Analysis of Flow-Through Entities Using Social Network Analysis Techniques

Ashish Agarwal and Shannon Chen, McCombs School of Business, University of Texas, Austin,
and Ririko Horvath, Larry May, and Rahul Tikekar, Research, Analysis, and Statistics,
Internal Revenue Service

Introduction

The tax law allows flow-through tax treatment for certain legal entities so that income is subject to tax only once—at the partner or shareholder level. The information return that flow-through entities file includes a Schedule K-1 that shows income, deductions, credits, and other items that are allocated to the owners. Flow-through treatment introduces complexity in auditing tax returns as flow-through entities can be associated with many owners and vice versa. A flow-through entity can also be associated with many other flow-through entities. Additionally, the financial flows can vary across owners—both in type and magnitude. For example, owners can receive different types of income and deductions from a flow-through entity, and each owner may not receive the same proportion of the income and deductions. Due to this complexity, there is a need to represent the associations between different types of entities and the related financial flows in a form that is quantifiable and that can be used to evaluate these associations for tax compliance risk. This is especially important considering that more than 20 million Schedule K-1s are issued every year.

In its simplest form, a tax structure that includes flow-through entities is a network where entities can be represented as nodes and the linkages between entities can be represented as edges. Further, the financial flow associated with the linkages can represent the strength or weight of the link. Thus, social network analysis (SNA) is a potential tool that can be used to represent the economic structures resulting from the use of flow-through entities. SNA has been successfully used to represent the complex associations and flows across heterogeneous entities such as individuals, products, and firms. Many commercial platforms such as Facebook, Twitter, and Google+ use SNA to capture the dynamic linkages between individuals and associate individual behavior and preferences with their network characteristics. These platforms use these insights to provide targeted information and services to these individuals. Prior research has also investigated interconnections between user networks and product networks. Similarly, SNA has been used to represent industry structures and the associated outcomes.

In this paper, we investigate how SNA can be applied to characterize the complex associations between different entities within a flow-through structure and develop measures to quantify these associations.[1] To achieve this, we consider two different samples of Schedule K-1 data obtained from the Internal Revenue Service (IRS) yK1 database. We also consider different types of linkages occurring within enterprises, namely K-1 linkages, primary-secondary associations (i.e., spousal linkages), and parent-subsidiary linkages. For our analysis, we first examine existing enterprises and develop SNA measures at both the enterprise and node levels. Next, we construct graphs representing the different enterprises in our two yK1 samples and calculate SNA measures. We also investigate the potential application of these graphs to identify economically important nodes and unusual enterprises.

Our investigation shows that SNA can be used to represent the tax structures associated with flow-through entities. This includes the ability to capture different types of nodes, different types of linkages across these nodes, and the ability to represent various types of financial flows associated with the entities. We also illustrate how network measures can be used to characterize different enterprises and compare these enterprises. More specifically, we show how SNA measures can be used to determine if enterprises conform to the expected network structure and to flag any exceptions. Finally, we demonstrate how a combination of network and node level measures can be used to identify economically important nodes in an enterprise.

[1] In future work, we plan to investigate whether these SNA measures are predictive of tax noncompliant behavior.

Prior Literature

Prior academic work documents that certain firm characteristics, such as magnitude of book-tax differences, firm size, industry, and multinationality, are associated with corporate tax noncompliance (Mills (1998); Hanlon, Mills, and Slemrod (2007)). There is also evidence that greater organizational complexity and financial complexity are associated with higher levels of corporate tax avoidance (Wagener and Watrin (2013); Balakrishnan, Blouin, and Guay (2012)). However, these prior studies are limited to drawing conclusions about consolidated corporate entities where publicly filed financial statement data are available. In general, academic tax research has focused on corporate tax issues due to these data constraints.

