Board Interlocking Network and the Design of Executive Compensation Packages*

LING HENG WONG ANDRÉ F. GYGAX

21 November 2009

ABSTRACT

Directors often sit on multiple company boards. Two boards are said to be interlocked when they have at least one common director. A board interlocking network is a set of companies together with all the interlocks that exist among them. We propose that such an interlocking network is an important inter-corporate setting, which has bearing on how company boards make important corporate decisions. Using a sample of large U.S.-based public companies, board member information, the CEO compensation information, and the exponential random graph modeling techniques for social networks, we present evidence that board interlocks are positively linked with similarity in the design of CEO compensation packages in interlocked firms, particularly the proportions of stock-based components.

Keywords: Board interlocking network; Imitation; Structural p* modeling

JEL Classification Codes: C50; D03; D85; G30; M52

***** Ling Heng Wong is with the Victorian Funds Management Corporation. André Gygax is with the Department of Finance, Faculty of Economics and Commerce, The University of Melbourne. We would like to thank Peng Wang for patiently answering many of our questions about PNet. The authors also thank Lauren Cohen, Pipp Pattison, Garry Robins, and Denis Schweizer, as well as participants at the 2008 SNA Workshop "The Interdependence of Networks and Behavior: Psychological, Sociological, Political and Economic Perspectives," Melbourne Law School, and the 2009 CFF Research Conference for helpful comments. All errors are our own.

I. Introduction

Executive compensation has received an exponential growth of interest in recent years, both from the public and from the regulators. Articles in newspapers and magazines on this topic can be seen regularly[1]. A large part of this attention is because many people believe compensation packages for top executives are excessive and are getting out of line with the average worker's compensation level. For instance, according to the 2005 edition of *Executive Excess* by the Institute of Policy Studies/United for a Fair Economy (Anderson *et al.*, 2005), the ratio of an average CEO pay to an average production worker pay in the U.S. has increased from 107:1 in 1990 to 431:1 in 2004 (based on 367 leading U.S. corporations). While, after adjusting for inflation, the average CEO pay rose more than 300% during 1990–2004, the average worker's salary rose less than a mere 5% over the same period. Although CEO compensation appears excessive, Gabaix and Landier (2008) show that this pay increase can be explained by a similar increase in market capitalization of the CEO's firm during the period 1980 to 2003.

Besides the general increase in the level of executive compensation, there has also been a significant increase in the proportion of stock options in an average executive's compensation package. During the 1980s, only one fifth of total CEO compensation was in form of stock options and it grew to about one third by the mid-1990s (Hall and Liebman, 1998; Murphy 1998). Previously, this part of compensation packages were not transparent to outsiders due to a lack of relevant disclosure requirements in accounting standards and from the regulatory bodies (hence a possibility of collusion). The growing importance of stock options has subsequently led to changes in accounting standards (SFAS 123) in the U.S.[2]; effective from March 2006, SFAS 123 now requires fair values of options to be disclosed (Carter *et al.*, 2006). Also, effective from November 2006, new rulings by the SEC in the U.S. require much more stringent disclosure for executive compensation packages in general[3].

The increase in public interest in executive compensation has been accompanied by a parallel increase

[1] For example, "Taken for a ride", The Economist (July 11, 2002); "Targeting CEO Comp: The fight over executive pay obscures bigger boardroom problems", Fortune (December 12, 2005); "Investors focus on governance issues", The Age (April 20, 2006); "Risks in rewards", BRW (29 June 2006); "Inquiry at Apple – option irregularities / Stock grants to executives under scrutiny", San Francisco Chronicle (June 30, 2006); "In a free market, you have to pay top salaries to keep top players", Australian Financial Review (November 15, 2006); just to name a few.

[2] Available at `http://www.fasb.org/`

[3] See Federal Register Vol. 71, No. 174, Friday, September 8, 2006, Rules and Regulations. Available at `http://www.sec.gov/`

in the volume of academic literature on executive compensation (Hallock & Murphy, 1998). A main thrust in this academic area now is to understand why compensation packages have changed so much recently (both the increase in level and in the option proportion). Indeed, these recent trends challenge the classical principal-agent theory (see Eisenhardt (1989) for a review) because it has little power in explaining them. Some alternative theories such as the managerial power theory (Bebchuk *et al.*, 2002) and a market-based theory (Murphy and Zábojník, 2004) have been proposed. Inspired by the classical work by Granovetter (1985) on social embedding, the aim of this study is to investigate if the interlocking relationships among company boards have any bearing on the compensation package that these boards' design. As far as the authors are aware of, no previous studies have been taken in this context. Our analysis here has been made possible by some very recent and remarkable developments in social network analysis techniques (Robins *et al.*, 2001; Robins, *et al.*, 2006).

We investigate whether the existence of a board interlock between two boards is correlated with the similarity in compensation packages of the executives in the two companies. Using exponential random graph techniques, we find evidence that board interlocks are positively linked with similarity in the design of CEO compensation packages in interlocked firms, particularly the proportions of stock-based components.

Our study contributes to the corporate finance, governance and social network literature in a number of ways. First, our study demonstrates that corporate governance has to be investigated considering not only traditional economic explanations but also behavioral explanations. Second, we demonstrate that the modeling of board interlocks needs to be performed using techniques that take into account the specifics of corporate interlocking networks. Neglecting the essence of how a network behaves will result in unreliable statistical inference and incorrect conclusions. To address such concerns we use new techniques that have recently been developed in statistical network analysis. Third, our study contributes to a growing literature in finance and economics that has applied social network arguments to financial networks and contagion (Leitner, 2005; Gale and Kariv, 2007; Schweitzer *et al.*, 2009), fund performance, fund governance and asset pricing (Cohen and Frazzini, 2008; Kuhnen, 2009), option backdating (Bizjak, Lemmon and Whitby, 2009) and venture capital and entrepreneurial ventures (Hochberg, Ljungqvist and Lu, 2007; Hallen, 2008; Sorenson and Stuart, 2008).

The remainder of this study is structured as follows. Section II. reviews agency theory and optimal contracting and the managerial power theory. Section III. presents the main hypothesis. In Section IV. we discuss the methodology used in this study. In Section V. we analyze the data and present the results. Section VI. concludes the study.

II. Literature Review

A. Agency Theory and Optimal Contracting

Large modern corporations are typically characterized by the separation of ownership and control (Berle and Means, 1932; Alchian and Demsetz, 1972; Jensen and Meckling, 1976; Fama, 1980; Fama and Jensen, 1983; Jensen, 1983; etc.). The firms' shareholders (i.e. owners), who we also refer to as the principals, provide capital for the firm and own its residual claims; they do not actively participate in the firms' management and monitoring functions. On the other hand, the firms' actual controlling managers/executives, including the CEO, who we refer collectively to as the agents, are contracted by the owners to conduct the firm's day-to-day decision making and control (Fama and Jensen, 1983). In return, these agents are paid rent for their labor and time. This separation of roles has been argued as an efficient form of economic organization (Alchian and Demsetz, 1972; Fama, 1980).

The control of a firm can be further sub-divided into two main aspects (Fama and Jensen, 1983): (i) *decision management*, which involves the initiations and implementations of the firm's projects; and (ii) *decision control*, which involves the ratification of the management's initiatives and the monitoring of managements, including setting the appropriate compensation packages for executives. By default, the owners of the firm have the rights to decide on the firm's actions; but the owners often delegate all decision functions to a board of directors to act on their behalf. In turn, the board delegates the decision management function to the senior management, while retaining the decision control function.

Because (i) the objectives of the principals (shareholders) and the agents (managers) are not perfectly aligned (Ross, 1973; Jensen and Meckling, 1976), (ii) there is a difference in risk preferences between the principals and the agents (Arrow, 1971)[4]; and because (iii) the principal cannot have perfect monitoring information about the actions of agents, agency problems are said to arise because the agents do not always act in the interest of the principal. There are economic costs to the organization associated with agency problems (Alchian and Demsetz, 1972; Jensen and Meckling, 1976)[5]. The optimal contracting theory works within the principal-agency framework and argues that, by assuming the board is to seek to maximize shareholder's value, the problem of executive compensation design is equivalent to the problem of minimizing the cost of agency problems (Mirrlees, 1976; Holmstrom, 1982; Grossman and Hart, 1983). Empirical testing for the principal-agent framework has been difficult and sparse (Garen, 1997) and in general only weak evidence were obtained (see Garen (1997) and Jensen and Murphy (1990) and references therein).

