution. Galeotti et al. (2020) considers, as an extension, a setting in which the network-intervening planner has incomplete information about agents' standalone marginal returns; the planner's objective, in this work, is to design an intervention to perturb agents' standalone marginal returns to maximize utilitarian welfare.

The present work considers a new, important departure from full information for the decision-making observer in a networked economy. The observer simply does not know how the attributes of the economy sit on the network's nodes. The present work develops a comprehensive set of theoretical tools that a general observer intervening in a general interaction network can use. These tools can then be straightforwardly applied to specific types of observers to elucidate specific networked environments.

1.2 Outline of Paper

The plan for this paper is as follows. Section 2 begins by establishing the theoretical framework, describing the environment of the observer, and developing a comprehensive set of theoretical tools that the observer can use to better understand the economy and make decisions in the absence of full information. These theoretical tools are then applied in three different settings in Sections 3-5. Section 6 concludes.

2 Theoretical Framework

2.1 General Setup

We consider an economy with a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, which has N nodes. We index the network's nodes using the integer $i = 1, \ldots, N$. Accompanying the network is an $N \times 1$ vector of network constants:

$$\mathbf{w}_N = \begin{pmatrix} w_N(1) & \cdots & w_N(i) & \cdots & w_N(N) \end{pmatrix}^T \in \mathbb{R}^N.$$

We associate the constant $w_N(i)$ with network node i. The economy also separately has a collection of attributes. These attributes are indexed by the integer $j = 1, \ldots, N$, and together, they form the $N \times 1$ vector:

$$\mathbf{a}_N = \begin{pmatrix} a_N(1) & \cdots & a_N(j) & \cdots & a_N(N) \end{pmatrix}^T \in \mathbb{R}^N.$$

Each attribute sits on a network node. The $N \times 1$ vector

$$\boldsymbol{\chi}_N = \begin{pmatrix} \chi_{N1} & \cdots & \chi_{Ni} & \cdots & \chi_{NN} \end{pmatrix}^T,$$

a permutation of the indices $(1, \ldots, N)$, is the key that tells us how to map these attributes to network nodes. Attribute χ_{N1} maps to node 1, attribute χ_{Ni} maps to node i, and attribute χ_{NN} maps to node N. The scalar quantity of interest in this economy is:

$$z_N^* = \sum_{i=1}^{N} w_N(i)\, a_N(\chi_{Ni}).$$

Observers of this economy know $\mathbf{w}_N$, the vector of network constants, and they know $\mathbf{a}_N$, the set of attributes. However, they do not know how these attributes specifically sit on the nodes of the network, that is, they do not know $\boldsymbol{\chi}_N$. Without knowledge of $\boldsymbol{\chi}_N$, observers are unable to directly compute z_N^*.

Nonetheless, what the observers can do is construct a probability distribution of possible values for z_N^*. To generate this probability distribution, the observers must consider *all* $N!$ possible arrangements of the attributes on the network. The observers begin by allocating the attributes, in any manner desired, to the nodes on the network; given that allocation, the observers compute the scalar quantity. The observers then rearrange the attributes on the network, and for this new arrangement, they similarly compute the scalar quantity. The observers continue rearranging the attributes on the network and computing the scalar quantity until they have exhausted all $N!$ arrangements. The observers can then construct a probability distribution of possible values of z_N^*.

Mathematically, what is going on in the background is the following: there is an $N \times 1$ random vector:

$$\mathbf{X}_N = \begin{pmatrix} X_{N1} & \cdots & X_{NN} \end{pmatrix}^T.$$

Each realization of this random vector is a permutation of the indices $(1, \ldots, N)$. The permutation specifies how to map the economy's attributes to the network nodes; it is one possible arrangement. With $N!$ possible permutations of the indices $(1, \ldots, N)$, the random vector $\mathbf{X}_N$ specifies $N!$ possible arrangements of the attributes on the network, each one occurring with probability $\frac{1}{N!}$. The random variable

$$Z_N = \sum_{i=1}^{N} w_N(i)\, a_N(X_{Ni})$$

then generates the distribution of possible values for the scalar quantity z_N^*. The

CDF for this distribution is $G_{Z_N}(t) = \Pr[Z_N \leq t]$.

2.2 Alternative General Setup

We briefly present a parallel alternative setup which can be useful in certain applications. As before, we consider an economy with a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, which has N nodes indexed by $i = 1, \ldots, N$. Accompanying the network is an $N \times 1$ vector of network constants:

$$\mathbf{w}_N = \begin{pmatrix} w_N(1) & \cdots & w_N(i) & \cdots & w_N(N) \end{pmatrix}^T \in \mathbb{R}^N.$$

Constant $w_N(i)$ is associated with network node i. The economy also separately has a collection of attributes indexed by $j = 1, \ldots, N$:

$$\mathbf{a}_N = \begin{pmatrix} a_N(1) & \cdots & a_N(j) & \cdots & a_N(N) \end{pmatrix}^T \in \mathbb{R}^N.$$

Each attribute sits on a network node. The $N \times 1$ vector

$$\boldsymbol{\gamma}_N = \begin{pmatrix} \gamma_{N1} & \cdots & \gamma_{Nj} & \cdots & \gamma_{NN} \end{pmatrix}^T,$$

a permutation of the indices $(1, \ldots, N)$, instead tells us how to map *network nodes* to *attributes*. Node γ_{N1} maps to attribute 1, node γ_{Nj} maps to attribute j, and node γ_{NN} maps to attribute N. The scalar quantity of interest in this economy, z_N^*, is then:

$$z_N^* = \sum_{j=1}^{N} w_N(\gamma_{Nj})\, a_N(j)\,.$$

Observers of this economy know $\mathbf{w}_N$ and $\mathbf{a}_N$, but not $\boldsymbol{\gamma}_N$ (or $\boldsymbol{\chi}_N$). As a result, the observers are unable to compute z_N^*. The observers instead consider all possible arrangements of the attributes on the network. There is an $N \times 1$ random vector:

$$\mathbf{X}_N = \begin{pmatrix} X_{N1} & \cdots & X_{NN} \end{pmatrix}^T,$$

whose realizations are permutations of the indices $(1, \ldots, N)$. Each permutation specifies how to map the network nodes to the economy's attributes. The random variable

$$Z_N = \sum_{j=1}^{N} w_N\left(X_{Nj}\right) a_N\left(j\right)$$

generates the distribution of possible values for the scalar quantity z_N^*.

2.3 Characterizing the Distribution of Interest

We now proceed to characterize the probability distribution, $G_{Z_N}\left(t\right)$, and its statistical properties. Since the observers do not know $\boldsymbol{\chi}_N$ or $\boldsymbol{\gamma}_N$, they must rely on $G_{Z_N}\left(t\right)$ for making decisions that impact themselves and the agents in the economy. The statistical properties of Z_N are therefore crucial. To compute these statistical properties of Z_N, we first introduce the random variables W_N and A_N whose realizations are respectively $w_N\left(i\right)$ and $a_N\left(j\right)$. The first four population moments of W_N and A_N are defined as follows:

$$EW_N = \frac{1}{N}\sum_{i=1}^{N} w_N\left(i\right) \equiv \bar{w}_N \qquad EA_N = \frac{1}{N}\sum_{j=1}^{N} a_N\left(j\right) \equiv \bar{a}_N$$

$$\mathrm{Var}\, W_N = \frac{1}{N}\sum_{i=1}^{N}\left(w_N\left(i\right)-\bar{w}_N\right)^2 \qquad \mathrm{Var}\, A_N = \frac{1}{N}\sum_{j=1}^{N}\left(a_N\left(j\right)-\bar{a}_N\right)^2$$

$$\mathrm{Skew}\, W_N = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(w_N\left(i\right)-\bar{w}_N\right)^3}{\left(\mathrm{Var}\, W_N\right)^{3/2}} \qquad \mathrm{Skew}\, A_N = \frac{\frac{1}{N}\sum_{j=1}^{N}\left(a_N\left(j\right)-\bar{a}_N\right)^3}{\left(\mathrm{Var}\, A_N\right)^{3/2}}$$

$$\mathrm{Kurt}\, W_N = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(w_N\left(i\right)-\bar{w}_N\right)^4}{\left(\mathrm{Var}\, W_N\right)^{2}} \qquad \mathrm{Kurt}\, A_N = \frac{\frac{1}{N}\sum_{j=1}^{N}\left(a_N\left(j\right)-\bar{a}_N\right)^4}{\left(\mathrm{Var}\, A_N\right)^{2}}$$

We next present expressions that explicitly characterize the first four moments of Z_N:

Theorem 1 $EZ_N = N\bar{w}_N\bar{a}_N$.

The first moment of Z_N directly depends on the first moments of W_N and A_N. The higher the average network constant and/or the higher the value of the average attribute, the larger the first moment of Z_N. Also note that $EZ_N = N\bar{w}_N\bar{a}_N =$

$\frac{1}{N}\left(\sum_{i=1}^{N} w\,(i)\right)\left(\sum_{j=1}^{N} a\,(j)\right)$. If the sum of the attributes in the economy is held fixed, EZ_N will stay the same regardless of the actual values in $\{a\,(j)\}_{j=1}^{N}$. This feature will become important in this work's third application; observers to the economy will have a fixed, finite amount of resources for monitoring individual agents to reduce total crime, and the observers must choose how to allocate this fixed set of resources among agents. Here, the amount of resources used to monitor each agent is the attribute and the sum of the attributes in this economy is fixed. As a result, every single monitoring policy yields, on average, the same level of total crime, provided that all resources are expended. The mayor and the police must design the best monitoring policy given this important constraint.

Theorem 2 $\operatorname{Var} Z_N = \frac{1}{N-1}\left(N \operatorname{Var} W_N\right)\left(N \operatorname{Var} A_N\right)$.

The variance of Z_N directly depends on the variances of W_N and A_N. The greater the variation in the set network constants and/or attributes, the greater the variation in Z_N. Larger variance of Z_N means that it is more difficult to pinpoint z_N^*.

Theorem 3 $\operatorname{Skew} Z_N = \frac{(N-1)^{1/2}}{N-2} \operatorname{Skew} W_N \operatorname{Skew} A_N$.

The skewness of Z_N depends on the skewness of W_N, the skewness of A_N, and a factor that asymptotically tends toward $\frac{1}{\sqrt{N}}$. The greater the skewness in the set of network constants and/or attributes, the greater the skewness of Z_N. In certain applications, observers choose the set of attributes for the economy, and in so doing, they *design* the shape of the probability distribution, $G_{Z_N}(t)$. The skewness of the distribution then becomes one of those key statistical properties that observers consider during the design process precisely because, in unimodal settings, it dictates the extent to which the median of the probability distribution diverges from the mean.

Theorem 4 $\operatorname{Kurt} Z_N = \frac{3(N-1)}{N+1} + \frac{(N+1)(N-1)}{N(N-2)(N-3)}\left(\operatorname{Kurt} W_N - \frac{3(N-1)}{N+1}\right)\left(\operatorname{Kurt} A_N - \frac{3(N-1)}{N+1}\right)$.

The kurtosis of Z_N directly depends on the kurtosis of both W_N and A_N. The more heavy-tailed the set of network constants and/or attributes, the more heavy-tailed is Z_N. For large N, we have:

$$\text{Excess Kurt}\, Z_N \approx \frac{1}{N}\left(\text{Excess Kurt}\, W_N\right)\left(\text{Excess Kurt}\, A_N\right).$$

We then see that, for large N, excess kurtosis of Z_N directly depends on the excess kurtosis of W_N, the excess kurtosis of A_N, and a factor that tends toward $\frac{1}{N}$. The more leptokurtic the distribution of constants and/or attributes, the more leptokurtic is Z_N. For certain economies, the distribution of network constants or the distribution of attributes is naturally leptokurtic, and that increases the kurtosis of $G_{Z_N}(t)$. For example, in the second application that we study in this work, the distribution of attributes in the economy equals the wealth distribution, which has a stylized set of distributional features that namely includes leptokurtosis. Depending on the environment, heavy-tailedness of Z_N can affect an observer's decision-making.