With respect to flow-through entities, a separate stream of academic literature examines the choice of overall business structure given the tax and nontax costs and benefits (e.g., Guenther (1992); Ayers, Cloyd, and Robinson (1996); Gordon and MacKie-Mason (1994); MacKie-Mason and Gordon (1997)). For example, Ayers, *et al.* (1996) find that in choosing between corporate and noncorporate structures, business risk, number of owners, firm size, and firm age influence the choice of organizational form. However, again, these studies are limited to examining the organizational form decisions of top-level entities and cannot provide evidence regarding the use of flow-through entities embedded within corporate structures. Recent work on the use of special purpose entities, which include LLCs, LLPs, trusts, and other entities, is a first step in addressing how enterprises use flow-through entities within their business structure (Feng, Gramlich, and Gupta (2009); Demere, Donohoe, and Lisowsky (2015)). Nonetheless, without more detailed data, these studies can draw only high-level conclusions.

In this paper, we have access to a unique dataset collected by the IRS, which gives us visibility into the underlying organizational structure of business enterprises. Given the complexity of the data, which prior studies could not observe, we propose that SNA techniques can be useful in quantifying the many dimensions of complexity. For example, not only are we interested in measuring the number and types of entities within an enterprise, we also aim to measure and quantify the shape of enterprise structures, how entities are related, and the magnitude of these relationships. Our study contributes to the literature by proposing a new methodology to measure and quantify business structures.

Data

Our unit of analysis is an enterprise that contains two or more entities and includes at least one flow-through entity. These enterprises are defined by IRS using a 50-percent ownership rule. Specifically, a flow-through entity is considered to be part of an enterprise only if the taxpaying entities associated with the enterprise have—directly or indirectly—at least 50-percent ownership of the flow-through entity.

We consider two different samples of enterprises to conduct our proposed social network analysis. The first sample is based on an intersection of enterprises associated with entities that also appear in the proposed deficiency database in Tax Year 2009. The entities in the proposed deficiency database file Form 1120. This database flags entities that have deficiencies in their tax filings and also reports the proposed deficiency amounts. Thus, SNA measures of this sample of enterprises can potentially allow us in future work to determine the association between enterprise network characteristics and tax noncompliance. We also consider another sample, which represents all enterprises with flow-through entities. For our second sample, we randomly select 5,000 enterprises from Tax Year 2009 that have between 5 and 15 nodes, and we consider all entities associated with these enterprises. The objective of selecting this second sample is to compare the network characteristics of the proposed deficiency sample with a random sample of entities. Table 1 provides a summary of the two data samples. We extract the information about linkages between the sample entities. We consider parent-subsidiary, primary-secondary, and K-1 linkages. We also extract the financial flows associated with each link, such as gains, losses, income, interest, capital gains, rent, real estate income, etc. If there are multiple entries associated with a particular type of financial flow for a given link, we sum these values.