B. Managerial Power Theory and its Descendants

Indeed, the principal-agency/optimal contracting theory cannot be readily used to explain recent compensation trends like the general steep increase in compensation levels, recent widespread use of stock options, the "good timing" of stock option awards with company announcements (Yermack, 1997), among other puzzles in the analysis of compensation packages (Jensen and Murphy, 1990; Garen, 1997). Subsequently, the managerial power theory (or the "Fat Cat Theory") has been proposed as a compliment to the agency framework by incorporating the power relationship between the board and the management (Bebchuk *et al.*, 2002; Bebchuk and Fried, 2003). Here, we no longer assume the independence between the two parties. The idea of the theory is that, the more relative power the managers have over the board, the more likely it is for opportunistic managers to extract excessive rents. The basic rent-extraction mechanism proposed by Bebchuk *et al.* (2002) and Bebchuk and Fried (2003) is that, first of all, managers would try to extract as much rent as possible from their firm. However, the managers very well know that a very big pay check would normally create outrage among outsiders. Therefore the directors are unlikely to approve those proposals. As a result, managers will exercise their influencing power to obscure or legitimize their extraction.

[4]Since the principal of the firm who provides capital out of their pocket (or at least arranged the relevant financing under their names) to the firm are willing to be the risk-bearer of the firm, they are assumed to be less risk-averse than the managers, who only provide management expertise in return of some relatively stable stream of income.

[5]For a succinct review on the development and on wide-ranging applications of the agency theory, see Eisenhardt (1989).

These camouflage activities are intended to result in less effective corporate governance structures permitting further increase in their relative managerial power (and hence their pay!). Compensation packages can be thought as a compromise between opportunistic rent-extraction and minimizing outrage.

One of the many ways to diminish the relative power of the board, the executive, particularly the CEO, can be involved in the selection of new board members (Shivdasani and Yermack, 1999). Studies including Mace (1971), Lorsch and MacIver (1989), and Shivdasani and Yermack (1999) have provided evidence on the influence of the CEO in selecting new board members. Shivdasani and Yermack (1999) have shown that when a CEO serves on the nominating committee for new board member (or when no such committee exists), the firm will employ fewer independent outside directors and more gray outside directors with conflicts of interest. These would then lead to a board with less monitoring power and hence increase the relative power of the management (see also Hermalin and Weisbach (1998)). Another possible way for the management to diminish the relative power of the board is to get rid of the directors that are not in the CEO's favor. Indeed, the longevity of director's board membership has been shown to be dependent on the CEO's recommendations (Bebchuk and Kahan, 1990).

The managerial power theory is useful in explaining many phenomena that cannot be fully explained by optimal contracting arguments. For instance, it has been observed in Core *et al.* (1999) that when the board of a company is more effective (i.e. when there is a larger board, more outside directors, the CEO is not the chair of the board, and outside directors are not appointed by the CEO, etc.), the CEO's pay is generally lower. Agency theory alone does not provide much explanatory power here since it merely spells out the general principal-agent conflict of interest; whereas the managerial power theory puts a "power" perceptive on top of the agency model, making this phenomena rather trivial to understand.

The most interesting situation not easily explainable by optimal contracting is the almost uniform use of at-the-money options (Bebchuk *et al.*, 2002). Murphy (1998) noted that about 95 percent of all options granted in 1992 are at-the-money. Bebchuk *et al.* (2002) argued that the opportunistic managers would naturally want stock options granted to them to be in-the-money. However outsiders would obviously be "outraged" because these options do not provide much incentive for the managers. Boards, on the other hand, would want options to be out-of-money so as to give managers more incentive. The power play between the executive and the board would result in the awards of at-the-money options (i.e. the middle ground), which are unlikely to cause much outrage but yet, favor more the manager than predicted by optimal contracting arguments. Further examples of the power of managerial theory can be found in Bebchuk and Fried (2003).

However, the managerial power theory is not universally adopted as a full explanation for recent trends in CEO compensations. The main skeptics include Kevin Murphy and his collaborators (see for instance Murphy (2000), Hall and Murphy (2002), Murphy and Zábojník (2003, 2004)). Murphy and Zábojník (2004) argued that if managerial power theory is to be used in explaining the general increase in compensation levels, we need to show that boards are becoming more and more under the captive influence of the management over recent years; however, the opposite may be happening (Holmstrom and Kaplan, 2001). Holmstrom and Kaplan (2001) argued that boards are now becoming increasingly active in monitoring: a trend that is accompanied by an increased usage of incentive-based compensation for board members (Perry, 2000). On the other hand, the managerial power theory also predicts that internally-hired CEOs should earn more than externally-hired CEOs since the former should have closer ties with the board (hence more influence). However, Murphy and Zábojník (2003) claimed to have evidence (unpublished) that externally-hired CEOs actually have higher pay. Murphy and Zábojník (2004) therefore proposed an complimentary market-based theory. Their main idea is that, due to the advent of technology, the relative importance of a CEO's firm-specific knowledge over his/her general management skills has diminished. As a result, since only general and transferable skills are priced by the labor market, the demand for CEOs with general management skills pushes their price (i.e. compensation) up. We note that this market-based theory and the managerial power theory are in fact not in contraction with each other and we can use both to understand the current compensation trends.

Even though the managerial power theory may not have the full explanatory power in explaining some current compensation trends, it is definitely an important step in understanding compensation package designs because it goes beyond of just using rational economic arguments (as in the optimal contracting theory) and puts the specific social/political contexts of the directors into the limelight. Indeed, an insightful work by Granovetter (1985) pointed out that economic actions may not always be entirely determined by economic incentives but also by the social embeddings as well. Thus, in this study, we want to understand how directors, as social beings, make decisions on compensation package designs, relative to their social settings. The particular social setting we want to focus on is the board interlocking network. These are complex social networks among company boards created by directors sitting on multiple boards (Pennings, 1980; Stokman *et al.*, 1985; Mizruchi, 1996; Robins and Alexander, 2004). By all means, board interlocking networks are not trivial: in the U.S., roughly every third director sits on more than just one company board. In the following subsections we discuss what a board interlocking network is, how it comes about, and what effects it has on connected firms. In Section 3, we shall explain why we believe that such a network may have bearing on the design of executive compensation packages.

C. Origins of Board Interlocking Networks

We define a board interlock to be a relationship created between two company boards when they share at least one common director. A board interlocking network is a collection of company boards together with all the interlocks that exist among them. For a very long time, board interlocking networks have been a contentious phenomenon because they are thought to compromise on the effectiveness of corporate governance. Indeed, it is reasonable to wonder how directors who have to put on different hats on different boards can effectively act in the interest of every company that they serve (Mizruchi, 1996).

How do board interlocks come about in the first place? Why do so many directors sit on more than just one board? There is not a single answer for such general questions. Instead there are a number of alternative, non-mutually-exclusive models and theories on interlocks, which we now discuss briefly. Based on the observation that banking and financial institutions were often at the center of a board interlocking network, the earliest interlock theory argues that it was the allocation of financial capital that drives the formation of the interlocks (Hilferding, 1910; Dooley, 1969). Directors of financial institutions sit on other boards in order to facilitate and coordinate the systematic mobilization of capital in the broad economy (Perlo, 1957). There is also evidence that companies with higher need for capital (e.g. higher debt-to-equity ratio) tend to interlock with banks (Pfeffer, 1972; Mizruchi and Stearns, 1988).

Another interlock model is the so-called control model (Allen, 1974; Zeitlin, 1974; Allen, 1978; Kotz, 1978; Mizruchi, 1982; Scott, 1982; Burt, 1983). This model describe interlocks as a result of firms wishing to control others by seeking seats on others' boards. Hence, the model predicts that firms in the economy will organize themselves to form competing "in-groups" so that these groups can increase the competitive advantage of the dominating organizations (Stokeman *et al.*, 1985). These might in fact lead to collusion and widespread poor corporate governance in the economy. As a result, since the early 20th century, the *Clayton Act* (1914) in the U.S. prohibits individuals to become directors in two organizations which are deemed competitors (see for instance, Mizruchi (1982), Mizruchi (1996), and Fich and White (2005)).