Now that we have evaluated the first four moments of Z_N, a couple of quick remarks are in order. First, we see that each moment of Z_N depends on moments of W_N and A_N that are of the exact same order. Second, in many settings, and for all three applications explored in this work, the vector of constants is explicitly network-derived. The statistical features of $G_{Z_N}(t)$ then directly depend on the topological features of the network. Third, note that we could swap W_N and A_N in Theorems 1-4, and we would end up with the exact same expressions for the first four moments of Z_N. Why? The mapping of attributes to network nodes and then rearranging these attributes among the fixed set of nodes ends up being mathematically equivalent to the reverse, that is, mapping network nodes to attributes and then rearranging the network nodes among the fixed set of attributes.

We next identify the necessary and sufficient conditions for Z_N to be invariant to rearrangement, meaning that Z_N realizes the same value for every possible arrangement of attributes on the network:

Theorem 5 *Z_N is invariant to rearrangement if and only if (1) $w_N(i) = \bar{w}_N$ $\forall i \in \{1, \ldots, N\}$ and/or (2) $a_N(j) = \bar{a}_N$ $\forall j \in \{1, \ldots, N\}$. Then $Z_N = N\bar{w}_N\bar{a}_N$ with probability 1.*

When $Z_N = N\bar{w}_N\bar{a}_N$ with probability 1, the distribution $G_{Z_N}(t)$ becomes degenerate, and $Z_N = z_N^*$ with probability 1. Even though the observers of this economy do not know $\boldsymbol{\chi}_N$ or $\boldsymbol{\gamma}_N$, they are able to use their knowledge of $\mathbf{w}_N$ and $\mathbf{a}_N$ to compute z_N^* with certainty.

In the more general setting, we can also bound the support of Z_N:

Theorem 6 *Construct the* $N \times 1$ *vector* $\widetilde{\mathbf{w}}_N$ *by ordering the elements of* $\mathbf{w}_N$*, so that* $\widetilde{w}_N(i) \leq \widetilde{w}_N(k)$ *whenever* $i \leq k$*. Construct the* $N \times 1$ *vector* $\widetilde{\mathbf{a}}_N$ *by ordering the elements of* $\mathbf{a}_N$*, so that* $\widetilde{a}_N(j) \leq \widetilde{a}_N(k)$ *whenever* $j \leq k$*. Then*

$$\operatorname{minsupp} Z_N = \sum_{i=1}^{N} \widetilde{w}_N(i)\, \widetilde{a}_N(N+1-i) \quad \textit{and} \quad \operatorname{maxsupp} Z_N = \sum_{i=1}^{N} \widetilde{w}_N(i)\, \widetilde{a}_N(i)\,.$$

The lower and upper bounds on the support of Z_N determine the complete range of possible values for z_N^*.

We now study the asymptotic behavior of $Z_N(t)$ as the number of network constants and the number of attributes in the economy, both equal to N, tend to infinity. Define $\Phi(t)$ as the standard normal cumulative distribution function, and $\phi(t)$ as the standard normal probability density function. We arrive at the following central limit theorem-type result:

Theorem 7 *Suppose that*

$$\lim_{N\to\infty} \frac{\max_{1\leq i\leq N} \left(w_N(i) - \bar{w}_N\right)^2}{\sum_{i=1}^{N} \left(w_N(i) - \bar{w}_N\right)^2} = 0 \quad \textit{and} \quad \lim_{N\to\infty} \frac{\max_{1\leq j\leq N} \left(a_N(j) - \bar{a}_N\right)^2}{\sum_{j=1}^{N} \left(a_N(j) - \bar{a}_N\right)^2} = 0.$$

Then $\lim_{N\to\infty} G_{\frac{Z_N - EZ_N}{(\operatorname{Var} Z_N)^{1/2}}}(t) = \Phi(t)$ *if and only if* $\lim_{N\to\infty} \frac{1}{N} \sum \sum_{\left|\delta_{N,ij}\right| > \tau} \delta_{N,ij}^2 = 0$ *for any* $\tau > 0$*, where*

$$\delta_{N,ij} = \frac{\left(w_N(i) - \bar{w}_N\right)\left(a_N(j) - \bar{a}_N\right)}{\left[\frac{1}{N} \sum_{i=1}^{N} \left(w_N(i) - \bar{w}_N\right)^2 \sum_{j=1}^{N} \left(a_N(j) - \bar{a}_N\right)^2\right]^{1/2}}.$$

As $N \to \infty$, Z_N tends to a normal distribution provided that the elements in the vectors $\mathbf{w}_N$ and $\mathbf{a}_N$ are both asymptotically well-behaved.

Theorem 7 allows the observers to strongly approximate $G_{Z_N}(t)$ as a normal distribution with mean EZ_N and variance $\operatorname{Var} Z_N$ in the limit of large N. This central limit theorem-type result is indeed informative, but it does have its limits. The next theoretical result overcomes these limits; it shows how to greatly enrich the mathematical description of the probability distribution, $G_{Z_N}(t)$, by offering an asymptotic expansion:

Theorem 8 *Define* $\kappa_{3N} = \frac{1}{\sqrt{N}} \operatorname{Skew} W_N \operatorname{Skew} A_N$ *and define* $\kappa_{4N} = \frac{1}{N} \operatorname{Kurt} W_N \operatorname{Kurt} A_N - \frac{3}{N} \operatorname{Kurt} W_N - \frac{3}{N} \operatorname{Kurt} A_N + \frac{3}{N}$*. Then, the following relation approximately holds as* $N \to \infty$*:*

$$\sup_{t \in \mathbb{R}} \left| G_{\frac{Z_N - EZ_N}{(\operatorname{Var} Z_N)^{1/2}}}(t) - K_N(t) \right| = o\left(N^{-1}\right), \text{ where}$$

$$K_N(t) = \Phi(t) - \phi(t) \left[\frac{\kappa_{3N}}{6} \left(t^2 - 1\right) + \frac{\kappa_{4N}}{24} \left(t^3 - 3t\right) + \frac{\kappa_{3N}^2}{72} \left(t^5 - 10t^3 + 15t\right) \right].$$

The asymptotic expansion is given by the function $K_N(t)$. According to this theorem, $G_{\frac{Z_N - EZ_N}{(\operatorname{Var} Z_N)^{1/2}}}(t) \approx K_N(t)$ and therefore $G_{Z_N}(t) \approx K_N\left(\frac{t - EZ_N}{(\operatorname{Var} Z_N)^{1/2}}\right)$. The first term of this asymptotic expansion is the normal distribution and the additional terms represent deviations away from the normal distribution. Note that, as $N \to \infty$, these additional terms tend to zero. For finite N, the extent to which $G_{Z_N}(t)$ deviates from a normal distribution depends on the values of κ_{3N} and κ_{4N}. The term, κ_{3N}, is approximately equal to the expression for Skew Z_N in Theorem 3, and the term, κ_{4N}, is nearly equal to the expression for Excess Kurt Z_N listed below Theorem 4. This asymptotic expansion is a critical tool that enables the observers to compute the distribution of possible values for z_N^*, thereby helping observers better understand their economy. The observers in all three applications studied in this work make extensive use of the asymptotic expansion.

3 Peer Networks and Education Outcomes

3.1 Model

We consider a population of N students in a school. Students' decisions are guided by those of their peers. We therefore have a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ of peer influence. Both the students and the network nodes are indexed by the integer $i = 1, \ldots, N$, with student i sitting on node i. The $N \times N$ adjacency matrix $\mathbf{G}$ captures students' relations with peers. $[\mathbf{G}]_{ik} = 1$ if student i considers student k to be a peer (i.e., friend), and otherwise $[\mathbf{G}]_{ik} = 0$. Peer relations need not be reciprocal. n_i is the out-degree for node i; it is the total number of individuals that student i considers to be his/her peers. We set $[\mathbf{G}]_{ii} = 0 \ \forall i \in \{1, \ldots, N\}$.

We also have a collection of demographic characteristics for the student population. $\mathbf{c}(j) = \begin{pmatrix} c_j^1 & \cdots & c_j^h & \cdots & c_j^H \end{pmatrix}^T$ is a vector of H demographic characteristics for a single student. The set $\{\mathbf{c}(1), \ldots, \mathbf{c}(j), \ldots, \mathbf{c}(N)\}$ is then the collection of demographic characteristics for the entire student body. The vector $\boldsymbol{\chi}_N$ tells us how to assign demographic characteristics to a particular student; student 1 has demographic characteristics $\mathbf{c}(\chi_{N1})$, student i has demographic characteristics $\mathbf{c}(\chi_{Ni})$, and student N has demographic characteristics $\mathbf{c}(\chi_{NN})$.

Define $e_N(i) \geq 0$ to be the effort put forth by student i. Similar to Calvó-Armengol, Patacchini, and Zenou (2008), this effort is chosen to maximize the following utility function:

$$u_i(\mathbf{e}_N, \mathbf{G}) = [\mu n_i + a_N(\chi_{Ni})] \, e_N(i) - \frac{1}{2}(e_N(i))^2 + \phi \sum_{k=1}^{N} [\mathbf{G}]_{ik} \, e_N(i) e_N(k),$$

where $\phi > 0$, $\mu > 0$, and n_i is the number of out-edges for student i. We assume that

$$a_N(\chi_{Ni}) = \boldsymbol{\beta}^T \mathbf{c}(\chi_{Ni}),$$

where $\mathbf{c}(\chi_{Ni})$ is the vector of demographic characteristics for student i, and $\boldsymbol{\beta}$ is an $H \times 1$ vector of parameters. For simplicity, we exclude contextual effects (Manski, 1993).

Student utility is represented here as a linear-quadratic function with continuous efforts and strategic complementarities ($\phi > 0$). We compute the Nash equilibrium of this game in which students simultaneously choose effort levels. Defining $\rho(\mathbf{G})$ as the spectral radius of matrix $\mathbf{G}$, we have:

Theorem 9 *Suppose that $\phi\rho(\mathbf{G}) < 1$. Then the unique and interior Nash equilibrium in pure strategies is given by*

$$\mathbf{e}_N^* = (\mathbf{I} - \phi\mathbf{G})^{-1} (\mu\mathbf{G}\mathbf{1} + \mathbf{a}_N(\boldsymbol{\chi}_N)),$$

where

$$\mathbf{a}_N\left(\boldsymbol{\chi}_N\right)=\begin{pmatrix}\boldsymbol{\beta}^T\mathbf{c}\left(\chi_{N1}\right) & \ldots & \boldsymbol{\beta}^T\mathbf{c}\left(\chi_{Ni}\right) & \ldots & \boldsymbol{\beta}^T\mathbf{c}\left(\chi_{NN}\right)\end{pmatrix}^T.$$

The condition $\phi\rho\left(\mathbf{G}\right)<1$, discussed by Debreu and Herstein (1953, Theorem 3), establishes an upper bound on network complementarities. It prevents the positive feedback loops induced by such complementarities from escalating without bound and causing student efforts to never reach an equilibrium level.

Note that the equilibrium profile of student effort,

$$\mathbf{e}_N^*=\left(\mathbf{I}-\phi\mathbf{G}\right)^{-1}\left(\mu\mathbf{G}\mathbf{1}+\mathbf{a}_N\left(\boldsymbol{\chi}_N\right)\right)=\sum_{k=0}^{\infty}\left(\phi\mathbf{G}\right)^k\left(\mu\mathbf{G}\mathbf{1}+\mathbf{a}_N\left(\boldsymbol{\chi}_N\right)\right),$$

is equal to a weighted Katz-Bonacich centrality measure. The higher the student's weighted Katz-Bonacich centrality, the greater that student's equilibrium effort because the student is the recipient of a larger amount of direct and indirect peer influence.