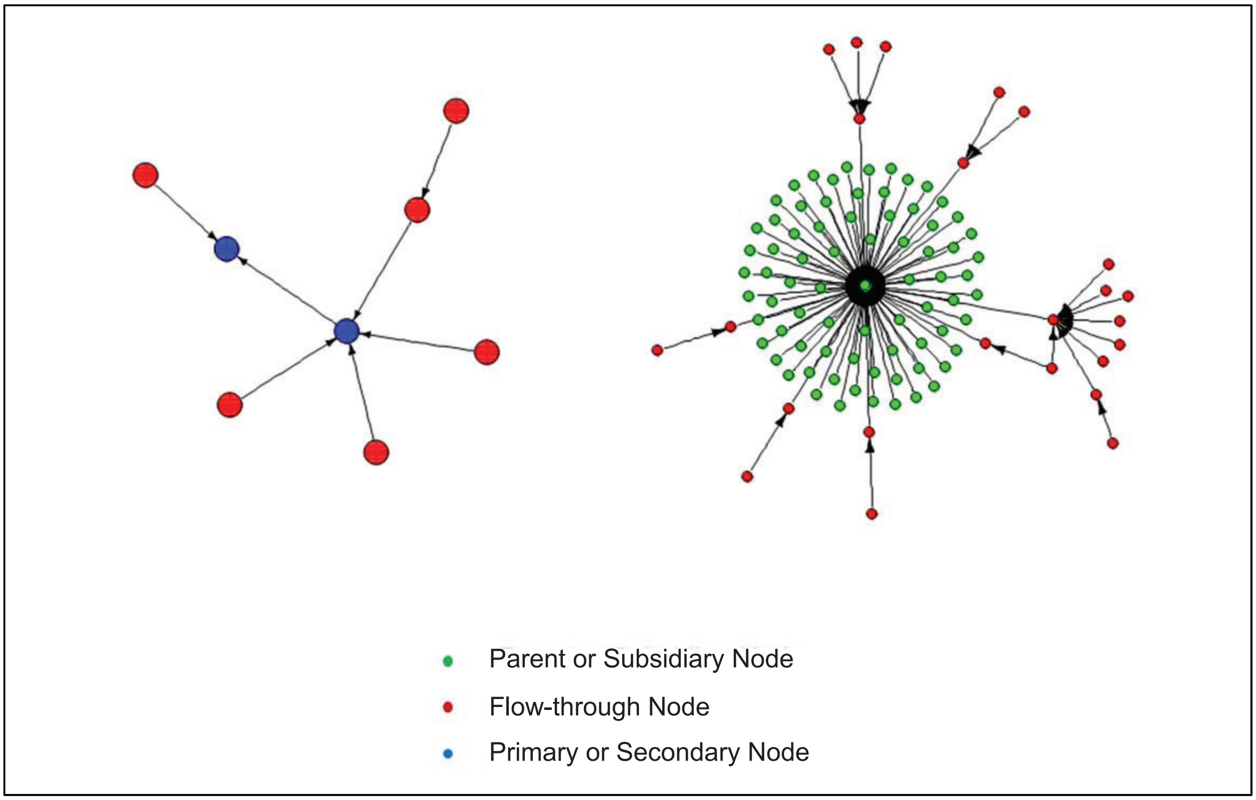
TABLE 1. Data Samples

	Sample Based on Proposed Deficiency Database	Random Sample
Year	2009	2009
Number of enterprises	5,913	5,000
Entities	107,638	31,884
K-1 links	411,644	28,210
Parent-Sub links	75,832	1,225
Primary-Secondary links	55	2,590

Social Network Analysis

We use the igraph library in R CRAN for building the network and calculating SNA measures. We treat each enterprise as an independent network and create separate graphs for each enterprise. Each enterprise graph is both a directed and weighted network. In a directed network, transactions always flow from the payer to the payee. The weights represent different types of flows where each weight can vary in magnitude. Each type of flow (income, gain, loss, dividend, interest, etc.) can have a separate weight based on the magnitude. We tag nodes based on the type of connections. Figure 1 shows two sample enterprises. We create enterprise-level SNA measures that represent the entire network associated with the enterprise. We also create node-level SNA measures that are associated with individual entities within each enterprise. These measures are described in the next section.

FIGURE 1. Sample Networks Representing Enterprises



Network Measures

Our primary objective is to determine if the network obtained by the use of flow-through entities and the associated economic flows represent tax-noncompliance risk. Recent studies suggest that complex enterprises are associated with higher levels of tax avoidance than more simple structures (Balakrishnan, *et al.* (2012), Wagener and Watrin (2013)). This suggests that complexity of the economic structure associated with the use of flow-through entities could potentially reveal tax avoidance (and perhaps evasion) behavior. To represent this complexity, we use standard network theory to develop different measures to characterize the network associated with individual entities and the enterprise. Our measures represent different combinations of the attributes of the entities involved in an enterprise and the economic flows associated with these nodes.