Another version of the above control model is that, instead of having firms wanting to gain control of each other and to collude, interlocks are simply manifestations of inter-dependence among companies (Pfeffer and Salancik, 1978). Here, the interlocks signify the non-competitive liaison that are explicitly for the good of all parties involved (Stokman et al., 1985). By having directors sitting on each other's boards, they can facilitate the sharing of information, experience, and other techniques (Mariolois and Jones, 1982). Useem (1984) argued that the greater the centrality a firm has in the interlock network, the more access to information it has.

The last model we are to discuss here is the upper-class cohesion model (Zeitlin, 1974; Useem, 1982; Palmer, 1983). Company directors are often from the upper class and they share similar educational backgrounds, club memberships, social circles, etc. This model proposes that interlocks are formal instruments that the people from such a "upper class" use in order to enhance the cohesion of their class.

D. Effects of Board Interlocking Network

What are the general effects of board interlocks on firms' activities? Many prior studies have shown that firms' actions are indeed related to their interlocks. For instance, in the context of corporate strategy, Davis (1991) showed that there is an interlock network diffusion process with regards to the use of poison pill as a defense strategy; Haunschild (1993) demonstrated that corporate acquisition activities are also spread over the interlocking network by some imitation processes; further, Palmer *et al.* (1993) showed that interlocks are associated with the adoption of multidivisional form in corporations, and Galaskiewicz and Wasserman (1989) showed that interlocks are linked with how corporations decide on their charitable recipients.

Under what circumstances do interlock networks affect firm's outcomes? To answer that question, we can turn to some valuable literature in organizational learning (Levitt and March, 1988). DiMaggio and Powell (1993) theorize that firms copy each other's practices in order to gain legitimacy in uncertain situations. Haunschild and Miner (1997) and Haunschild and Beckman (1998) later demonstrated that imitation behaviors over the interlocking network are more likely to happen when there is high level of uncertainty and ambiguity and when there is a lack of alternate sources of information.

III. Hypothesis

Compensation package design is an ambiguous exercise for the board. Indeed, designing a package is more of an art than a science: although optimal contracting theory (e.g. Mirrlees (1976) and Grossman and Hart (1983)) stipulate the existence of an "optimal contract" under suitable assumptions (for instance, the knowledge of manager's utility function) and other simplifying formulations, there are no demonstrated practical rules on how to map the characteristics of an executive and of the company to such an optimal package. The board has to weigh a large amount of factors such as the executive's age, his/her work style, the current stage of economic cycle, current industry environments, company's strategic directions, new business opportunities, etc.; and select from a range of cash-based and stock-based instruments in order to reach a design decision. To make this decision even more ambiguous, there is no way to determine the effectiveness of such a design since a package is specific to a certain executive and to a specific (non-repeatable) period of time. Therefore, following from the imitation theory (Haunschild (1993); Haunschild and Miner (1997); Haunschild and Beckman (1998)) discussed in the last Section, it is highly conceivable that, in making such an ambiguous but important design decision, a board may turn to its interlocking network neighbors to borrow and imitate ideas regarding their compensation practices and design approaches.

On the other hand, because package design is an ambiguous exercise, the outcome of the package design can also reflect some fundamental beliefs or characteristics of the board members. For instance, a board that puts a large emphasis on stock-based compensation may be more inclined to focus on creating overall shareholder's values and not so much on devising a "fair" plan that precisely measures the contributions of each executive. Whereas a board that puts emphasis on bonus plans may be relatively less aggressive in generating greater shareholders return and instead more inclined to devising a plan that rewards fairly the efforts and contributions of executives. What we want to argue is that design outcomes can be a proxy for important characteristics of the board. By the principle of homophyly (McPherson *et al.*, 2001) (i.e. birds of the same feather flocks) boards with similar fundamental characteristics/beliefs may be more inclined to interlocked. Further boards may form interlocks with similar-minded firms in order to enhance class-cohesion (Useem, 1982).

In the above, we have argue that interlocks may imply similarity in package designs (imitation) and that similarity in package design may imply interlocks (homophyly). Based on these, the main hypothesis we propose is that the existence of a board interlock between two boards is correlated with the similarity in compensation packages of the executives in the two companies. In other words, we expect to observe similar compensation package designs in connected firms. To test this hypothesis empirically, we focus on the compensation package of the most senior executive, namely the CEO. To proxy for the design of a compensation package, we focus on the proportion of the four most common components (described in Section 1) out of the value of the whole compensation package.

IV. Methodology

A. Representing Networks with Graphs

Interlocking networks can be conveniently represented by mathematical objects called *graphs* (e.g. see standard texts like Bollabás (1998) and Diestel (2000)). A graph $G = (\mathbb{V}, \mathbb{E})$ is composed of a set of *vertices* (or *nodes*) $\mathbb{V} = \{v_i\}_{i=1}^{N}$ representing the N firms within the interlocking network and a set of *edges* (or equivalently *links*) $\mathbb{E} = \{e_k\}_k \subset \{(v_i, v_j); v_i \neq v_j \in \mathbb{V}\}$, representing the existence of interlocks between pairs of firms. Graphs we consider here are “undirected”, i.e. (v_i, v_j) simply represent an edge between v_i and v_j: the direction is irrelevant, hence (v_i, v_j) is equivalent to (v_j, v_i). There is a one-to-one correspondence between a graph and an $N \times N$ *adjacency matrix* $X = [x_{ij}]$, where

$$x_{ij} = \begin{cases} 1, & \text{if there is a link between } v_i \text{ and } v_j \\ 0, & \text{otherwise} \end{cases}, \qquad (1)$$

and N is the number of vertices in the network. For undirected graphs, it is implicit that $x_{ij} = x_{ji}$ for all i, j, since direction is irrelevant; hence X is always *symmetric*. We subsequently will use the adjacency matrix representation in our statistical network model.

In this study, we only introduce the terms to be used in this study.[6] Two vertices are *neighbors* if there is a link between them. The set of neighbors of vertex v is $\mathcal{N}(v) = \{v_i : (v, v_i) \in \mathbb{E}\} = \{w : x_{vw} = 1\}$. The *degree* of a vertex v (call it d_v) is the number of vertices that it has a link with, so $d_v = |\mathcal{N}(v)| = \sum_{i=1}^{N} x_{iv}$[7]. A *path* from vertex a to vertex b is an ordered sequence of vertices $\{a, v_1, v_2, \ldots, b\}$ such that each successive pair of vertices in the sequence are neighbors. The *path length* is defined to be the number of edges in the ordered sequence. If there exists at least a path between two vertices, the two vertices are said to be *connected*. If there is no path between two vertices, they are *disjoint* or *disconnected* (and the path lengths of disjoint vertices are defined to be ∞). The shortest path (if it exists) between two vertices is often used as a measure of how closely they are connected within the network; such a distance is called the *geodesic* distance.

The *density* of a network is defined to be the number of realized edges in the network divided by the total number of *possible* edges in the network; hence it is always between 0 and 1. If a network has density 1 (all possible edges are realized), it is called a *complete* network. If a network has density 0, it is called an *empty* network. The *degree distribution* is defined to be the sequence $\{D_0, D_1, \ldots, D_{N-1}\}$, where D_i denotes the number of vertices with degree i, and is one of the ways to characterize a network (e.g. see Farkas *et al.* (2004)).

Now let us consider some important *local* network structures. These local structures are the building

[6] Note that there are many good introduction to terminology in graph theory in standard texts, e.g. Bollabás (1998) and Dietel (2000).

[7] If X is a set, we denote the size of X by $|X|$.

blocks of the exponential random graph (p^*) models that we shall employ.

Counts of these local structures in a network are important in our network model. As an example, for a given graph we could count the number of 2-stars, 3-stars, and triangles. Lastly, we define the *clustering coefficient* C of a network to be $C = 3S_2/T$, where S_2 and T denote the number of 2-stars and triangles in the network respectively. $0 \leq C \leq 1$ is used to measure tendency for triangles to form.