We are interested in student performance. As in Calvó-Armengol et al. (2009) and Jackson and Zenou (2015), the academic performance for the entire population of students, $\mathbf{z}_N^*$, is:

$$\mathbf{z}_N^*=\mathbf{e}_N^*+\mathbf{a}_N\left(\boldsymbol{\chi}_N\right).$$

We see that a student's demographic attributes and effort level shape that student's performance. Plugging in the expression for $\mathbf{e}_N^*$ yields the following:

$$\mathbf{z}_N^*=\left(\mathbf{I}-\phi\mathbf{G}\right)^{-1}\left(\mu\mathbf{G}\mathbf{1}\right)+\left[\left(\mathbf{I}-\phi\mathbf{G}\right)^{-1}+\mathbf{I}\right]\mathbf{a}_N\left(\boldsymbol{\chi}_N\right);$$

academic performance for student i is:

$$z_{N,i}^*=\left[\left(\mathbf{I}-\phi\mathbf{G}\right)^{-1}\right]_{i*}\left(\mu\mathbf{G}\mathbf{1}\right)+\left[\left(\mathbf{I}-\phi\mathbf{G}\right)^{-1}+\mathbf{I}\right]_{i*}\mathbf{a}_N\left(\boldsymbol{\chi}_N\right),$$

where $i*$ designates the i^{th} row of the corresponding matrix.

Now, the principal would like to assess student performance in the absence of any educational interventions. The principal knows the topology of the peer influence network and where each student sits on the network from observations of student interactions by faculty and staff. Moreover, the principal knows the demographic characteristics of students from an anonymous survey of the student body.[2] Specifically, the principal knows the vectors $\mathbf{c}(1), \ldots, \mathbf{c}(N)$ of demographic characteristics. With knowledge of $\boldsymbol{\beta}$, the principal also knows $\mathbf{a}_N$, which is the vector of attributes in this application. However, since the survey is anonymous, the principal does not know $\boldsymbol{\chi}_N$, that is, he does not know how to assign these vectors of demographic characteristics to specific students. The principal therefore considers all possible ways to arrange the attributes on the nodes of the peer influence network.

This setting maps to the theoretical environment in Section 2. For each student i, we have:

$$constant_{N,i} = \left[(\mathbf{I} - \phi \mathbf{G})^{-1} \right]_{i*} (\mu \mathbf{G} \mathbf{1}),$$

$$\mathbf{w}_{N,i}^T = \left[(\mathbf{I} - \phi \mathbf{G})^{-1} + \mathbf{I} \right]_{i*},$$

and

$$\mathbf{a}_N = \begin{pmatrix} a_N(1) & \cdots & a_N(N) \end{pmatrix}^T = \begin{pmatrix} \boldsymbol{\beta}^T \mathbf{c}(1) & \cdots & \boldsymbol{\beta}^T \mathbf{c}(N) \end{pmatrix}^T.$$

Student i's performance under the mapping $\boldsymbol{\chi}_N$ is:

$$z_{N,i}^* = constant_{N,i} + \sum_{k=1}^{N} w_{N,i}(k)\, a_N(\chi_{Nk}) \quad \forall i \in \{1, \ldots, N\}.$$

The principal knows $constant_{N,i}$ and $\mathbf{w}_{N,i}$, $\forall i \in \{1, \ldots, N\}$, from knowledge of the network. He also knows $\mathbf{a}_N$ from the anonymous survey. However, since the principal

[2] Such anonymous surveys are fairly common. Some examples of these surveys include the Illinois Youth Survey, the Texas School Survey, the California Healthy Kids Survey, the Florida Youth Survey, and the Connecticut School Health Survey. Often, an individual school report gets generated from the survey, which is then made available directly to the school principal. The demographic characteristics collected from the survey are key for helping the principal understand the student body and its capacity to perform academically. Such demographic characteristics include the following: whether the student is in foster care, identifies as homeless, is an English language learner, has a disability, qualifies for food stamps through his/her family, drinks alcohol, uses illicit drugs, uses tobacco, and/or regularly participates in extracurricular activities. The survey also often inquires about the student's parents, such as their highest level of schooling, and whether they are migrants, immigrants, and/or connected to the military.

does not know $\boldsymbol{\chi}_N$, we introduce the random variable $Z_{N,i}$, which captures possible levels of academic performance for student i:

$$Z_{N,i} = constant_{N,i} + \sum_{k=1}^{N} w_{N,i}(k)\, a_N(X_{Nk}) \quad \forall i \in \{1, \ldots, N\},$$

with $\mathbf{X}_N$ an $N \times 1$ random vector whose realizations are permutations of the indices $(1, \ldots, N)$.

For any given student, the principal can compute $G_{Z_{N,i}}(t)$, the distribution of possible levels of academic performance for student i sitting on node i of the network. Student i fails if his composite grade is below $\underline{f}$; thus, the probability that student i fails is $G_{Z_{N,i}}\left(\underline{f}\right)$. The principal is interested in the expected failure rate in the student body in the absence of any educational interventions:

$$expected\ failure\ rate = \frac{1}{N} \sum_{i=1}^{N} G_{Z_{N,i}}\left(\underline{f}\right).$$

3.2 Student Performance in a Sample School

We consider a population of 120 high school students from Lycée Thiers in Marseilles, France, who are undertaking their "classes préparatoires" (Mastrandrea et al., 2015). In December 2013, these students were asked to keep contact diaries. At the end of four days, these students listed all of the individuals that they encountered over that time period.[3] We consider all individuals recorded in a student's contact diary to be that students' peers.

The resulting network, depicted in Figure 2, has 120 nodes, 501 edges, and a density of 0.0351. We set $\phi = 0.10$ and $\mu = 0.10$. We are interested in the academic performance of student 1 and student 3. We find that $constant_{N,1} = 0.92$ and $constant_{N,3} = 0.15$. We also compute $\mathbf{w}_{N,1}$ and $\mathbf{w}_{N,3}$. The left panel of Figure 3 plots the counter-cumulative distribution function of elements in $\mathbf{w}_{N,1}$. This distribution has a mean of 0.024, a standard deviation of 0.19, skewness of 10.55, and kurtosis of 114.14, with a minimum of zero and a maximum of 2.03. The right panel of Fig-

[3] The dataset is available at http://www.sociopatterns.org/datasets/high-school-contact-and-friendship-networks/.

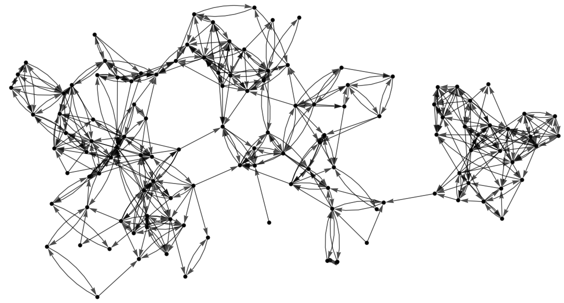

Figure 2: The network of high school students.

ure 3 plots the counter-cumulative distribution function of elements in $\mathbf{w}_{N,3}$. This distribution has a mean of 0.018, a standard deviation of 0.18, skewness of 10.77, and kurtosis of 117.35, with a minimum of zero and a maximum of 2.00.

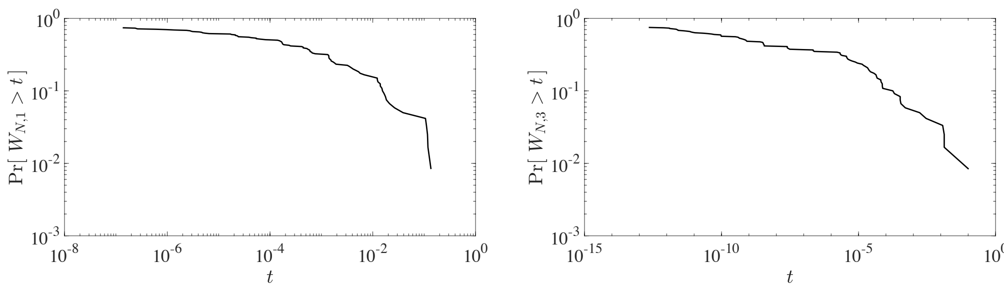

Figure 3: Counter-cumulative distribution function of network constants for student 1 (left) and student 3 (right).

As for $\mathbf{a}_N$, purely out of simplicity, we draw its elements from a normal distribution with a mean of 2 and a standard deviation of 0.5. Now that we have computed $constant_{N,i}$ and $\mathbf{w}_{N,i}$ for $i = 1, 3$ as well as $\mathbf{a}_N$, we can study these two students' potential levels of academic performance.

The students are assessed on a 15-point scale, with failure occurring if the student's grade is at or below 5. Student 1's grade on average equals 7.00 (follows from Theorem 1, after adding $constant_{N,1}$ to the existing expression), with a standard deviation of 1.05 (follows from Theorem 2), skewness of -0.013 (follows from Theorem 3), and kurtosis of 2.42 (follows from Theorem 4). Student 1's minimum possible grade is 3.80, and student 1's maximum possible grade is 10.46 (follows from Theorem 6, after adding $constant_{N,1}$ to the existing expressions). The probability that student 1 fails is 2.77 percent (follows from Theorem 8). The left panel of Figure 4 plots the distribution of academic performance for student 1. The threshold for failure, $\underline{f}$, is positioned at the grade of 5. We see that student 1 is very likely to pass. Student 3's grade on average, however, equals 4.62, with a standard deviation of 1.03, skewness of -0.013, and kurtosis of 2.41. Student 3's minimum possible grade is 2.06, and student 3's maximum possible grade is 7.48. The probability that student 3 fails is 63.20 percent. The right panel of Figure 4 plots the distribution of academic performance for student 3. We see that student 3 is more likely to fail than pass. The expected failure rate for the entire population of students is 21.36 percent. The principal can then use all of this information to design appropriate educational interventions.

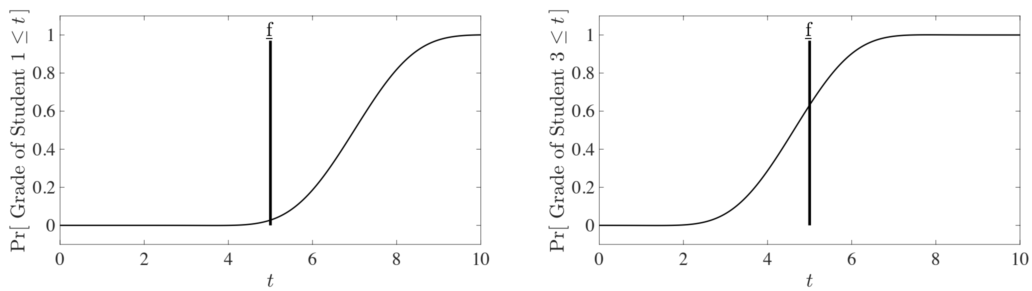

Figure 4: Cumulative distribution function of academic performance for student 1 (left) and student 3 (right).

4 Tax Revenue in an Environment with Consumption Network Effects

4.1 Model

We have: (1) a population of N consumers and (2) a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with N nodes. The N consumers are indexed by the integer $j = 1, \ldots, N$ with consumer j having a wealth attribute $a_N(j) \equiv a_{Nj}$. The network nodes are separately indexed by the integer $i = 1, \ldots, N$. Corresponding to the network is an $N \times N$ adjacency matrix $\mathbf{G}$ that captures interconnections among nodes. Matrix element $[\mathbf{G}]_{k\ell} = 1$ if there is a linkage from node k to node ℓ, and otherwise matrix element $[\mathbf{G}]_{k\ell} = 0$. We set $[\mathbf{G}]_{kk} = 0 \; \forall k \in \{1, \ldots, N\}$.