Network-Level Measures

Studies show that flow-through entities facilitate multistate tax avoidance and have been widely used as special purpose vehicles (Fox and Luna (2005), Feng, *et al.* (2009)). Therefore, the characteristics and distribution of flow-through entities in an enterprise may represent risk for tax noncompliance. To capture this, we include network-level measures based on the characteristics of the flow-through entities. These are described below.

1. Density

A denser network with a greater number of connections between nodes is more complex than a network with an equal number of nodes and fewer connections. Thus, we may expect a network or an enterprise with high density to have higher compliance risk than other enterprises.

Density can be defined as the ratio of the number of links present in the network to the maximum number that are logically possible, given the size of the network. It can be captured as the following ratio:

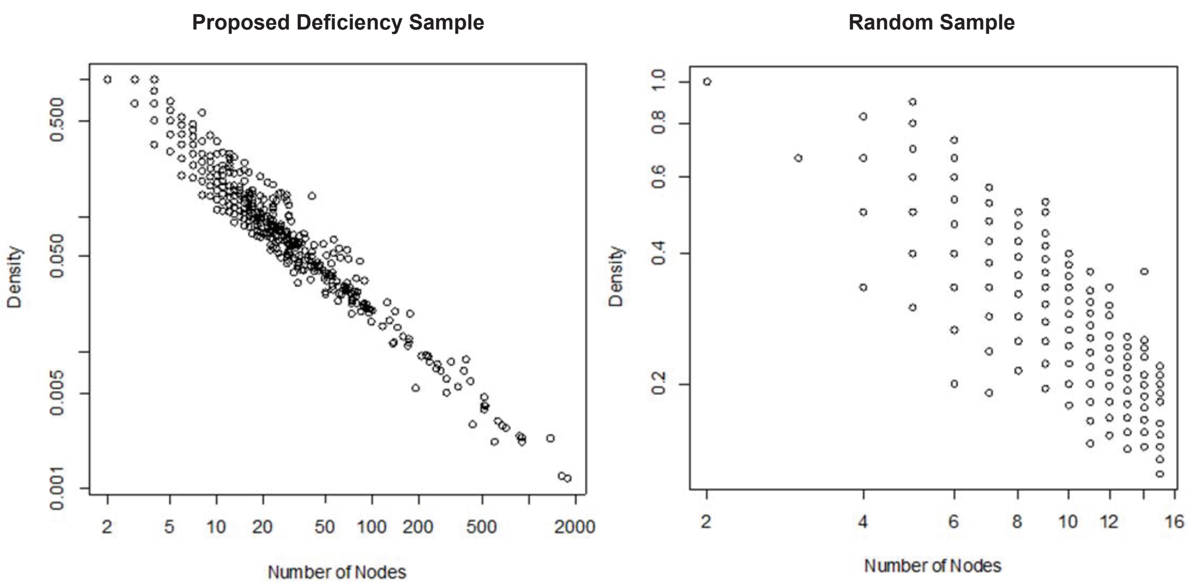
$$\frac{\text{Total Number of Links}}{\frac{1}{2}n(n-1)}$$

where

n = number of nodes within an enterprise.

Note that enterprises with fewer nodes will have higher density. In that case, to compare across enterprises of different sizes, this measure can be normalized by size. We can use either the number of nodes or total enterprise assets to represent the size. Figure 2 compares the density values associated with enterprises in our two data samples as a function of the number of nodes. (Refer to the descriptions of these two samples above.)

FIGURE 2. Enterprise Density as a Function of the Number of Nodes



2. Diversity of Nodes

A network including different types of nodes is more complex and may reflect greater tax planning than a homogenous network. Thus, we may expect a network or an enterprise with high diversity of nodes to have higher compliance risk than other enterprises. Diversity of nodes can be defined as a degree of concentration where nodes are of different types. It can be captured as the Simpson index or Herfindahl index. The measure equals the probability that two entities taken at random from the network represent the same type. It equals:

$$\sum_i p_i^2$$

where p_i represents the proportion of node of type i. Lower values of the index indicate lower concentration and thus greater diversity.