B. Basic Statistical Network Modeling

The standard approach to modeling interlocking networks in the finance/economics literature is to treat each interlock x_{ij} as an *independent* data point (e.g. Haunschild and Beckman (1998)). However, such an approach ignores the complex inter-dependency among interlocks. In fact, treating x_{ij} as independent variables ignores the essence of a network. Here, we use a more "holistic" approach which takes into account of such inter-dependency. Since our methodology here is non-standard in related literature, we give a brief introduction in the following. We assume that the network $[x_{ij}]$ comes from an distribution of networks that is specified by a parameterized *joint probability distribution* and our aim is to estimate the parameters for the probability distribution that *best fit* the empirically observed network.

The statistical network model we shall use is the class of *exponential random graph models* (also known as p^* models – we used these two names interchangeably) with early developments by Frank and Strauss (1986) and later generalized by Wasserman and Pattison (1996), Pattison and Wasserman (1999) and Robins *et al.* (1999). Let $\mathbf{X}$ be a random variable generated from the set of all $N \times N$ adjacency matrices representing the set of *all* possible realization of networks with size N; let $\mathbf{x}$ denote a realized (observed) network drawn from $\mathbf{X}$, then the most general form of the exponential random graph model is specified by the following distribution:

$$\text{Prob}(\mathbf{X} = \mathbf{x}) = \frac{1}{\kappa} \exp\left(\sum_g \lambda_g \times z_g(\mathbf{x})\right), \qquad (2)$$

where g's denote some network configurations such as edge, 2-stars, 3-stars, triangles, etc.; $z_g(\mathbf{x})$'s are network statistics relating to g in $\mathbf{x}$, such as the *count* of g in $\mathbf{x}$; λ_g is the parameter corresponding to the statistic $z_g(\mathbf{x})$ to be estimated; and finally

$$\kappa = \sum_{\mathbf{x}\in\mathbf{X}} \exp\left(\sum_g \lambda_g \times z_g(\mathbf{x})\right) \tag{3}$$

is the *normalizing constant* which ensures that $0 \leq \text{Prob}(\mathbf{X}=\mathbf{x}) \leq 1$ for all possible $\mathbf{x}$ and $\sum_{\mathbf{x}\in\mathbf{X}} \text{Prob}(\mathbf{X}=\mathbf{x}) = 1$. Park and Newman (2004) provided some mathematical justifications for the *exponential* form of the distribution using entropy maximization arguments.

A modeling task to *specify* what g's and z_g's are in Eq. 2. The choice should reflect our hypothesis on what the underlying social processes in the interlocking network are. In the following, we consider a series of different p^* model specifications in the order of complexity – the aim is to discuss why many simpler specifications are inadequate for our purpose and we would indeed need the more involved specifications to do the job.

A. Erdös-Rényi Random Graph Model

The simplest model in the class of exponential random graph models is the Erdös-Rényi or *Bernoulli* random graph model, initiated independently by Erdös and Rényi (1959) and Anatol Rapaport (1953)[8]. The probability distribution of the model is:

$$\text{Prob}(\mathbf{X}=\mathbf{x}) = \frac{1}{\kappa}\exp\left(\theta \times L(\mathbf{x})\right), \tag{4}$$

where $L(\mathbf{x})$ is the number of edges in $\mathbf{x}$ and θ is corresponding parameter (usually called the edge parameter). An alternative convenient way of specifying a p^* model is to specify the *conditional log-odd* ω_{ij} of

[8]Due to its simplicity, this model amenable to rigorous mathematical treatments and have received much attention in mathematics and physics literature. For an extensive review of these results refer to Bollobás (1985) and Janson *et al.* (2000).

each edge x_{ij}, which is defined as

$$\omega_{ij} = \log \frac{\text{Prob}(\mathbf{X}_{ij} = 1 | \text{everything else in network other than } \mathbf{X}_{ij})}{\text{Prob}(\mathbf{X}_{ij} = 0 | \text{everything else in network other than } \mathbf{X}_{ij})}. \tag{5}$$

Intuitively, the log-odd measures the likelihood for the realization of an edge x_{ij} conditional on everything else in the network. Starting from Eq. 4 we can show that for the Erdös-Rényi model,

$$\omega_{ij} = \log \frac{\exp(\theta(l+1))}{\exp(\theta l)} = \theta, \tag{6}$$

where l is the number of edges in the network without x_{ij} being present. This is saying that every edge exists with equal likelihood (constant θ) and independently of each other (ω_{ij} is not dependent on any other network statistic). Indeed, the name "Bernoulli" comes from the fact that each entry in the adjacency matrix is an identical and independently distributed (i.i.d.) *Bernoulli* trial, i.e. $\forall i, j$,

$$x_{ij} = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } 1-p \end{cases}, \tag{7}$$

for some constant $p \in [0,1]$ called the *edge probability*.

To get some intuitions for this model, let us consider the effect of the parameter θ in Eq. (4) on the actual distribution of networks: (1) when $\theta < 0$, networks with *very few* edges are favored in the distribution since the smaller $L(\mathbf{x})$ is, the less negative $\theta L(\mathbf{x})$ is and thus the larger the probability weight $\mathbf{x}$ would have ; (2) when $\theta = 0$, every network occur with *equal* probability; (3) similarly when $\theta > 0$, networks with *many* edges will be favored in the distribution.

B. The Independence Assumption

We shall demonstrate that the assumption in the Erdös-Rényi model that edges are *independent* of each other is too restrictive and unreasonable in modeling interlocking networks. Let us assume that there are three firms *a*, *b*, and *c* within an interlocking network. For this network, we have the following information: (a) a director (call him/her X) sits on the boards of both firms *a* and *b* thus creating an interlock between the firms, and (b) a director (call him/her Y) sits on the boards of both firms *b* and *c* thus creating an interlock between the firms, then based on these two pieces of information, do we know anything about whether there exists an interlock between firms *a* and *c*?

Clearly, the answer is *yes*. First of all, there is the special case that X and Y are in fact the same person, then all this information would *necessarily* imply that X (or Y!) sits on both firm *a* and *c* and there *is* an interlock between firms *a* and *c*. Secondly, even if X and Y are not the same person, various interlocking network models introduced in Section 2 provide economic/sociological arguments that supports that the two piece of information tell us something about the interlock between *a* and *c*. For instance, we consider again the "class-cohesion" model (Zeitlin, 1974; Useem, 1982; Palmer, 1983). As X and Y sit on the board of firm *b*, it is conceivable that they have a strong tie between these individuals. Say, on occasions when firm *a* want to admit new directors to its board, X may utilize the social link with Y to gain access to possible candidates. The pool of candidates that Y, because of the desire to form a cohesive group, may well come from firm *b* (or any other board that Y sits on). Thus there is a tendency to close the triangle structure among firms *a*, *b*, and *c*. Also under the "resource-dependent" model (Pfeffer and Salancik, 1978), firms have a tendency to form close triangles in order to facilitate the sharing of information and experience.

C. Markov Random Graph Model

Based on what we have just discussed, a more general and flexible assumption for a model for interlocking networks would be that an edge (v_i, v_j) is *conditionally dependent* on all other edges involving vertices v_i and v_j. Using the last example, this assumption is saying that the interlock between *a* and *c* is conditionally dependent on all interlocks involving firms *a* and firm *c*, which obviously include the interlocks with *b* created by directors X and Y. Using the Hammersley-Clifford theorem (Besag, 1974), Frank and Strauss

(1986) showed that models satisfying this assumption would admit the following exponential form:

$$\text{Prob}(\mathbf{X} = \mathbf{x}) = \frac{1}{\kappa} \exp\left(\theta L(\mathbf{x}) + \sum_{i=2}^{\infty} \sigma_i S_i(\mathbf{x}) + \tau T(\mathbf{x})\right), \tag{8}$$

where $L(\mathbf{x})$ is the number of edges in $\mathbf{x}$ and θ is its associated parameter (as in the Erdös-Rényi model); $S_i(\mathbf{x})$ is the number of i-stars in $\mathbf{x}$ and σ_i is its associated parameter; and $T(\mathbf{x})$ is the number of triangles in $\mathbf{x}$ and τ is its associated parameter. In particular, the tendency to close triangles (as demonstrated in the last subsection) can now captured by the $\tau T(\mathbf{x})$ term[9]. Further, the intuitive meaning of the star term parameters is that they will align themselves to match the degree distribution of the network. We call the model specified in Equation (8) the *Markov* random graph model.