The population of consumers sits on the nodes of the network. The $N \times 1$ vector $\boldsymbol{\gamma}_N = \begin{pmatrix} \gamma_{N1} & \cdots & \gamma_{Nj} & \cdots & \gamma_{NN} \end{pmatrix}^T$ carries out this mapping; it maps network node γ_{Nj} to consumer j. For any given mapping $\boldsymbol{\gamma}_N$, there exists an $N \times N$ permutation matrix $\mathbf{P}$ that transforms the original adjacency matrix $\mathbf{G}$ to a new adjacency matrix $\mathbf{PGP}^T$. Matrix element $\left[\mathbf{PGP}^T\right]_{k\ell}$ indicates the existence of a linkage from consumer k to consumer ℓ under the mapping $\boldsymbol{\gamma}_N$. The set $\mathcal{N}(j)$ of network neighbors for consumer j consists of those indices $k \in \{1, \ldots, N\}$ for which $\left[\mathbf{PGP}^T\right]_{jk} = 1$. The number of neighbors for consumer j is $n_j = |\mathcal{N}(j)|$.

We next describe the economic environment. There are two consumption goods: x and y. Let x_{Nj} denote the consumption of good x by agent j and let y_{Nj} denote the consumption of good y by agent j. Each consumer j cares about his own consumption of good x, his own consumption of good y, and the consumption of good y by his immediate neighbors, that is, $\{y_{Nk}\}_{k \in \mathcal{N}(j)}$. Similar to Ghiglino and Goyal (2010), the utility of consumer j takes the following Cobb-Douglas form:

$$u_j\left(x_{Nj}, \Phi_j\left(y_{Nj}, \{y_{Nk}\}_{k \in \mathcal{N}(j)}\right)\right) = x_{Nj}^{\sigma}\left(\Phi_j\left(y_{Nj}, \{y_{Nk}\}_{k \in \mathcal{N}(j)}\right)\right)^{1-\sigma},$$

where $0 < \sigma < 1$ and

$$\Phi_j\left(y_{Nj}, \{y_{Nk}\}_{k \in \mathcal{N}(j)}\right) = y_{Nj} + \alpha n_j \left[y_{Nj} - \frac{1}{n_i} \sum_{k \in \mathcal{N}(j)} y_{Nk}\right]$$

with $\alpha > 0$. When $\mathcal{N}(j) = \{\}$, we set $\Phi_j = y_{Nj}$. When agent j's neighbors all consume the same quantity of good y as agent j, there is no net externality and $\Phi_j\left(y_{Nj}, \{y_{Nk}\}_{k \in \mathcal{N}(j)}\right) = y_{Nj}$. For agent j, subutility Φ_j is increasing in own consumption of good y, and it is decreasing in neighbors' consumption of good y. Agents care about their own consumption of good y relative to the average consumption of their neighbors. The magnitude of this effect linearly scales with the size of an agent's neighborhood, n_j. Setting $\alpha > 0$ means that an agent's utility is negatively affected when average consumption of one's neighbors exceeds own consumption. The more that one's neighbors consume, the worse off that agent is. There is pressure to keep up with one's neighbors.

Set good x as the numéraire good, and let p be the price of good y. There is a government-imposed consumption tax on good y. Defining T as the tax rate, the tax revenue collected from consumer j is Tpy_{Nj}. The optimization problem for consumer j is then as follows:

$$\max_{x_{Nj}, y_{Nj}} u_j \left(x_{Nj}, \Phi_j \left(y_{Nj}, \{y_{Nk}\}_{k \in \mathcal{N}(j)} \right) \right) \quad \text{s.t.} \quad x_{Nj} + p\,(1+T)\, y_{Nj} = a_{Nj}.$$

For each consumer j, the pair $\left(x^*_{Nj}, y^*_{Nj} \right)$ solves the optimization problem. Define $\mathbf{I}$ as the $N \times N$ identity matrix. Also define the $N \times N$ matrix $\mathbf{G}^N$ for which $\left[\mathbf{G}^N \right]_{jk} = \frac{1}{1+\alpha[\mathbf{G}]_{j*}\mathbf{1}}$ if $[\mathbf{G}]_{jk} = 1$ and otherwise $\left[\mathbf{G}^N \right]_{jk} = 0$. ($[\mathbf{G}]_{j*}$ is equal to the j^{th} row of matrix $\mathbf{G}$.) The $N \times 1$ vector $\mathbf{y}^*_N$ is the population vector of equilibrium good-y consumption, and z^*_N is total tax revenue raised given the mapping $\boldsymbol{\gamma}_N$.

Theorem 10 *Given the mapping $\boldsymbol{\gamma}_N$, total tax revenue raised is:*

$$z^*_N = \frac{(1-\sigma)\,T}{1+T} \mathbf{1}^T \left[\mathbf{I} - \alpha\sigma\mathbf{G}^N \right]^{-1} \mathbf{P}^T \mathbf{a}_N.$$

The expression for z^*_N contains the term $\left[\mathbf{I} - \alpha\sigma\mathbf{G}^N \right]^{-1}$. With $\sigma < 1$ by assumption, $\mathbf{I} - \alpha\sigma\mathbf{G}^N$ is invertible.

The observer to this economic system, the government, would like to raise a minimum amount of tax revenue $\underline{r}$, by levying a tax on individual consumers' purchases of good y. The government knows each agent's wealth and consumption preferences, and it has a sense of the topology of the underlying interaction network. The government does not know how the agents sit on the network. The government therefore has no knowledge of z^*_N, the tax revenue that would result in the end from imposing tax rate T.

This setting maps to the theoretical environment in Section 2. The vector of wealth attributes is $\mathbf{a}_N$, and given Theorem 10, the vector of network constants in this economy is:

$$\mathbf{w}_N = \frac{(1-\sigma)\,T}{1+T} \left[\mathbf{I} - \alpha\sigma \left(\mathbf{G}^N \right)^T \right]^{-1} \mathbf{1}.$$

Accordingly, for a given mapping $\boldsymbol{\gamma}_N$, total tax revenue raised is $z^*_N = \mathbf{w}^T_N \mathbf{P}^T \mathbf{a}_N$. For a different mapping $\boldsymbol{\gamma}'_N$, the permutation matrix $\mathbf{P}$ adjusts and the wealth attributes get rearranged on the network. The government knows $\mathbf{a}_N$ and $\mathbf{w}_N$, but it has no knowledge of the mapping $\boldsymbol{\gamma}_N$ or $\boldsymbol{\chi}_N$.

What the government can do is consider random variable Z_N from Section 2. Each realization of random variable Z_N is tax revenue raised for a particular arrangement of wealth on the network's nodes. $G_{Z_N}(t)$ is then the distribution of all possible levels of tax revenue raised for a given tax rate after considering all possible allocations of the wealth attribute to the network's nodes. The government chooses a tax rate T so that $G_{Z_N}(\underline{r}) = \Pr[Z_N \leq \underline{r}] = 0.05$, that is, it chooses a tax rate so that there is only a five-percent probability that the tax revenue raised is less than the minimum desired threshold, $\underline{r}$. The government utilizes the asymptotic expansion from Theorem 8 to make this calculation.

4.2 Setting the Optimal Tax Rate in a Sample Economy

Let us begin by constructing the network of neighbors. We use the well-known Southern Women dataset (Davis et al., 1941), which features $N = 18$ women observed over a nine-month period. During this time window, there were 14 informal social events that different subsets of women attended. We consider two of the Southern women to be connected if they attended at least two events together. Figure 5 depicts the resulting interaction network, with its 18 nodes, 194 edges, and network density of 0.6340.

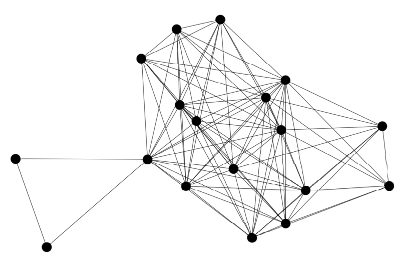

Figure 5: The network of Southern women.

We next endow each of the women with wealth so that they can consume goods x and y. We draw each woman's wealth from a Pareto distribution with a scale parameter of 100 and a shape parameter of 2. The resulting empirical distribution of agent wealth, $G_{A_N}(t)$, has a mean of 184.86 units, a standard deviation of 129.51

units, skewness equal to 1.95, and kurtosis equal to 5.38. The heavy-tailedness of the wealth distribution is evident in the left panel of Figure 6, which plots the counter-cumulative distribution function of consumer wealth. The lowest wealth of any woman in the population equals 105.54 units, and the highest wealth of any woman in the population equals 527.86 units.

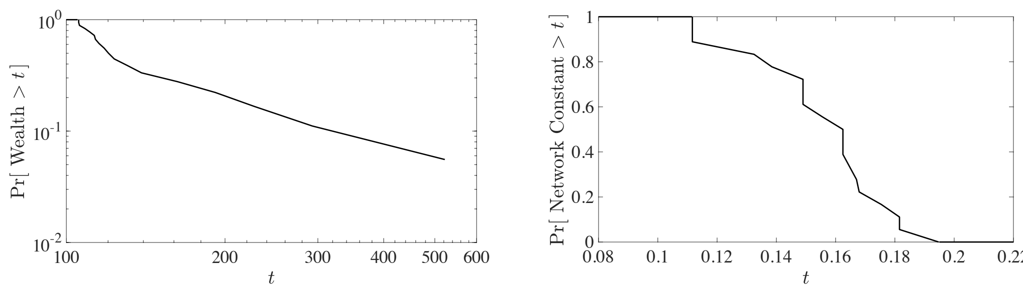

Figure 6: Empirical counter-cumulative distribution function of consumer wealth (left), and empirical counter-cumulative distribution function of network constants when tax rate $T = 0.2254$ (right).

We set the parameters $\alpha = 0.5$ and $\sigma = 0.5$, and we set the minimum desired tax revenue to be raised by the government equal to 500 units: $\underline{r} = 500$. Given these primitives and consumers' preferences, we are then able to compute the distribution of possible levels of tax revenue raised in this economy for any specified tax rate, T.

Let us set the tax rate equal to 22.54 percent. By fixing the tax rate, we can compute the set of network constants in $\mathbf{w}_N$. These network constants have a mean of 0.16, a variance of 0.022, skewness of -0.56, and kurtosis of 2.83. The smallest network constant is 0.11, and the largest network constant is 0.19. The counter-cumulative distribution function of network constants is plotted in the right panel of Figure 6.

With $\mathbf{w}_N$ and $\mathbf{a}_N$ both constructed, we can now employ the theoretical results from Section 2 to characterize the resulting tax revenue distribution. When the tax rate equals 22.54 percent, the tax revenue distribution has a mean of 521.22 units (follows from Theorem 1), a standard deviation of 12.38 units (follows from Theorem 2), skewness of -0.28 (follows from Theorem 3), and kurtosis of 2.71 (follows from Theorem 4). The minimum possible amount of tax revenue that can be raised is 476.07 units, and the maximum possible amount of tax revenue that can be raised is 556.51 units (follows from Theorem 6). The probability that the tax revenue raised will exceed 500 units is 95 percent (follows from Theorem 8). The left panel of Figure 7 plots, via asymptotic expansion, the cumulative distribution function of tax revenue raised when the tax rate equals 22.54 percent. We see that, for this tax rate, there is a five-percent probability that tax revenue raised will not meet desired revenue, $\underline{r}$.

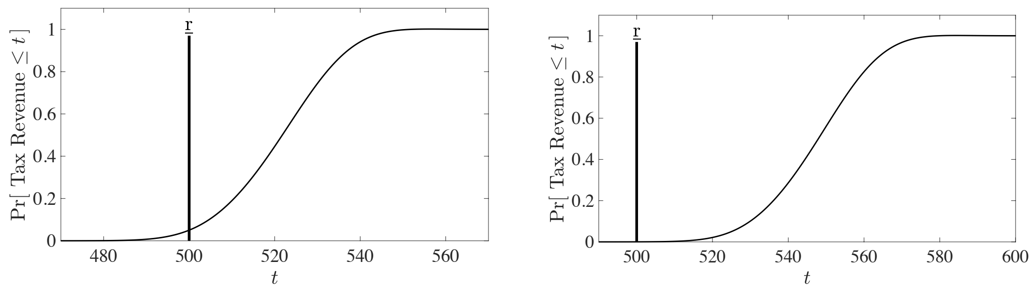

Figure 7: Cumulative distribution function of tax revenue when tax rate $T = 0.2254$ (left) and tax rate $T = 0.2394$ (right).