3. Loss Nodes

A network with an abnormally high proportion of flow-through entities incurring losses may reflect greater tax planning than a network with a smaller proportion of flow-through loss nodes. Economically, we would not expect an enterprise to continue operating unprofitable nodes. Thus, the presence of a high number of loss nodes could be indicative of noncompliance as the losses would provide tax savings to taxpaying entities.

Loss nodes can be defined as the proportion of flow-through entities incurring losses within the network to the total number of nodes. It can be captured as the following ratio:

$$\frac{\textit{Number of Flow-through Loss Nodes}}{\textit{Total Number of Nodes}}.$$

Note that larger enterprises are expected to have lower proportions of loss nodes due to diversification.

4. External Degree Centrality

The interaction between enterprises, perhaps through joint ventures or minority partnership interests, is more complex than a self-contained network. Hence, we may expect a network with a high number of external linkages to have higher compliance risk than other networks.

External degree centrality can be defined as the proportion of links within a network that are connected to entities external to the immediate network. It can be captured as the following ratio:

$$\frac{\begin{array}{c}\textit{Total Number of External Links}\\ \textit{Associated with Other Enterprises}\end{array}}{\textit{Total Number of Links}}.$$

5. Graph Centralization

Graph centralization represents the variation in the centrality scores of the nodes in a network. Centrality of a node reflects its importance locally (degree) or relative to the rest of the network (closeness, betweeness, etc.). A highly centralized graph represents a structure where only a few nodes are the focus of the economic activity.

Centralization can be expressed as:

$$\frac{\sum_i [C_D(\max) - C_D(i)]}{(n-1)(n-2)}$$

where $C_D(i)$ is the centrality of node *i* (see below) and *n* is the number of nodes within an enterprise.

Node-Level Measures

The purpose of developing these measures is to highlight economically important entities within an enterprise. These measures can be aggregated further across all the entities within an enterprise to come up with a composite score at the enterprise level.

1. Degree Centrality (Standardized)

Nodes with high degree centrality are connected to a greater number of entities within the network structure. This represents a higher level of complexity than nodes with few connections. Thus, we may expect a node with high degree centrality to have higher compliance risk than other nodes.

Degree centrality can be defined as the number of linkages present at each node. It can be captured as:

$$\frac{\textit{Number of Links per Node}}{n-1}$$

where *n* = number of nodes within an enterprise.

A node with a large asset balance is expected to have high degree centrality. Thus, to compare across nodes, we normalize this measure by total node-level assets.

2. Weighted Degree Centrality

Nodes with high degree centrality adjusted for their level of activity are more likely to be economically important entities within the network structure. We may expect a node with greater activity to have higher compliance risk than other nodes.

Weighted degree centrality can be defined as the number of linkages present at each node weighted by various types of economic flows. Our primary economic flows of interest are profits and losses. We can also weight this measure by interest and royalty payments as these flows are associated with tax-planning strategies using intercompany loans and intellectual property transfers. Weighted degree centrality can be captured as:

$$\textit{Number of Links per Node}^{1-\alpha} \times \textit{Node Strength}^{\alpha}$$

where

Node Strength for a node $i = \sum_{j \neq i}^{n} w_{ij},$

w_{ij} is the weight representing the economic flow between two nodes i and j, and

α is a tuning parameter determining the relative importance of number of links compared to the weight of links.

The tuning parameter can be determined by estimating the relative impact of each attribute on the desired output. Also note that the weight of a link is equal to the magnitude of the economic flow.