In Equation (8), the exponent has an infinite number of terms due to the infinite sum of star terms. In order to practically fit an observed network to the model, we have to "finitize" the model. One way is to enforce an additional working assumption that the parameters associated with some higher-order star terms to be zero (this is consistent with the observation that the parameters associated with these higher-order star terms have very small magnitudes even if we fit them in). Practically, it is common to retain only the 2-star and 3-star terms, so Equation (8) can be reduced to

$$\text{Prob}(\mathbf{X} = \mathbf{x}) = \frac{1}{\kappa} \exp\left(\theta L(\mathbf{x}) + \sigma_2 S_2(\mathbf{x}) + \sigma_3 S_3(\mathbf{x}) + \tau T(\mathbf{x})\right). \tag{9}$$

From this equation, we can also easily derive the conditional log-odd ω_{ij} for this simplified model (cf. Eq. (6)):

$$\omega_{ij} = \theta + \sigma_2 \Delta_{ij}(S_2) + \sigma_3 \Delta_{ij}(S_3) + \tau \Delta_{ij}(T), \tag{10}$$

[9] In sociology literature, this closure effect is also known as the *transitivity* effect (Wasserman and Faust, 1994), i.e. "friends of friends are also likely to be your friends", which had been a main motivation for the development of this model.

where $\Delta_{ij}(\#)$ denotes the extra number of # if the interlock x_{ij} is realized. From Eq. (10), it is easy to see that edge dependency comes from the fact that ω_{ij} depends on how many local structures it creates if it is realized *in relation* to other edges. For instance, if τ is a large positive number, then this distribution of networks prefers to realign edges that have large $\Delta_{ij}(T)$, i.e. there is a tendency to complete triangle configurations.

The major difficult in obtaining maximum likelihood estimates for the parameters in Equation (9) (i.e. $\theta, \sigma_2, \sigma_3$ and τ) is that the normalizing constant κ, which itself is dependent on $\theta, \sigma_2, \sigma_3$ and τ, cannot be known *a priori* and so we cannot explicitly construct the likelihood function. Previous attempts in obtaining parameter estimates involve the use of *pseudo-likelihood* estimators (Strauss and Ikeda, 1990). However the theoretical behaviors of these estimators are not well-known and we can only accept these estimates with a certain leap of faith. In recent times, because of the advances in theoretical understanding of *simulation-based* estimations (Geyer and Thompson, 1992), the dramatic increase in cheap computing power available, and many efforts into developing efficient simulation algorithms for exponential random graph models (like PNet[10], Stocnet[11], and statnet[12]), obtaining true maximum likelihood parameters can now be efficiently done through Markov Chain Monte Carlo (MCMC) simulations. Currently this is the only feasible mean of obtaining maximum likelihood estimates for general exponential random graph (p^*) models. We shall not go into the technical details of these methods here (a good reference for which is Snijders (2002)), but the basic idea involves *simulating* a distribution of networks using some parameter estimates and then iteratively adjusting the parameter estimates until the relevant statistics of the observed network and the statistics of the simulated distribution match.

This estimation by simulation procedure works fine in many observed networks, but unfortunately for many others, this procedure fails to converge to proper estimates. This is due to the so-called *near-degeneracy* problem during a simulation (Snijders *et al.*, 2006). A brief discussion of this technical problem is necessary in order to appreciate the specification we shall use in Section 5 for our analysis.

[10] Available at `http://www.sna.unimelb.edu.au/`.
[11] Available at `http://stat.gamma.rug.nl/stocnet/`.
[12] Available at `http://www.csde.washington.edu/statnet/`.

D. Near-Degeneracy and Models with k-Statistics

Near-degenerate distributions are distributions where the vast majority of probability mass concentrates on a very small number of network configurations (Snijders *et al.*, 2006). In other words, $\text{Prob}(\mathbf{X} = \mathbf{x}) \approx 0$ for almost all $\mathbf{x}$. Although these near-degenerate distributions arise in many interesting critical phenomena problems in the theoretical physics literature like the well-known Ising model (Binney *et al.*, 1992), these distributions are problematic in our context here. To illustrate the issue, we follow a numerical example in Robins *et al.* (2006) and consider a simple exponential random graph model with only an edge term and a triangle term (i.e. the Markov model with no star terms):

$$\text{Prob}(\mathbf{X} = \mathbf{x}) = \frac{1}{\kappa} \exp\left(\theta L(\mathbf{x}) + \tau T(\mathbf{x})\right). \tag{11}$$

If we fix $\theta = -3.0$ and simulate a number of distributions for τ ranging from 0.0 to 1.5, we can observe that (1) when $\tau \to 0.0$, the distribution *degenerates* and concentrates its probability mass on the completely empty network ($x_{ij} = 0, \forall i \neq j$) (or nearly empty networks); (2) as $\tau \to 1.5$, the distribution *degenerates* and concentrates its probability mass on the complete network (i.e. $x_{ij} = 1, \forall i \neq j$) (or nearly complete networks); (3) when τ is between but away from these two ends, there exists a *coexisting* phase, meaning that the simulation either produces a degenerate distribution full of empty (or nearly empty) graphs, or a degenerate distribution full of complete (or nearly complete) graphs. Which distribution it gets "stuck in" depends on where we start the Markov Chain simulation and the particular evolution path of the Markov chain.

Such behaviors are highly undesirable in fitting observed networks because observed networks are seldom nearly empty (density $\to 0$) or nearly complete (density $\to 1$)! For instance, in our board interlocking network in the next section, the density is moderate at 0.019. Unfortunately, in many applications including ours, we would encounter combinations of parameter values (as demonstrated above) that step around the degenerate regions of the parameter space. Snijders *et al.* (2006) have proposed new types of model specifications which have been shown empirically (together with some justifications) to eliminate the likelihood of having near-degenerate distributions in simulations. They have proposed to include the so-called k-statistics into the model specification (i.e. as one of the g's and z_g's in Equation (2)). We here discuss k-star and k-triangle statistics.

The k-star statistic is defined as

$$u_S(\lambda, \mathbf{x}) = S_2(\mathbf{x}) - \frac{S_3(\mathbf{x})}{\lambda} + \frac{S_4(\mathbf{x})}{\lambda^2} + \ldots + (-1)^{N-2} \frac{S_{N-1}(\mathbf{x})}{\lambda^{N-3}} \tag{12}$$

$$= \sum_{k=2}^{N-1} (-1)^k \frac{S_k(\mathbf{x})}{\lambda^{k-2}}, \tag{13}$$

where $\lambda \in \mathbb{R}$ is an arbitrary tunable parameter, i.e. $u_S(\lambda, \mathbf{x})$ is a geometrically weighted sum of all possible "star" terms. This statistic can be used to replace the usual assumption in the Markov model that the parameters with higher-order star terms are zero (in Equation (9)). Instead the k-statistic is based on an alternative assumption that all star parameters are related in the following geometric manner,

$$\sigma_k = -\frac{\sigma_{k-1}}{\lambda}, \qquad \text{for some } \lambda \in \mathbb{R}. \tag{14}$$

This assumption means that the degree distribution is geometrically distributed and has the effect of keeping the degree distribution in check and preventing average degree to "explode" (Snijders *et al.*, 2006).