If the government wants to instead ensure that tax revenue raised exceeds $\underline{r} = 500$ units with probability 1, it will need to set a higher tax rate. The minimum necessary tax rate would equal 23.94 percent, which is six percent greater than the 22.54-percent tax rate computed above. The right panel of Figure 7 plots, via asymptotic expansion, the cumulative distribution function of tax revenue raised when the tax rate is 23.94 percent. Tax revenue exceeds the minimum required amount, $\underline{r} = 500$, with probability 1.

5 Managing Crime

5.1 Model

We have a city with a subpopulation of N individuals potentially engaging in criminal activity. These individuals are indexed by the integer $j = 1, \ldots, N$. Meanwhile, there is a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with N nodes indexed by the integer $i = 1, \ldots, N$. Accompanying this network is the $N \times N$ adjacency matrix $\mathbf{G}$. $[\mathbf{G}]_{k\ell} = 1$ if there is a linkage between nodes k and ℓ, and otherwise $[\mathbf{G}]_{k\ell} = 0$.

The $N \times 1$ vector $\boldsymbol{\gamma}_N = \begin{pmatrix} \gamma_{N1} & \cdots & \gamma_{Nj} & \cdots & \gamma_{NN} \end{pmatrix}^T$ specifies how these individuals sit on the nodes of the network. Node γ_{N1} maps to agent 1, node γ_{Nj}

maps to agent j, and node γ_{NN} maps to agent N. Corresponding to the mapping $\boldsymbol{\gamma}_N$ is an $N \times N$ permutation matrix $\mathbf{P}$. We are then able to transform the original adjacency matrix $\mathbf{G}$ to a new $N \times N$ adjacency matrix $\mathbf{PGP}^T$. Whereas $[\mathbf{G}]_{k\ell} = 1$ indicates the existence of a linkage from node k to node ℓ, $\left[\mathbf{PGP}^T\right]_{k\ell} = 1$ indicates the existence of a linkage from individual k to individual ℓ under the mapping $\boldsymbol{\gamma}_N$. We have $\left[\mathbf{PGP}^T\right]_{kk} = 0 \ \forall k \in \{1, \ldots, N\}$. We define $\mathcal{N}(j)$ as the set of network neighbors for individual j; $\mathcal{N}(j)$ consists of those indices $k \in \{1, \ldots, N\}$ for which $\left[\mathbf{PGP}^T\right]_{jk} = 1$. A linkage from individual k to individual ℓ means that individual ℓ exerts criminal influence over individual k. Individual ℓ might be individual k's superior in an organized crime environment or perhaps a partner in crime. Linkages need not be reciprocal.

Define $e_N(j) \geq 0$ to be the criminal effort put forth by individual j and $\mathbf{e}_N$ to be the population's effort profile. Similar to Liu et al. (2014), this effort is chosen to maximize the following utility function:

$$
\begin{aligned}
u_j\left(\mathbf{e}_N, \mathbf{PGP}^T\right) = (\pi_N(j) + \xi)\, e_N(j) - \frac{1}{2}\left(e_N(j)\right)^2 - p_N(j)\, f e_N(j) \\
+ \phi \sum_{k=1}^{N} \left[\mathbf{PGP}^T\right]_{jk} e_N(j)\, e_N(k),
\end{aligned}
$$

where $\phi > 0$. The first term, $(\pi_N(j) + \xi)\, e_N(j)$, represents the proceeds from carrying out a crime. $\pi_N(j)$ is a term that depends on the characteristics of individual j, such as gender, age, and education level. The vector of H demographic characteristics for individual j is $\mathbf{c}(j) = \begin{pmatrix} c_j^1 & \cdots & c_j^h & \cdots & c_j^H \end{pmatrix}^T$. With the $H \times 1$ vector of parameters, $\boldsymbol{\beta}$, common to all individuals, we have $\pi_N(j) = \boldsymbol{\beta}^T \mathbf{c}(j)$. The parameter ξ, meanwhile, depends on characteristics that are common to all individuals, such as the wealth of the neighborhood in which the criminal activity is being committed. The vector of L common characteristics is $\mathbf{d} = \begin{pmatrix} d^1 & \cdots & d^\ell & \cdots & d^L \end{pmatrix}^T$, and the $L \times 1$ vector of parameters is $\boldsymbol{\delta}$, so that $\xi = \boldsymbol{\delta}^T \mathbf{d}$.

There is a direct quadratic cost to committing crime, $\frac{1}{2}\left(e_N(j)\right)^2$, and there is an additional cost of being caught, $p_N(j)\, f e_N(j)$. This latter term equals the probability of being caught, $p_N(j) \in (0, 1)$, multiplied by the corresponding punishment or fine, $fe_N(j)$. Note that this cost increases with own effort; as one's involvement in the crime increases, so does the severity of the punishment. The final term, $\phi \sum_{k=1}^{N} \left[\mathbf{PGP}^T\right]_{jk} e_N(j)\, e_N(k)$, captures the network effect. There are strategic complementarities in effort among connected individuals. With the absence of formal criminal training, network linkages provide the means by which individuals learn criminal behavior as well as the hard and soft skills needed to successfully engage in criminal activity.

To purely examine the role of the network, we are going to assume that individuals are identical except for their network position, so we set $\pi_N(j) = \pi \ \forall j \in \{1, \ldots, N\}$. We proceed to compute the Nash equilibrium of this game in which individuals simultaneously choose their criminal effort levels. Defining $\rho\left(\mathbf{PGP}^T\right) = \rho(\mathbf{G})$ as the spectral radius of matrix $\mathbf{PGP}^T$, we can solve for the equilibrium effort profile, $\mathbf{e}_N^*$, and aggregate criminal effort, e_{agg}^*:

Theorem 11 *Suppose that $\phi\rho(\mathbf{G}) < 1$. Then the unique and interior Nash equilibrium in pure strategies is given by:*

$$\mathbf{e}_N^* = \mathbf{P}(\mathbf{I} - \phi\mathbf{G})^{-1}\mathbf{P}^T\left[(\pi + \xi)\mathbf{1} - \mathbf{p}_N f\right]$$

and

$$e_{agg}^* = \mathbf{1}^T(\mathbf{I} - \phi\mathbf{G})^{-1}\mathbf{P}^T\left[(\pi + \xi)\mathbf{1} - \mathbf{p}_N f\right].$$

Note that the equilibrium profile of criminal effort is equal to a weighted Katz-Bonacich centrality measure:

$$\mathbf{e}_N^* = \left(\mathbf{I} - \phi\mathbf{PGP}^T\right)^{-1}\left[(\pi + \xi)\mathbf{1} - \mathbf{p}_N f\right] = \sum_{k=0}^{\infty}\left(\phi\mathbf{PGP}^T\right)^k\left[(\pi + \xi)\mathbf{1} - \mathbf{p}_N f\right]$$

The higher an individual's weighted Katz-Bonacich centrality, the greater that individual's criminal effort because the individual is the recipient of a larger amount of direct and indirect criminal influence from other individuals.

Now, these individuals are being monitored by the police. The mayor and the police select a monitoring policy $\mathbf{m}_N = \begin{pmatrix} m_N(1) & \cdots & m_N(j) & \cdots & m_N(N) \end{pmatrix}^T \geq \mathbf{0}$, in which police use $m_N(j)$ resources to monitor individual j. By monitoring an individual, the police increase the probability of that individual being caught carrying out a criminal activity. Specifically,

$$p_N(j) = p_0 + m_N(j).$$

The probability that an individual j is caught equals the sum of a baseline probability, p_0, and the amount of monitoring, $m_N(j)$. The baseline probability depends on factors such as whether the community is willing to report crimes, and whether members of the community are willing to install personal monitoring devices like surveillance cameras. Monitoring reduces the effort that an individual decides to accord towards crime. However, there is a maximum amount, $\bar{m}$, by which any individual can be monitored. We can now rewrite the expressions for $\mathbf{e}_N^*$ and e_{agg}^*:

$$\mathbf{e}_N^* = \mathbf{P}\left(\mathbf{I} - \phi\mathbf{G}\right)^{-1}\mathbf{P}^T\left[\left(\pi + \xi - p_0 f\right)\mathbf{1} - \mathbf{m}_N f\right]$$

and

$$e_{agg}^* = \mathbf{1}^T\left(\mathbf{I} - \phi\mathbf{G}\right)^{-1}\left(\pi + \xi - p_0 f\right)\mathbf{1} - \mathbf{1}^T\left(\mathbf{I} - \phi\mathbf{G}\right)^{-1}\mathbf{P}^T\mathbf{m}_N f.$$

The following constraints hold:

$$p_0 + \bar{m} \in (0,1) \text{ and } \left(\pi + \xi - p_0 f\right)\mathbf{1} - \mathbf{m}_N f > \mathbf{0}.$$

With criminal activity directly proportional to criminal effort, the mayor and the police both care about reducing aggregate criminal effort. However, the police are resource-constrained: $\sum_{j=1}^{N} m_N(j) \leq M$ and $M < \bar{m}N$. The police must choose a monitoring policy that respects this upper bound, M. Aggregate criminal effort must also be below $\bar{e}_{agg}$, which is the maximum level of criminal effort that the public is willing to tolerate. Since the mayor is seeking reelection, he strongly desires $e_{agg}^* \leq \bar{e}_{agg}$; otherwise, the mayor will lose his office.

The police know the topology of the criminal influence network because they are targeting specific crimes, and these crimes generate naturally occurring network structures. The characteristics of the individuals and their environment, that is, π and ξ, are moreover observable. Unfortunately, without expending additional resources, the police do not know how these individuals sit on the network.

This criminal setting maps to the theoretical environment in Section 2. If we set

$$constant_N = \mathbf{1}^T (\mathbf{I} - \phi\mathbf{G})^{-1} (\pi + \xi - p_0 f) \mathbf{1},$$

$$\mathbf{w}_N^T = \mathbf{1}^T (\mathbf{I} - \phi\mathbf{G})^{-1},$$

$$\mathbf{a}_N = -\mathbf{m}_N f,$$

and

$$z_N^* = e_{agg}^*,$$

we have:

$$z_N^* = constant_N + \mathbf{w}_N^T \mathbf{P}^T \mathbf{a}_N = constant_N + \sum_{j=1}^{N} w_N (\gamma_{Nj}) \, a_N (j) \, .$$

The relevant set of attributes in the criminal setting is the collection of monitoring amounts, $\{m_N (j)\}_{j=1}^N$, multiplied by a constant.

Given the lack of knowledge of $\boldsymbol{\gamma}_N$, the police can monitor everyone equally, setting $\mathbf{m}_N = \frac{M}{N}\mathbf{1}$. However, under this policy,

$$e_{agg}^* = \mathbf{1}^T (\mathbf{I} - \phi\mathbf{G})^{-1} (\pi + \xi - p_0 f) \mathbf{1} - \mathbf{1}^T (\mathbf{I} - \phi\mathbf{G})^{-1} \frac{M}{N} \mathbf{1} f > \bar{e}_{agg}.$$

Aggregate criminal effort is still above the level that the public is willing to tolerate. Thus, the police must try monitoring ex ante identical individuals unequally and see if this approach enables $e_{agg}^* \leq \bar{e}_{agg}$.