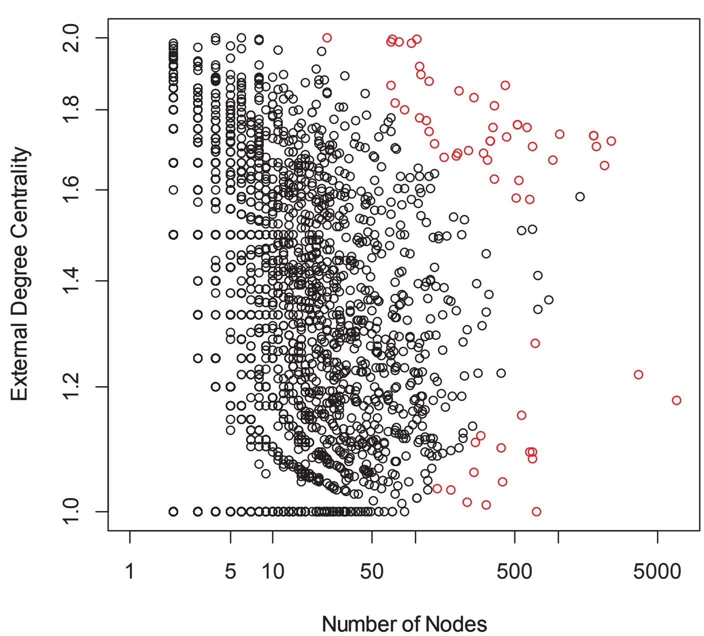
Other node-level measures are explained in the Appendix.

Application of SNA

Outlier Analysis

One of the applications of SNA is to create common measures to describe enterprises/nodes and to use these measures to find exceptions. Such exceptions in our context can potentially point to noncompliant tax behavior. For example, we can expect that the external degree centrality of an enterprise decreases as the number of nodes associated with the enterprise increases. Large deviations from the expected value of this measure can be flagged as exception. To determine such exceptions or outliers, we can carry out simple regression analysis to model expected behavior. Outliers can be identified as deviations from the fitted values obtained from the regression coefficient. Figure 3 shows the outlier enterprises (marked in red) using Cook's Distance, a commonly used measure to identify outliers.