We also define k-triangle. A 2-triangle is defined to be the combination of two triangles sharing a common edge (which is called the base edge). In general, a k-triangle is defined to be the combination of k triangles all sharing a common base edge and let $u_T(\lambda, \mathbf{x})$ be its number in $\mathbf{X}$. Let $g_{ij}(\mathbf{x})$ be the number of 2-paths connecting v_i and v_j, then a useful and convenient way to combine all k-triangles $t_k(G)$ into a single measure is as follows:

$$u_T(\lambda,\mathbf{x}) = 3T_1(\mathbf{x}) - \frac{T_2(\mathbf{x})}{\lambda} + \frac{T_3(\mathbf{x})}{\lambda^2} - \ldots + (-1)^{n-3}\frac{T_{N-2}(\mathbf{x})}{\lambda^{N-3}} \tag{15}$$

$$= \lambda \sum_{i<j} x_{ij}\left[1 - \left(1 - \frac{1}{\lambda}\right)^{g_{ij}(\mathbf{x})}\right], \tag{16}$$

where $\lambda \in \mathbb{R}$ is an arbitrary tunable constant. This statistic measures the "clumpiness" of triangles and it is an aspect that the Markov model is unable to take into account. In fact, it is very likely that interlocking networks display such clumpiness (as we shall demonstrate the significance of k-triangle statistics in our models in the next section) because there are some "high-profile" directors sitting on a large number of boards. For example, one such director sitting on 5 boards would create $\binom{5}{2} = 10$ interlocks among the 5 companies (a full sub-network). Close alliance groups (Pfeffer and Salancik, 1978) may tend to form these clumpy patterns as well. So, including k-triangle statistics is likely to increase model fit in interlocking networks. For both $u_S(\lambda,\mathbf{x})$ and $u_T(\lambda,\mathbf{x})$, λ is commonly set at 2 even though its exact value is not critical in modeling.

C. Modeling Network with Firm Attributes

So far, we have considered modeling the *structure* of the interlocking networks, which is only part of the story. Our main aim in this study is to model the relationship between network structure and the *attributes* of the firms. The attribute that we are interested are the characteristics of the CEO compensation packages; let the vector $\mathbf{a} = [a_i]^T_{1\le i\le N}$ denotes a certain characteristic that we want to model and a_i is the characteristic for firm i. For instance, a_i can denote the *proportion* of stock options granted to the CEO in firm i in his/her entire compensation package. Let $\mathbf{A}$ be the random variable that generates $\mathbf{a}$ within the model. To date, there are two way on how we can model the attributes within the exponential random graph (p^*) framework:

- The first way is given by Robins *et al.* (2001a): we treat attributes $\mathbf{A}$ as *independent* variables and the interlocks $\mathbf{X}$ as dependent on $\mathbf{A}$. We specify the model by specifying the form for $\text{Prob}(\mathbf{X}=\mathbf{x}|\mathbf{A}=\mathbf{a})$. This approach is used to interpret the relationship observed between attributes and network ties as *social selection* effects: a process that establish or kill network links *because of* the attributes that the firms have.

- The second way is given by Robins *et al.* (2001b): we treat interlocks $\mathbf{X}$ as *independent* variables and the attributes $\mathbf{A}$ as dependent on $\mathbf{X}$. We specify the model by specifying the form for $\text{Prob}(\mathbf{A} = \mathbf{a}|\mathbf{X} = \mathbf{x})$. This approach is used to interpret the relationship observed between attributes and network ties as *social influence* effects: a process that affects the firm's attributes because of the links that the firms have.

Now the question for us is: which way shall we use? First, we note that it is beyond the power of *static* models[13] like the exponential random graph models to truly disentangle network influence and network selection processes[14]. Only dynamic models[15], such as Huisman and Snijders (2003), Steglich *et al.* (2005) or Koskinen and Snijders (2007), can have such a distinguishing power. As a result, the main difference between the above two modeling approaches is mainly on the interpretation (Robins *et al.*, 2001a). They merely interpret the correlation observed in the network by enforcing a particular causal direction. Since we only want to demonstrate a correlation, in this study we shall arbitrarily use the social selection approach (Robins *et al.*, 2001a).

Understanding this social selection modeling approach is best done by the way of an example. Say, we conjecture that our interlocking network has Markov edge dependency and on top of that firms select each other based on the similarity and combined level of an attribute $\mathbf{a} = [a_i]$. Recall the conditional log-odd in the (simplified) Markov model *without* attribute in Equation (10). We can extend the log-odd expression by adding the extra attribute-dependent terms:

$$\omega_{ij} = \theta + \underbrace{\sigma_2\Delta_{ij}(S_2) + \sigma_3\Delta_{ij}(S_3) + \tau\Delta_{ij}(T)}_{\text{as in Eq. (10)}} + \underbrace{\alpha^-|a_i - a_j| + \alpha^+(a_i + a_j)}_{\text{attribute interactions}} \tag{17}$$

[13] Models that use network data at one time epoch.

[14] A classical example (Steglich *et al.*, 2005) is that if we observe two smokers are friends at a particular time, it is impossible for us to use this information alone to determine whether (1) they are smokers in the first place, and because of smoking they become friends, or (2) they are friends in the first place and one friend is a smoker and the other is not, but because they are friends now, the smoker influence the non-smoker to smoke as well.

[15] Models that use network data at more than one time epoch.

For instance, if α^- is very positive, then the log-odd ω_{ij} would increase if $|a_i - a_j|$ is very large as well. This means that the network distribution prefers to have edges that have large differences in the attribute. Conversely, if α^- is very negative, it means that the distributions do not prefer large differences and prefer more *similarity*. Similar interpretation can be applied to the $\alpha^+(a_i + a_j)$ term which accounts for the combined level effect. For completeness, we note that the conditional probability distribution of this example is given by

$$\text{Prob}(\mathbf{X} = \mathbf{x} | \mathbf{A} = \mathbf{a}) = \frac{1}{\kappa} \exp\left(\underbrace{\theta L(\mathbf{x}) + \sigma_2 S_2(\mathbf{x}) + \sigma_3 S_3(\mathbf{x}) + \tau T(\mathbf{x})}_{\text{as in Eq. (9)}} + \underbrace{\alpha^- A^-(\mathbf{x}) + \alpha^+ A^+(\mathbf{x})}_{\text{attribute interactions}} \right), \qquad (18)$$

where

$$A^-(\mathbf{x}) = \sum_{i<j} x_{ij} |a_i - a_j|, \qquad \text{and} \qquad A^+(\mathbf{x}) = \sum_{i<j} x_{ij} (a_i + a_j). \qquad (19)$$

Specific details of the social selection models we shall use will be given in the next Section.

V. Data Analysis and Results

This section is divided into four parts: first, we describe our sample details; second, we look at some basic network statistics. Then we report our structural p^* modeling results; and finally we report our social selection p^* modeling results. We conclude with a brief summary.

A. Details of Sample

The companies included in our sample board interlocking network are the 291 U.S.-based public companies that are simultaneously in **(a)** the Standard & Poors 500 constituent list (as of 30th June 2003), and **(b)** the 2003 Fortune 500 constituent list (as of April 2003). The motivation for these criteria is to ensure the companies we consider are the largest and the most representative companies in the U.S. This ensures a level of homogeneity in our sample. We note that the criterion for inclusion in each of the above two lists are different. According to the S&P 500 Fact Sheet[16], the “weighted" S&P 500 index is a proxy for the “U.S. equities, reflecting the risk and return characteristics of the broader large-cap universe on an on-going basis" and the selection is based on somewhat *subjective* representativeness of the U.S. large-cap world. On the other hand, Fortune 500 is a list of the largest U.S. companies based *objective* measures such as market capitalizations and asset sizes, etc.

Information regarding board composition in our sample companies were obtained from an SQL database collated by a well-known web site *They Rule*[17]. The database contains details of board members in 2003 Fortune 500 companies. Based on the fact that the website is widely cited and on a sanity check performed by the authors, we find little reason to question the accuracy of the data set. We have written a small C++ utility program which generates the network adjacency matrix file from a pre-processed SQL dump file from the above database. On the other hand, we obtained the CEO compensation information and other relevant company information from the widely-used Standard & Poor’s Execucomp database. The database contains compensation package details of the top five executives (including CEO) for large U.S.-based large companies, inclusive of all Fortune 500 companies.