The police have the option of expending C resources to learn where individuals sit on the network, that is, to learn $\boldsymbol{\gamma}_N$. Upon learning $\boldsymbol{\gamma}_N$, the police can implement an optimal monitoring policy. The optimal policy, $\mathbf{m}_N^{opt} (M - C)$, involves allocating $\bar{m}$ monitoring resources to the individual with the highest network constant, allocating $\bar{m}$ monitoring resources to the individual with the next highest network constant, and continuing in this fashion until all $M - C$ resources are expended. Provided that $e_{agg}^* \leq \bar{e}_{agg}$ under $\mathbf{m}_N^{opt} (M)$, it will be optimal to learn $\boldsymbol{\gamma}_N$ for a small enough C. If C is relatively large, though, the police will not have enough monitoring resources remaining to combat crime.

What the police can do instead is randomly implement the optimal monitoring policy using all available resources, M. The police can allocate $\bar{m}$ monitoring resources towards k individuals in a selected subset of the population and then allocate the remainder, $M - k\bar{m}$ to one additional individual. Since the police do not know where these individuals actually sit on the network, it is as if the police are randomly targeting network nodes. Given this monitoring policy, the police can construct the distribution of possible resulting aggregate effort levels by considering all possible ways that the individuals, with their monitoring amounts, can be arranged on the network. The random variable of interest is Z_N, the level of aggregate criminal effort, defined as follows:

$$Z_N = constant_N + \sum_{j=1}^{N} w_N\left(X_{Nj}\right) a_N\left(j\right),$$

with $\mathbf{X}_N$ an $N \times 1$ random vector whose realizations are permutations of the indices $(1, \ldots, N)$, and the objects $constant_N$, $\mathbf{w}_N$, and $\mathbf{a}_N$ previously defined.

The mayor and the police are interested in selecting a monitoring policy that accords more mass to the left tail of the aggregate criminal effort distribution. By doing so, the mayor and the police are able to increase the probability that total crime is below the public's threshold. However, it is important to note that, when total resources expended are equal to M, *any* monitoring policy that is randomly implemented yields a distribution of aggregate criminal effort with the *exact same* mean. Moreover, the mean of the distribution is equal to the level of aggregate criminal effort that results under the uniform monitoring policy (follows from Theorem 1). If total criminal effort under the uniform policy exceeds the public's threshold, then for any randomly implemented monitoring policy, total criminal effort also exceeds the public's threshold, on average. The mayor and the police must therefore design a monitoring policy that transfers as much mass as possible to the left tail of the distribution, with the knowledge that they will never be able to shift the distribution's mean.

5.2 Managing Crime in a Sample Economy

In the early part of 1993, members of a large cocaine trafficking organization in New York City were wiretapped (Natarajan, 2000); 151 telephone conversations were recorded. These telephone conversations represent flows of information and more general exchanges of communication between members of the organization. We can use records of these conversations to generate a network of influence among the members of the cocaine trafficking organization.[4] To generate the influence network, we assume that telephone conversations enable the mutual propagation of influence. Accordingly, the linkages in this network are undirected. Figure 8 depicts the influence network for members of the cocaine trafficking organization. There are 28 nodes and 40 edges, and the network has a density of 0.106.

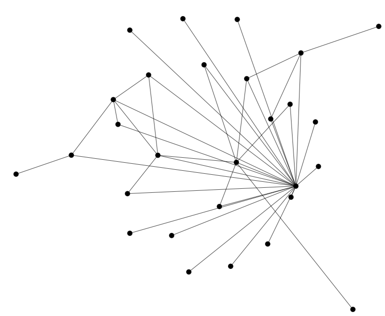

Figure 8: The New York City cocaine dealing network.

Now, a separate United States city also happens to have a local cocaine trafficking ring. The police in that city are trying to contain the crime that the ring generates, but they currently know neither the topology of the network nor the network position of each trafficker. The police can reasonably assume that the topology of the criminal interaction network for the local cocaine trafficking ring is the same as the one from New York City. The local police will later have the option of expending resources to learn individuals' network positions.

Before we go further, let us assign values to the key parameters shaping our economic environment. We set $\pi = 31$ and $\xi = 30$; π and ξ are the two parameters

[4] The network data is available at: https://sites.google.com/site/ucinetsoftware/datasets/covert-networks/cocainedealingnatarajan.

that determine an individual's proceeds from carrying out a crime. We also set ϕ equal to 0.15; ϕ captures the extent to which individuals' choices of effort are strategic complements. The parameter, f, which determines an individual's cost from getting caught by the police, equals 100. The baseline probability of a criminal being caught in the absence of police monitoring, p_0, is 0.10. The police have $M = 5$ total units of resources for monitoring. The maximum amount by which any individual can be monitored is: $\bar{m} = 0.5$. Finally, the maximum level of criminal effort that the public is willing to tolerate is: $\bar{e}_{agg} = 4500$ units.

Given the topology of the interaction network, the local police can compute the vector of network constants, $\mathbf{w}_N$. The counter-cumulative distribution function of network constants is plotted in Figure 9. The average network constant takes a value of 5.13, with a standard deviation of 3.07, skewness of 3.13, and kurtosis of 14.44. The smallest network constant is 1.78, and the largest network constant is 18.85. There is tremendous heterogeneity in network constants within the population, which means that aggregate criminal effort will fluctuate widely depending on which individuals the police target.

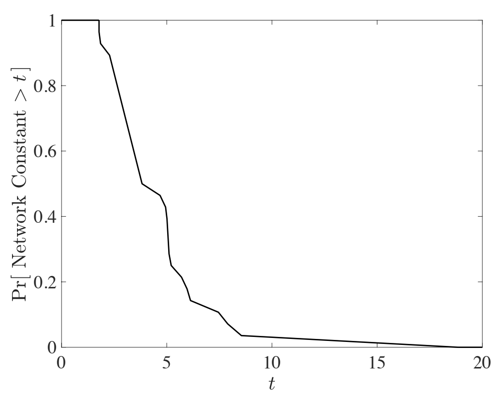

Figure 9: The counter-cumulative distribution function of network constants.

The mayor and the police consider different monitoring policies given their available resources. They first consider monitoring every individual equally, setting $\mathbf{m}_N = \frac{M}{N}\mathbf{1}$. Unfortunately, this policy results in aggregate criminal effort of 4764 units, which exceeds $\bar{e}_{agg}$, the threshold that the public is willing to tolerate (see Figure 10, top left). The police therefore need to monitor ex ante identical individuals unequally. Let us suppose that the cost of learning where individuals sit on the network, C, equals 1 unit. The police then implement the following optimal monitoring policy, $\mathbf{m}_N^{opt}(M-C)$: they fully monitor the eight individuals with the highest network constants, and they accord zero monitoring to the remaining agents. The aggregate criminal effort that results equals 4043 units, which is below the public's threshold (see Figure 10, top right). However, if it turns out that the cost of learning where individuals sit on the network, C, is much higher, say, it equals 3 units, $\mathbf{m}_N^{opt}(M-C)$ no longer yields a sufficiently low crime level. When $C=3$, the optimal monitoring policy, $\mathbf{m}_N^{opt}(M-C)$, involves fully monitoring the four individuals with the highest network constants, and according zero monitoring to the remaining agents. The aggregate criminal effort that results equals 5193 units, which greatly exceeds the public's threshold, $\bar{e}_{agg}$ (see Figure 10, bottom left).

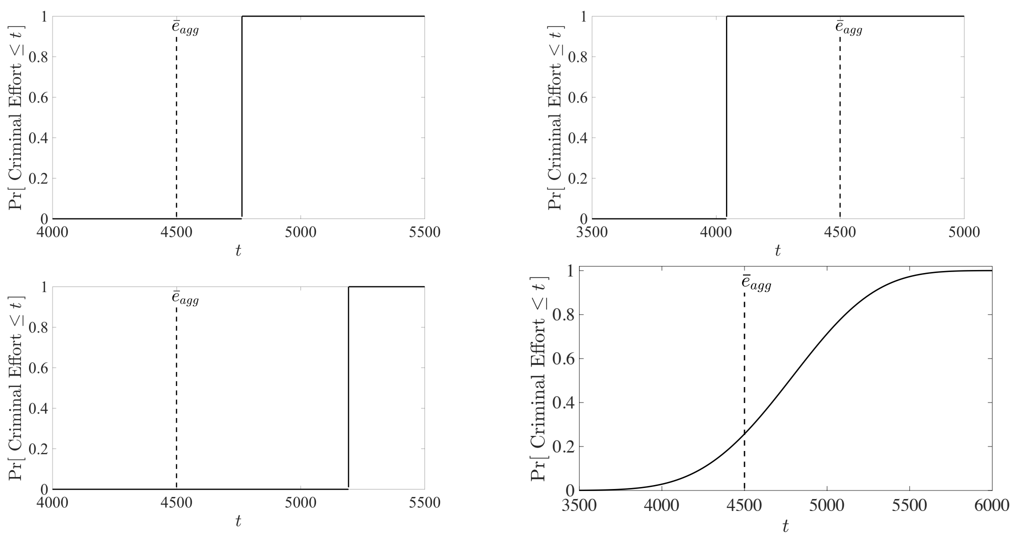

Figure 10: The distribution of aggregate criminal effort under an equal monitoring policy, $\mathbf{m}_N = \frac{M}{N}\mathbf{1}$ (top left). The distribution of aggregate criminal effort under an optimal monitoring policy, $\mathbf{m}_N^{opt}(M-C)$, setting C = 1 (top right). The distribution of aggregate criminal effort under an optimal monitoring policy, $\mathbf{m}_N^{opt}(M-C)$, setting C = 3 (bottom left). The distribution of aggregate criminal effort after randomly implementing $\mathbf{m}_N^{opt}(M)$ (bottom right).

The mayor and the police thus decide, in this high-cost environment, to not expend any resources learning each individual's network position. Rather, they choose to randomly implement $\mathbf{m}_N^{opt}(M)$ using all available resources, M. Ten individuals, selected at random, are fully monitored, and all other agents receive zero monitoring. The distribution of aggregate criminal effort that results (see Figure 10, bottom right) has a mean of 4764 units (follows from Theorem 1), with a standard deviation of 396.31 units (follows from Theorem 2), skewness of -0.37 (follows from Theorem 3), and kurtosis of 2.07 (follows from Theorem 4). The lowest possible level of aggregate criminal effort that can result is 3532 units, while the highest possible level of aggregate criminal effort that can result is 5695 units (follows from Theorem 6). The probability that aggregate criminal effort is below the public's threshold is 25.72 percent (follows from Theorem 8). Without knowing individuals' network positions, there is still at least a one-in-four chance that the police can contain crime to a desirable level.

Any random monitoring policy that fully expends the $M = 5$ resources generates a distribution of aggregate criminal effort with a mean that *always* equals 4764 units (follows from Theorem 1). Regardless of how agents are targeted at random, the mean of this distribution is fixed. On average, for any random monitoring policy, aggregate criminal effort therefore always exceeds the public's threshold. The question is whether the police can design a monitoring policy that transfers as much mass as possible below the public's threshold while holding the mean of the distribution fixed.

6 Conclusion

This work develops theoretical tools so that an observer to a networked economy can accurately characterize the economy and make decisions in the absence of full information. We focus on the following realistic departure from full information: the observer knows the population of attributes comprising the economy as well as the topology of the network, but not how these attributes specifically sit on the network's nodes. This work puts forth a methodological approach and develops an accompanying set of comprehensive theoretical tools for the decision-making observer. The observer can therefore overcome this lack of information and still make decisions despite the uncertain environment. Departing from the full-information environment is important in the study of networked economies because it achieves greater realism;

the present work develops a complete theory in this space.

References

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

Battaglini, M. and E. Patacchini (2018): "Influencing connected legislators," *Journal of Political Economy*, 126, 2277–2322.

Belhaj, M. and F. Deroïan (2018): "Targeting the key player: An incentive-based approach," *Journal of Mathematical Economics*, 79, 57–64.

——— (2019): "Group targeting under networked synergies," *Games and Economic Behavior*, 118, 29–46.

Belhaj, M., F. Deroïan, and S. Safi (2020): "Costly agreement-based transfers and targeting on networks with synergies," *Working Paper*.