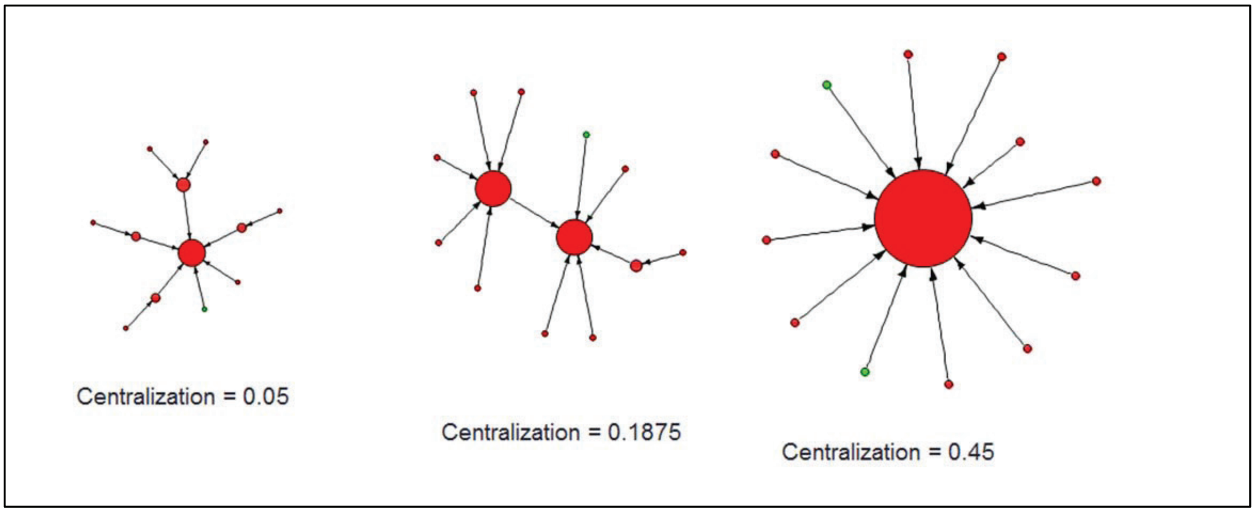
FIGURE 3. External Degree Centrality



Identifying Economically Important Nodes

Graph centralization and degree centrality of the individual nodes can be combined to determine economically important nodes. A graph with a high degree of centralization indicates that certain nodes are dominant in the enterprise. These dominant nodes can be identified using their degree centrality values. A dominant node is expected to have a higher degree centrality value. Figure 4 shows sample graphs for different enterprises, which vary in their centralization values. The node size in each graph is proportional to the degree centrality of the node. The enterprise with a high centralization value also has nodes with a much higher degree centrality value as compared to the other nodes. In these graphs, we have used standard degree centrality values. However, the same analysis can be repeated using weighted degree centrality, where one can use any of the previously defined weights.

FIGURE 4. Enterprises With Different Centralization Values



Conclusion and Future Work

We investigate how SNA techniques can be used to analyze tax structures that include flow-through entities. We show that SNA can be a useful approach to characterize these structures. It allows us to represent diverse enterprises consisting of different types of taxpaying entities and flow-through entities. Further, we can capture different types of financial flows associated with these entities. SNA provides measures at the enterprise level and the node level within an enterprise. This allows efficient comparison of these entities across several different metrics. SNA can be used further to identify exceptions both at the enterprise level and at the node level within an enterprise.

While the current work illustrates the potential of using SNA techniques to analyze flow-through entities, there are several additional avenues for future work to establish the use of SNA in predicting tax noncompliance and to operationalize these measures for implementation. These include:

a) *Measure Validation:* Future work should conduct empirical analysis using existing noncompliance data for enterprises and identify relevant SNA measures that are indicators of noncompliance.

b) *Exception or Outlier Identification:* Future work should focus on defining robust measures for identifying outliers or exceptions. This involves establishing the correct association between the size of the enterprise and the SNA measures and investigating different metrics to identify outliers. Additionally, data-mining approaches can be explored to conduct the outlier analysis.

c) *Measure Refinement:* Future work should also dig deeper into the definition of the proposed measures. This includes validation of the measures using a training dataset. Additionally, there is significant overlap in the proposed measures. Thus, future work should investigate composite measures that can be used to characterize enterprises.

d) *Enterprise Definition:* Currently, we rely on the enterprise definition based on the 50-percent rule. Future work should investigate network structures without imposing this rule and conduct the outlier analysis.

e) *Multi-Year Analysis:* Currently, we focus on a single year to establish the use of SNA. Future work should validate the SNA measures across multiple years.

Besides establishing the role of SNA to characterize tax structures, another potential avenue of research is using SNA to evaluate these structures over time. Taxpaying entities may alter their structure in response to tax law changes or audits. In that case, the measure thresholds used to identify exceptions should change over time. Future work should also incorporate this dynamic aspect in the definition of SNA measures.

References

Ayers, B., C. Cloyd, and J. Robinson. 1996. "Organizational form and taxes: An empirical analysis of small businesses." *Journal of the American Taxation Association* (18): 49–67.

Balakrishnan, K., J. Blouin, and W. Guay. 2012. "Does tax aggressiveness reduce corporate transparency?" Working paper, University of Pennsylvania.

Demere, P., M. Donohoe, and P. Lisowsky. 2015. "The economic effects of special purpose entities on corporate tax avoidance." Working paper, University of Illinios.

Feng, M., J. D. Gramlich, and S. Gupta. 2009. "Special purpose vehicles: Empirical evidence on determinants and earnings management." *The Accounting Review* 84 (6): 1833–1876.

Fox, W. F., and L. Luna. 2005. "Do limited liability companies explain declining State corporate tax revenues?" *Public Finance Review* 33 (6): 690–720.