B. Basic Network Statistics

A. Interlocking Network Structure

Because of its sheer size, it is impractical to do any meaningful visual inspection on the network. Instead, we shall look at some of its basic statistics. Our sample network has 811 distinct between-firm interlocks, i.e. the number of distinct pairs of companies which share at least one common director. This gives us a fairly low network density at 0.019. At the same time, the clustering coefficient is very high at 0.368. A combination of a low network density and a high clustering coefficient is a key signature of human social networks; these networks are commonly known in literature as “small-world” networks (Watts, 1999). Refer to Davis *et al.* (2003) for a similar demonstration that the U.S. interlocking network among the largest companies is a small-world network.

The degree distribution is depicted in Figure 1 (a). It is positively skewed (skewness = 0.75) with a

[16] Available at `http://www2.standardandpoors.com`

[17] Available at `http://www.theyrule.net`.

typical company interlocking directly with an average of 5.6 other companies; this is compatible the results in Robins and Alexander (2004). The most connected company is interlocked with 20 other companies, while 17 companies are not interlocked with any other company at all. Furthermore, most pairs of firms are *indirectly* connected (i.e. there is a path of finite lengths connecting them). The full geodesic distribution of the network is depicted in Figure 1 (b). A typical pair of companies are connected to each other in 4 steps. While there is a total of $\binom{291}{2} = 42,195$ possible pairs of firms in this network, $4,794$ pairs (11%) pairs are not connected to each other. This means that a typical company is, on average, connected (directly or indirectly) to almost 89% of other companies in this network. As a result, we can regard our sample interlocking network is highly connected.

B. Compensation Package Designs within the Network

Next we look at the distribution of compensation package across different main class: salary, bonus, stock options[18], and restricted stocks. A histogram of the proportion for each of these compensation components can be found in Figure 2. In both cases for salary and cash bonus, there is a clear "normal" proportion across the sample; for instance the norm for salary is about 10%–20% of the whole compensation package, while the norm for cash bonus is about 15%–30% of the whole compensation package. The case for stock option is very different: 44 CEOs (i.e. about 15% of the sample) have no stock options at all[19]. For those CEOs who have stock options in their packages, stock options account for about 20%–75% of their whole package. Lastly, in the case for restricted stocks, 162 CEOs (i.e. 56% of the sample) have no restricted stocks at all. For those CEOs who have restricted stocks, restricted stocks account for about 10%–60% of their whole package.

Obviously, the proportions of each component is not independent of each other. A large proportion of one component necessarily means a smaller possible proportion of other components. Hence, one would naturally expect negative correlations between each pairs of component proportions. The correlation matrix among the four main components can be found in Table 1 and the result confirms the above expectation.

[18]To avoid any confusion, we note that stock options are measured in their Black-Scholes values rather than in dollar amount that the executive exercise in that fiscal year. It is because Black-Scholes value are more relevant to how the board of directors perceives the costs of the stock option; also, the board has no direct control over how the executive exercise his/her options.

[19]We assume those CEOs whose Black-Scholes values for their stock options equals zero are those without any stock option in their package. It is because stock option granted to executives are usually *at-the-money* (Murphy, 1998).

The strongest 3 correlations (in terms of magnitude) are: that between stock option and salary (-0.38), that between stock option and bonus (-0.45), and that between stock option and restricted stock (-0.43). The only pair of proportions with positive correlation is the pair between salary and bonus (0.14). We conjecture this to be an either-or situation where a package either has a sizable stock option component and little other components, or a package that has a small, or no, stock option component and a large other components.

C. Structural p^* Modeling

Next, using a hierarchy of p^* models (without firms' attributes) as described in Section 4, we study the *pure* network processes that underpin the interlocking network. These results shall provide us with essential information when we, later in the next subsection, analyze the similarity effect on compensation packages over such a network.

A. Notes on Estimation Specifications

We used PNet[20], a software package based on MCMC simulation for exponential random graph model estimations, to perform our analysis. For each model, we initially use (1) 1 consecutive estimation run, (2) multiplicative factor = 50, and (3) 2,000 in number of phase 3 steps. However, due to degeneracy issues noted in the last section, there may be difficulties to obtain estimate convergence in some particular models; in these cases, we have employed the following practical tricks to increase the likelihood of obtaining convergent estimates. In theory, they would not introduce bias into the estimates. (1) Increase the number of simulation steps (so-called stage 3 steps); (2) Increase the number of consecutive estimation runs such that the final estimates of the previous estimation run becomes the initial parameter guesses of the next run. All these tricks have the downside of increasing the simulation run time. Further, in some cases, having a very long simulation run does not necessarily ensure convergence because a longer run can also increases the chance of getting stuck in a degenerate region of the parameter space.

[20] Available at `http://www.sna.unimelb.edu.au`.

B. Network Modeling Results

A summary of model fitting specifications and results are tabulated in Table 2. First, as a baseline model, we fit our sample network to the simple Erdös-Rènyi model (St-1) given by Eq. 4 in Section 4. In this model, the edge parameter $\theta = -3.932$ is significant by default, i.e. the model fits the network very well simply because we are ignoring any possible network dependency. The next model we fit the network to is the classical Markov random graph model (St-2) with edge, 2- and 3-star, and triangle parameters (given by Eq. 9). All parameter estimates are statistically significant, however only the edge parameter θ has goodness-of-fitness measure, t-statistic[21], less than 0.1. Large t-statistic values for α_2, α_3, and τ suggest that our sample network does not fit the Markov model well. Although the stochastic nature of our estimation method may have been a possible reason for this poor fit, we have performed numerous but failed attempts to obtain better estimations (e.g. by change initial estimates, etc.). Hence, we can reasonably deduce that the specification of model St-2 is inadequate in describing our network.

Next, we now try to add new specifications, k-statistics, to our p^* model specifications. In the third model (St-3), we add the k-triangle parameter κ_T. Recall that the k-triangle term can help to explain the "clumpy" community structures in a network. Now, refer to Table 3, all parameter estimates are statistically significant (except the simple triangle parameter τ); also all network statistics have small t-statistics less than 0.1. This a major improvement over the St-2 model in model fit. The main advantage of this model is believed to come from the power to incorporate the clumpy community effects ($\kappa_T = 0.750$ is positive and statistically significant). Note again that, the simple triangle effect is not statistically significant, even though the relevant t-statistic is very small – this means that all clustering and clumping effect has already been accounted for by the k-triangle term.

Lastly, we attempted to include k-star parameter as well over the St-3 model in St-4 model to see if it may contribute any further improvement in model fit. However, after numerous attempts, we were unable to obtain convergence in the estimates. Hence, we conclude that model St-3 is the best network model describing the pure network process of our network. For the sake of later discussion, we write the

[21] it is defined as $t = (o - e)/\sigma$, where o is the number of observed network statistic, e is the expected number of network statistic in the model, and σ is the observed standard deviation of the statistic. The smaller the t-statistic, the better the model fits with regard to that particular network statistic.

probability distribution function for model St-3:

$$\text{Prob}(\mathbf{X}=\mathbf{x}) = \frac{1}{\kappa}\exp\left(\theta L(\mathbf{x}) + \sigma_2 S_2(\mathbf{x}) + \sigma_3 S_3(\mathbf{x}) + \kappa_T(2,\mathbf{x})\right). \tag{20}$$

D. Attributed p^* Modeling

Using a hierarchy of social selection exponential random graph models, we want separate the pure structural network effects and the *similarity* effects between packages of interlocked firms. Simply as a matter of modelling choice, we interpret this similarity effect as network selection effect.

A. Hierarchy of Social Selection p^* Models

We denote the proportion of the compensation package component in question (salary, bonus, stock option, or restricted stocks) for firm i by $[p_i]$. Then we fit the sample network for each type of package component with the following p^* models in order:

1. **Erdös-Rényi model with proportion** $[p_i]$. The model to fit is:

$$\text{Prob}(\mathbf{X}=\mathbf{x}|[p_i]) = \frac{1}{\kappa}\exp\left(\theta L(\mathbf{x}) + \alpha_y^- A^-(\mathbf{x}) + \alpha_y^+ A^+(\mathbf{x})\right) \tag{21}$$

where $L(\mathbf{x}) = \sum_{i<j} x_{ij}$ is the number of links in the network, and $A^-(\mathbf{x})$ and $A^+(\mathbf{x})$ are the total differences and sums of attributes between interlocked firms, i.e.,

$$A^-(\mathbf{x}) = \sum_{i<j} x_{ij}|p_i - p_j|, \qquad \text{and} \qquad A^+(\mathbf{x}) = \sum_{i<j} x_{ij}(p_i + p_j). \tag{22}$$

A positive α_y^- value would corresponds to the preference of network configurations in the distribution for interlocked firms having very *different* proportions for the package component. Hence, a significant *similarity effect* would correspond to a significantly negative α_y^- value. **Note that α_y^- is the main parameter we want to focus on in our analysis.** The three parameters θ, α_y^-, and α_y^+ are what we shall estimate in this model.