Bloch, F. and N. Quérou (2013): "Pricing in social networks," *Games and Economic Behavior*, 80, 243–261.

Breza, E., A. Chandrasekhar, T. McCormick, and M. Pan (2020): "Using aggregated relational data to feasibly identify network structure without network data," *American Economic Review*, 110, 2454–84.

Brown, W. and U. Patange (2020): "Targeted intervention in random graphs," *International Symposium on Algorithmic Game Theory*, 211–225.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2008): "Peer effects and social networks in education," *Working Paper*.

——— (2009): "Peer effects and social networks in education," *The Review of Economic Studies*, 76, 1239–1267.

Candogan, O., K. Bimpikis, and A. Ozdaglar (2012): "Optimal pricing in networks with externalities," *Operations Research*, 60, 883–905.

Dasaratha, K. (2020): "Distributions of centrality on networks," *Games and Economic Behavior*, 122, 1–27.

Davis, A., B. Gardner, and M. Gardner (1941): *Deep South*, Chicago: The University of Chicago Press.

Debreu, G. and I. Herstein (1953): "Nonnegative square matrices," *Econometrica*, 597–607.

Demange, G. (2017): "Optimal targeting strategies in a network under complementarities," *Games and Economic Behavior*, 105, 84–103.

Does, R. J. (1983): "An Edgeworth expansion for simple linear rank statistics under the null-hypothesis," *The Annals of Statistics*, 607–624.

Fainmesser, I. and A. Galeotti (2016): "Pricing network effects," *The Review of Economic Studies*, 83, 165–198.

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

Ghiglino, C. and S. Goyal (2010): "Keeping up with the neighbors: social interaction in a market economy," *Journal of the European Economic Association*, 8, 90–119.

Hájek, J. (1961): "Some extensions of the Wald-Wolfowitz-Noether theorem," *The Annals of Mathematical Statistics*, 506–523.

Jackson, M. and Y. Zenou (2015): "Games on networks," in *Handbook of Game Theory with Economic Applications*, Elsevier, vol. 4, 95–163.

La'O, J. and A. Tahbaz-Salehi (2020): "Optimal monetary policy in production networks," *Working Paper.*

Lee, L.-F., X. Liu, E. Patacchini, and Y. Zenou (2014): "Criminal networks: Who is the key player?" *Working Paper.*

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

MASTRANDREA, R., J. FOURNET, AND A. BARRAT (2015): "Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries, and friendship surveys," *PLoS ONE*, 10, 1–26.

NATARAJAN, M. (2000): "Understanding the structure of a drug trafficking organization: A conversational analysis," *Crime Prevention Studies*, 11, 273–298.

PARISE, F. AND A. OZDAGLAR (2019): "Graphon games," *Proceedings of the 2019 ACM Conference on Economics and Computation*, 457–458.

——— (2020): "Analysis and interventions in large network games," *Annual Review of Control, Robotics, and Autonomous Systems*, 4, 455–486.

VALENTE, T. (2012): "Network interventions," *Science*, 337, 49–53.

Appendix: Proofs

For proofs of Theorems 3, 4, 9, 10, and 11, please refer to the Online Appendix.

Proof of Theorem 1

$EZ_N = E\left[\sum_{i=1}^{N} w_N(i)\, a_N(X_{Ni})\right] = \sum_{i=1}^{N} w_N(i)\, E\left[a_N(X_{Ni})\right] = N\bar{w}_N\bar{a}_N.$ □

Proof of Theorem 2

Dropping the "N" subscripts, we have the following:

$$\operatorname{Var} Z = E\left[\sum_{i=1}^{N} w(i)\, a(X_i) - N\bar{w}\bar{a}\right]^2$$

$$= E\left(\sum_{i=1}^{N} w(i)\, a(X_i) \sum_{j=1}^{N} w(j)\, a(X_j)\right) - (N\bar{w}\bar{a})^2$$

$$= E\left(\sum_{i=1}^{N} (w(i))^2 (a(X_i))^2\right) + E\left(\sum_{i=1}^{N} \sum_{\substack{j=1\\ i\neq j}}^{N} w(i)\, w(j)\, a(X_i)\, a(X_j)\right) - (N\bar{w}\bar{a})^2$$

$$= E(a(X_i))^2 \sum_{i=1}^{N} (w(i))^2 + \left[\frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{\substack{j=1\\ i\neq j}}^{N} a(i)\, a(j)\right]\left[\sum_{i=1}^{N} \sum_{\substack{j=1\\ i\neq j}}^{N} w(i)\, w(j)\right] - (N\bar{w}\bar{a})^2$$

$$= \left(\operatorname{Var} A + \bar{a}^2\right)\left(N \operatorname{Var} W + N\bar{w}^2\right) + \frac{1}{N(N-1)}\left[(N\bar{a})^2 - \left(N \operatorname{Var} A + N\bar{a}^2\right)\right] \times \left[(N\bar{w})^2 - \left(N \operatorname{Var} W + N\bar{w}^2\right)\right] - (N\bar{w}\bar{a})^2$$

$$= \frac{1}{N-1}\left(N \operatorname{Var} W\right)\left(N \operatorname{Var} A\right). \ \square$$

Proof of Theorem 5

Suppose that Z_N is invariant to rearrangement. Then $\operatorname{Var} Z_N = 0$. $\operatorname{Var} Z_N = 0$ if and only if (1) $\operatorname{Var} W_N = 0$ and/or (2) $\operatorname{Var} A_N = 0$, that is if and only if $w_N(i) = \bar{w}_N \ \forall i \in \{1, \ldots, N\}$ and/or $a_N(i) = \bar{a}_N \ \forall i \in \{1, \ldots, N\}$.

Now suppose that (1) $w_N(i) = \bar{w}_N \ \forall i \in \{1, \ldots, N\}$ and/or (2) $a_N(i) = \bar{a}_N \ \forall i \in \{1, \ldots, N\}$. Then (1) $\operatorname{Var} W_N = 0$ and/or (2) $\operatorname{Var} A_N = 0$, and it follows that $\operatorname{Var} Z_N = 0$ and Z_N is invariant to rearrangement.

With probability 1, $Z_N = \sum_{i=1}^{N} w_N(i)\, a_N(X_{Ni}) = \sum_{i=1}^{N} \bar{w}_N \bar{a}_N = N \bar{w}_N \bar{a}_N$. $\square$

Proof of Theorem 6

By the rearrangement inequality,

$$\begin{aligned} &\widetilde{w}_N(1)\,\widetilde{a}_N(N) + \cdots + \widetilde{w}_N(i)\,\widetilde{a}_N(N+1-i) + \cdots + \widetilde{w}_N(N)\,\widetilde{a}_N(1) \\ &\quad \le \widetilde{w}_N(1)\,\widetilde{a}_N(X_{N1}) + \cdots + \widetilde{w}_N(i)\,\widetilde{a}_N(X_{Ni}) + \cdots + \widetilde{w}_N(N)\,\widetilde{a}_N(X_{NN}) \\ &\qquad\qquad \le \widetilde{w}_N(1)\,\widetilde{a}_N(1) + \cdots + \widetilde{w}_N(i)\,\widetilde{a}_N(i) + \cdots + \widetilde{w}_N(N)\,\widetilde{a}_N(N), \end{aligned}$$

where $\mathbf{X}_N = \begin{pmatrix} X_{N1} & \cdots & X_{Ni} & \cdots & X_{NN} \end{pmatrix}^T$ takes on all $N!$ permutations of the indices $(1, \ldots, N)$. $\square$

Proof of Theorem 7

As defined in Hájek (1961), let $(R_{\nu 1}, \ldots, R_{\nu N_\nu})$ be a random vector which takes on the $N_\nu!$ permutations of $(1, \ldots, N_\nu)$ with equal probabilities. Let $\{b_{\nu i}, 1 \le i \le N_\nu, \nu \ge 1\}$ and $\{a_{\nu i}, 1 \le i \le N_\nu, \nu \ge 1\}$ be double sequences of real numbers. Put $S_\nu = \sum_{i=1}^{N_\nu} b_{\nu i}\, a_{\nu R_{\nu i}}$, and define $\bar{a}_\nu = \frac{1}{N_\nu} \sum_{i=1}^{N_\nu} a_{\nu i}$ and $\bar{b}_\nu = \frac{1}{N_\nu} \sum_{i=1}^{N_\nu} b_{\nu i}$. Consider the following lemma:

Lemma 1 (Hájek (1961), Theorem 4.1) *Let us suppose that*

$$\lim_{\nu\to\infty} \frac{\max_{1\le i\le N_\nu} (a_{\nu i} - \bar{a}_\nu)^2}{\sum_{i=1}^{N_\nu} (a_{\nu i} - \bar{a}_\nu)^2} = 0 \quad \textit{and} \quad \lim_{\nu\to\infty} \frac{\max_{1\le i\le N_\nu} \left(b_{\nu i} - \bar{b}_\nu\right)^2}{\sum_{i=1}^{N_\nu} \left(b_{\nu i} - \bar{b}_\nu\right)^2} = 0.$$

Then S_ν has an asymptotically normal distribution with mean value ES_ν and variance $\operatorname{Var} S_\nu$ *if, and only if, for any $\tau > 0$*

$$\lim_{\nu\to\infty} \frac{1}{N_\nu} \sum_{|\delta_{\nu ij}|>\tau}\sum \delta_{\nu ij}^2 = 0,$$

where

$$\delta_{\nu ij} = \frac{\left(b_{\nu i} - \bar{b}_\nu\right)\left(a_{\nu j} - \bar{a}_\nu\right)}{\left[\frac{1}{N_\nu}\sum_{i=1}^{N_\nu}\left(b_{\nu i} - \bar{b}_\nu\right)^2 \sum_{j=1}^{N_\nu}\left(a_{\nu j} - \bar{a}_\nu\right)^2\right]^{1/2}} \qquad (1 \le i, j \le N_\nu,\ \nu \ge 1).$$ [5]

Theorem 7 follows immediately from Lemma 1. □

Proof of Theorem 8

To prove this theorem, we first provide the necessary background for Theorem 2.1 in Does (1983). We then present Theorem 2.1 in Does (1983) as a lemma, and from this lemma, we show how to prove Theorem 8.

As in Does (1983), with notation modified for the present work, let $\Delta_1, \Delta_2, \ldots, \Delta_N$ be independent and identically distributed random variables with a common continuous distribution function F. If $\Delta_{1:N} < \Delta_{2:N} < \cdots < \Delta_{N:N}$ denotes the sequence $\Delta_1, \Delta_2, \ldots, \Delta_N$ arranged in increasing order, then the rank R_{jN} of Δ_j is defined by $\Delta_j = \Delta_{R_{jN}:N}$ for $j = 1, 2, \ldots, N$. Does (1983) considers the simple linear rank statistic

$$T_N = \sum_{j=1}^{N} x_{jN} J\left(\frac{R_{jN}}{N+1}\right),$$

[5] There is a minor typographical error in Theorem 4.1 of Hájek (1961). We should have

$$\delta_{\nu ij} = \frac{\left(b_{\nu i} - \bar{b}_\nu\right)\left(a_{\nu j} - \bar{a}_\nu\right)}{\left[\frac{1}{N_\nu}\sum_{i=1}^{N_\nu}\left(b_{\nu i} - \bar{b}_\nu\right)^2 \sum_{j=1}^{N_\nu}\left(a_{\nu j} - \bar{a}_\nu\right)^2\right]^{1/2}}$$

and not

$$\delta_{\nu ij} = \frac{\left(b_{\nu i} - b_\nu\right)\left(a_{\nu j} - \bar{a}_\nu\right)}{\left[\frac{1}{N_\nu}\sum_{i=1}^{N_\nu}\left(b_{\nu i} - \bar{b}_\nu\right)^2 \sum_{j=1}^{N_\nu}\left(a_{\nu j} - \bar{a}_\nu\right)^2\right]^{1/2}}.$$

where $x_{1N}, x_{2N}, \ldots, x_{NN},\ N = 1, 2, \ldots$ is a triangular array of regression constants and J is a score generating function defined on $(0, 1)$.