Gordon, R., and J. MacKie-Mason. 1994. "Tax distortions to the choice of organizational form." *Journal of Public Economics* 55 (2): 279–306.

Guenther, D. 1992. "Taxes and organizational form: A comparison of corporations and master limited partnerships." *The Accounting Review* 67 (1): 17–45.

Hanlon, M., L. Mills, and J. Slemrod. 2007. "An empirical examination of big business tax noncompliance." In *Taxing Corporate Income in the 21st Century*, edited by A. Auerbach, J. Hines, Jr., and J. Slemrod, 171–210. Cambridge: Cambridge University Press.

MacKie-Mason, J., and R. Gordon. 1997. "How much do taxes discourage incorporation?" *Journal of Finance* 52 (2): 477–505.

Mills, L. 1998. "Book-tax differences and Internal Revenue Service adjustments." *Journal of Accounting Research* 36 (2): 343–356.

Wagener, T., and C. Watrin. 2013. "The relevance of complex group structures for income shifting and investors' valuation of tax avoidance." Working paper, University of Münster.

Appendix

Additional Node-Level Measures

1. Closeness Centrality (Standardized)

Nodes with low levels of closeness are likely entities within a multitier network structure, which is more complex than a flat structure. Hence, we may expect a top-level node with lower closeness to have higher compliance risk than other nodes. We may also use the closeness centrality measure of the top-level node to represent the overall closeness of the network or enterprise.

Closeness centrality can be defined as the inverse of the distance between a node and every other node in the network, where distance is measured as the number of links in the shortest path from one node to another. It can be captured as:

$$\left[\frac{\sum_{j=1}^{n} Distance\ (i,j)}{n-1}\right]^{-1}$$

where

n=number of nodes within an enterprise

i, j=node i, node j, etc.

Large enterprises can have lower closeness. To compare across enterprises we can normalize this measure by total enterprise assets.

2. Profit and Loss Asymmetry

Skewed allocations of profit and loss can be an indication of tax planning. For example, allocating a partner a high percentage of flow-through losses but a low percentage of flow-through profits does not appear to be an economically rational allocation rule. Thus, we may expect a node with a high level of profit and loss asymmetry to have higher compliance risk than other nodes.

Profit and loss asymmetry can be defined as the disparity between an entity's share of profits and share of losses. It can be captured using the "Node Strength" where:

$$Node\ Strength\ for\ a\ node\ i\ = \sum_{j \neq i}^{n} w_{ij}$$

where w_{ij} is the weight representing the economic flow between two nodes i and j.

In this case, the weight of a link is equal to the absolute value of the difference between percentage allocation of profits and percentage allocation of losses at each link. Use of absolute differences can allow us to capture all deviations.

3. Net Flows Asymmetry

Noncompliant behavior could include unusual patterns of flows between entities, particularly where net flows are not economically rational or lack economic substance. Thus, we may expect a flow-through node with a high level of net flows asymmetry to have higher compliance risk than other flow-through nodes.

Net flows asymmetry captures where inflows and outflows to and from flow-through nodes are mismatched in terms of sign (e.g., inflows are all positive but outflows are all negative). It can be calculated as:

$$Node\ Strength_{inflows} - Node\ Strength_{outflows}.$$

4. *Character Asymmetry*

Noncompliant behavior could also involve unusual patterns of types of flows between entities, such as disproportionate flows of tax-preferred income items and deductions against ordinary income. Thus, we may expect a flow-through node with a high level of character asymmetry to have higher compliance risk than other flow-through nodes.

Character asymmetry captures where inflows and outflows to and from flow-through nodes are mismatched in terms of the character of the income or loss (e.g., tax preferred vs. ordinary items). It can be calculated as:

$$\frac{\textit{Tax-Preferred Inflows (Dividends, Capital Gains)}}{\textit{Total Inflows}} - \frac{\textit{Tax-Preferred Outflows}}{\textit{Total Outflows}}$$