Assumption A (Does, 1983). The regression constants $x_{1N}, x_{2N}, \ldots, x_{NN}$ satisfy $\sum_{j=1}^{N} x_{jN} = 0$, $\sum_{j=1}^{N} x_{jN}^2 = 1$, and $\max_{1\leq j\leq N} |x_{jN}| = O\left(N^{-1/2}\right)$.

Assumption B (Does, 1983). The score generating function J is three times differentiable on $(0, 1)$ and

$$\limsup_{t\to 0,1} t\,(1-t)\left|\frac{J''(t)}{J'(t)}\right| < 2;$$

there exist positive numbers $\Gamma > 0$ and $\alpha < 3 + \frac{1}{14}$ such that the third derivative J''' satisfies

$$|J'''(t)| \leq \Gamma\,(t\,(1-t))^{-\alpha} \text{ for } t \in (0,1)\,.$$

Does (1983) also assumes $\int_0^1 J(t)\,dt = 0$ and $\int_0^1 J^2(t)\,dt = 1$. The variance σ_N^2 of T_N is given by $\sigma_N^2 = \sigma^2(T_N) = \frac{1}{N-1}\sum_{j=1}^{N}\left(J\left(\frac{j}{N+1}\right) - \bar{J}\right)^2$, where $\bar{J} = \frac{1}{N}\sum_{j=1}^{N} J\left(\frac{j}{N+1}\right)$. For each $N \geq 2$, define $T_N^* = \sigma_N^{-1} T_N$ and $G_{T_N^*}(t) = \Pr(T_N^* \leq t)$ for $-\infty < t < \infty$. Furthermore define for each $N \geq 2$ and real t the function $K_N(t)$ by

$$K_N(t) = \Phi(t) - \phi(t)\left[\frac{\kappa_{3N}}{6}\left(t^2 - 1\right) + \frac{\kappa_{4N}}{24}\left(t^3 - 3t\right) + \frac{\kappa_{3N}^2}{72}\left(t^5 - 10t^3 + 15t\right)\right],$$

where $\Phi(t)$ denotes the standard normal distribution function and $\phi(t)$ its density; the quantities κ_{3N} and κ_{4N} are given by

$$\kappa_{3N} = \sum_{j=1}^{N} x_{jN}^3\left[\int_0^1 J^3(t)\,dt\right] \text{ and } \kappa_{4N} = \sum_{j=1}^{N} x_{jN}^4\left[\int_0^1 J^4(t)\,dt - 3\right] - \frac{3}{N}\left[\int_0^1 J^4(t)\,dt - 1\right].$$

Then:

Lemma 2 (Does (1983), Theorem 2.1) *If the Assumptions A and B are satisfied, then as $N \to \infty$*

$$\sup_{t\in\mathbb{R}}\left|G_{T_N^*}(t) - K_N(t)\right| = o\left(N^{-1}\right).$$

Begin with the sequence $\{y_{jN}\}_{j=1}^{N}$, whose elements are ordered, without loss of generality, so that $y_{jN} \leq y_{j'N}$ whenever $j \leq j'$. We now proceed to define the function $J(t)$ on the interval $(0,1)$. With $\eta \in \left(0, \frac{1}{2N}\right)$,

$$J(t) = \begin{cases} p_{0N}(t) & \text{if } t \in (0, \eta) \\ y_{jN}(t) & \text{if } t \in \left[\frac{j-1}{N} + \eta, \frac{j}{N} - \eta\right), \forall j = 1, 2, \ldots, N \\ p_{jN}(t) & \text{if } t \in \left[\frac{j}{N} - \eta, \frac{j}{N} + \eta\right), \forall j = 1, 2, \ldots, N-1 \\ p_{NN}(t) & \text{if } t \in [1 - \eta, 1) \end{cases}$$

The objects $p_{0N}(t), p_{1N}(t), \ldots, p_{NN}(t)$ are interpolating polynomials with the following constraints:

$$\lim_{t \to \eta -} p_{0N}(t) = y_{1N}$$
$$p_{jN}\left(\frac{j}{N} - \eta\right) = y_{jN}, \quad \forall j = 1, 2, \ldots, N, \text{ and}$$
$$\lim_{t \to \left(\frac{j}{N} + \eta\right) -} p_{jN}(t) = y_{(j+1)N}, \quad \forall j = 1, 2, \ldots, N-1.$$

The polynomials $p_{0N}(t), p_{1N}(t), \ldots, p_{NN}(t)$ are constructed so that the function $J(t)$ satisfies Assumption B. As $\eta \to 0$, $J(t)$ becomes a step function on the domain $(0,1)$:

$$\lim_{\eta \to 0} J(t) = \sum_{j=1}^{N} y_{jN} \, \mathbb{1}_{\frac{j-1}{N} \leq t < \frac{j}{N}},$$

with $\mathbb{1}_{\frac{j-1}{N} \leq t < \frac{j}{N}}$ an indicator function.

For the remainder of this proof, we focus on the limit $\eta \to 0$. In the next lemma, we show that $J\left(\frac{j}{N+1}\right) = y_{jN}$ for all $j = 1, 2, \ldots, N$.

Lemma 3 *For all $j = 1, 2, \ldots, N$, we have $J\left(\frac{j}{N+1}\right) = y_{jN}$.*

Proof. $J\left(\frac{j}{N+1}\right) = y_{jN}$ if and only if $\frac{j}{N+1} \in \left(\frac{j-1}{N}, \frac{j}{N}\right)$. First demonstrating $\frac{j}{N+1} < \frac{j}{N}$:

$$\frac{j}{N+1} = \frac{j}{N} \frac{N}{N+1} < \frac{j}{N}.$$

Next demonstrating $\frac{j}{N+1} > \frac{j-1}{N}$: $\frac{j}{N+1} > \frac{j-1}{N}$ if and only if $\frac{j}{N+1} - \frac{j-1}{N} > 0$, that is, if and only if $jN - (j-1)(N+1) > 0$. $jN - (j-1)(N+1) = N + 1 - j$, which is greater than zero for all $j = 1, 2, \ldots, N$. ■

Given Lemma 3, we have $J\left(\frac{R_{jN}}{N+1}\right) = y_{R_{jN}N}$ and $T_N = \sum_{j=1}^{N} x_{jN}\, y_{R_{jN}N}$. With $\Delta_1, \Delta_2, \ldots, \Delta_N$ i.i.d., the vector $\begin{pmatrix} R_{1N} & R_{2N} & \cdots & R_{NN} \end{pmatrix}$ is uniformly distributed over all $N!$ permutations of $\begin{pmatrix} 1 & 2 & \cdots & N \end{pmatrix}$. Set

$$x_{jN} = \frac{w_N(j) - \bar{w}_N}{(N \operatorname{Var} W_N)^{1/2}} \quad \text{and} \quad y_{jN} = \frac{a_N(j) - \bar{a}_N}{(\operatorname{Var} A_N)^{1/2}}.$$

Then $\sum_{j=1}^{N} x_{jN} = \sum_{j=1}^{N} \frac{w_N(j) - \bar{w}_N}{(N \operatorname{Var} W_N)^{1/2}} = 0$ and $\sum_{j=1}^{N} x_{jN}^2 = \sum_{j=1}^{N} \frac{(w_N(j) - \bar{w}_N)^2}{N \operatorname{Var} W_N} = 1$, and Assumption A is satisfied. We also satisfy the conditions $\int_0^1 J(t)\, dt = 0$ and $\int_0^1 J^2(t)\, dt = 1$:

$$\int_0^1 J(t)\, dt = \sum_{j=1}^{N} \frac{1}{N} y_{jN} = \frac{1}{N} \sum_{j=1}^{N} \frac{a_N(j) - \bar{a}_N}{(\operatorname{Var} A_N)^{1/2}} = 0, \text{ and}$$

$$\int_0^1 J^2(t)\, dt = \sum_{j=1}^{N} \frac{1}{N} y_{jN}^2 = \frac{1}{N} \sum_{j=1}^{N} \frac{(a_N(j) - \bar{a}_N)^2}{\operatorname{Var} A_N} = 1.$$

We can set

$$\bar{J} = \frac{1}{N} \sum_{j=1}^{N} J\left(\frac{j}{N+1}\right) = \frac{1}{N} \sum_{j=1}^{N} y_{jN} = \frac{1}{N} \sum_{j=1}^{N} \frac{a_N(j) - \bar{a}_N}{(\operatorname{Var} A_N)^{1/2}} = 0, \text{ and}$$

$$\begin{aligned}
\sigma_N^2 = \sigma^2(T_N) = \frac{1}{N-1} \sum_{j=1}^{N} \left(J\left(\frac{j}{N+1}\right) - \bar{J}\right)^2 &= \frac{1}{N-1} \sum_{j=1}^{N} \left(J\left(\frac{j}{N+1}\right)\right)^2 \\
&= \frac{1}{N-1} \sum_{j=1}^{N} y_{jN}^2 = \frac{1}{N-1} \sum_{j=1}^{N} \frac{(a_N(j) - \bar{a}_N)^2}{\operatorname{Var} A_N} = \frac{N}{N-1}.
\end{aligned}$$

Therefore,

$$T_N^* = \sigma_N^{-1} T_N = \sum_{j=1}^{N} \frac{x_{jN}\, y_{R_{jN}N}}{\sqrt{\frac{N}{N-1}}} = \frac{\sum_{j=1}^{N} \frac{w_N(j)-\bar{w}_N}{\sqrt{N \operatorname{Var} W_N}} \frac{a_N\left(R_{jN}\right)-\bar{a}_N}{\sqrt{\operatorname{Var} A_N}}}{\sqrt{\frac{N}{N-1}}} = \frac{Z_N - EZ_N}{\left(\operatorname{Var} Z_N\right)^{1/2}},$$

given the expressions for EZ_N and $\operatorname{Var} Z_N$ in Theorems 1 and 2. We proceed to solve for κ_{3N} and κ_{4N}:

$$\begin{aligned}\kappa_{3N} &= \sum_{j=1}^{N} x_{jN}^3 \left[\int_0^1 J^3(t)\, dt\right] = \sum_{i=1}^{N} \left(\frac{w_N(i) - \bar{w}_N}{\sqrt{N \operatorname{Var} W_N}}\right)^3 \sum_{j=1}^{N} \frac{1}{N} \left(\frac{a_N(j) - \bar{a}_N}{\sqrt{\operatorname{Var} A_N}}\right)^3 \\ &= \frac{1}{\sqrt{N}} \operatorname{Skew} W_N \ \operatorname{Skew} A_N,\end{aligned}$$

and

$$\begin{aligned}\kappa_{4N} &= \sum_{j=1}^{N} x_{jN}^4 \left[\int_0^1 J^4(t)\, dt - 3\right] - \frac{3}{N}\left[\int_0^1 J^4(t)\, dt - 1\right] \\ &= \frac{1}{N} \operatorname{Kurt} W_N \ \operatorname{Kurt} A_N - \frac{3}{N} \operatorname{Kurt} W_N - \frac{3}{N} \operatorname{Kurt} A_N + \frac{3}{N}.\end{aligned}$$

Given Lemma 2,

$$\sup_{t \in \mathbb{R}} \left| G_{\frac{Z_N - EZ_N}{\left(\operatorname{Var} Z_N\right)^{1/2}}}(t) - K_N(t) \right| = o\left(N^{-1}\right)$$

approximately holds as $N \to \infty$ with $\kappa_{3N} = \frac{1}{\sqrt{N}} \operatorname{Skew} W_N \ \operatorname{Skew} A_N$ and $\kappa_{4N} = \frac{1}{N} \operatorname{Kurt} W_N \ \operatorname{Kurt} A_N - \frac{3}{N} \operatorname{Kurt} W_N - \frac{3}{N} \operatorname{Kurt} A_N + \frac{3}{N}$. $\square$