2. **Erdös-Rényi model with proportion $[p_i]$ controlling for industry effect $[d_i]$.** Let $d_i \in \{1 \ldots 9\}$ denotes the 1-digit SIC code of firm *i* has. The model to fit here is:

$$\text{Prob}(\mathbf{X} = \mathbf{x} | [p_i], [d_i]) = \frac{1}{\kappa} \exp\left(\theta L(\mathbf{x}) + \alpha_y^- A^-(\mathbf{x}) + \alpha_y^+ A^+(\mathbf{x}) + i^= \Delta(\mathbf{x})\right) \tag{23}$$

where $\Delta(\mathbf{x})$ is defined to be the number of distinct pairs of interlocked firms within the the same industry. In other words

$$\Delta(\mathbf{x}) = \sum_{i<j} x_{ij} \delta(d_i, d_j), \qquad \text{where } \delta(d_i, d_j) = \begin{cases} 1, & \text{if } d_i = d_j \\ 0, & \text{if } d_i \neq d_j \end{cases}$$

The parameters θ, α_y^-, α_y^+, and $i^=$ are what we shall estimate. This model is to control for the possibility that there is a social selection process for interlocks based on industry.

3. **Erdös-Rényi model with proportion $[p_i]$ controlling for (a) industry effect $[d_i]$ and (b) firm size effects $[b_i]$.** Let b_i is the log market capitalization of firm *i*. The model to fit is:

$$\begin{aligned} \text{Prob}(\mathbf{X} = \mathbf{x} | [p_i], [b_i]) &= \frac{1}{\kappa} \exp\left(\theta L(\mathbf{x}) + \alpha_y^- A^-(\mathbf{x}) + \alpha_y^+ A^+(\mathbf{x}) + i^= \Delta_i(\mathbf{x})\right. \\ &\qquad \left. + \beta_{MV}^- B^-(\mathbf{x}) + \beta_{MV}^+ B^+(\mathbf{x})\right) \end{aligned} \tag{24}$$

where $B^-(\mathbf{x})$ and $B^+(\mathbf{x})$ are the sum of differences and sums of log market capitalization between interlocked firms. In other words,

$$B^-(\mathbf{x}) = \sum_{i<j} x_{ij}|b_i - b_j| \qquad \text{and} \qquad B^+(\mathbf{x}) = \sum_{i<j} x_{ij}(b_i + b_j), \tag{25}$$

Indeed, a simple correlation analysis can reveal that the proportion of various compensation components are related to the market size – refer to Table 1, Panel B. There is a strong negative correlation between proportion of salary and log market size, whereas proportions of bonus and stock options are fairly positively correlated with market size. The parameters θ, α_y^-, α_y^+, $i^=$, β_{MV}^- and β_{MV}^+ are what we shall estimate.

4. **Markov random graph model (new specifications) with proportion $[p_i]$, controlling for (a) industry effect $[d_i]$ and (b) firm size effect $[b_i]$.** The model to fit is:

$$\begin{aligned}\text{Prob}(\mathbf{X} = \mathbf{x}|[p_i],[b_i]) \quad = \quad & \frac{1}{\kappa}\exp(\underbrace{\theta L(\mathbf{x}) + \sigma_2 S_2(\mathbf{x}) + \sigma_3 S_3(\mathbf{x}) + \kappa_T u_T(\lambda,\mathbf{x})}_{\text{as in Eq. (20)}} \\ & +\alpha_y^- A^-(\mathbf{x}) + \alpha_y^+ A^+(\mathbf{x}) + i^= \Delta_i(\mathbf{x}) \\ & +\beta_{MV}^- B^-(\mathbf{x}) + \beta_{MV}^+ B^+(\mathbf{x}))\end{aligned} \tag{26}$$

where S_2, S_3 and κ_T are already defined in the last section. λ is set to 2 and its particular value is not critical to the overall modelling.

E. Network Modeling Results

A. Salary Proportion

The results of the parameter estimates are tabulated in Table 4. In the first model (pSAL-1), where there is no network dependence structure, we do not see any significant similarity effect (α_{SAL}^- not bounded away from 0.00), however there is a significant negative component *size* effect (α^{+}_{SAL} = -0.858). The negative component size effect means that interlocked firms tends to have *smaller* combined salary components. In the second model (pSAL-2), we control for the industry effect on top of the last model (pSAL-1), however, the industry parameter does not have any statistical significance in explaining the network processes. In the third model (pSAL-3), we add in the log firm size parameters as well. The similarity effect $\alpha^{-}_{SAL} = -0.887$ now becomes significant; another significant factor in this model is the *positive* total market size effect, i.e. interlocked firms tend to have higher combined market sizes. This positive combined size effect can correspond to either the case that: (1) large firms tends to interlocked with smaller firms, or (2) two large firms tends to be interlocked, or (3) a bit of both. To determine which is the most likely case, we look the similarity parameter in terms of market size, β^{-}_{MV}, which has a very small insignificant estimate in the mode. This says that interlocked firms tend to be of similar sizes; as a result, case (2) is most likely to be the case. In summary, the similarity effect for salary component was previously confounded by the combined firm value effect. Now that the latter effect is controlled for, similarity for salary proportion effect becomes visible.

In the final full network model (pSAL-4), we control for network processes by fitting the network with the social selection Markov model with new specifications. Now, the similarity effect α^{-}_{SAL} has decreased in magnitude and has becomes statistically insignificant again. Also, the parameter estimates for the market size effects (β^{-}_{MV} and β^{+}_{MV}) have now decreased in magnitude. These suggested that part of the previous similarity and total size effects are in fact due to pure network dependencies. These dependency effects cannot be accounted for in the earlier 3 models (pSAL-1 to pSAL-3). The only statistically significant factors in the full network model are the total market size effect and various network effects.

Indeed, the above analysis (pSAL-1 to pSAL-4) highlights the importance of using the exponential random graph (p^{*}) approach. Had we used the standard approach found in the finance/economics literature (i.e. treating each individual edge as an independent variable), then we could have mistaken the pure network dependency effect as similarity effect in the proportion of salary.

B. Proportion of Bonus

Refer to Table 5 for a summary of results. In the first simple model (pBON-1), we have both significant negative similarity effect (α^-_{BON} = -0.593) and significant positive combined proportion size effect (α^+_{BON} = 0.575) on the proportion of bonus. In the second model (pBON-2), we control for the industry effect, however the above two effects remain significant and the industry effect is insignificant as in the salary case. In the third model (pBON-3), we control for the firm size effects as well; here both the similarity and combined effect for firm sizes are significant. Similar conclusion about these effects can be drawn as in the salary case. On top of that, we note that the combined size effect for bonus component is now a lot smaller in magnitude ($\alpha^+_{BON} = 0.556 \rightarrow 0.378$ from pBON-2 to pBON-3). This can be partly explained by the fact that firm sizes are positively correlated with the proportion of bonus component (we have estimated $\rho = 0.11$ in our network, see Table 2, Panel B). Finally in the full network model (pBON-4), surprisingly, the similarity effect on the proportion of bonus disappeared. This shows that the previous effect are simply due to pure network dependencies.

C. Proportion of Stock Options

Refer to Table 6 for a summary of results. In the first model (pSTO-1), there is clearly a strong similarity effect ($\alpha^-_{STO} = -0.540$), assuming all interlocks are independent. In the second model (pSTO-2), as can be inferred from the previous 2 models, there is no significant industry effects and the similarity effect remains significant. In the third model (pSTO-3), there is still a similar level of similarity effect for the proportion of stock options taking into account of the standard errors of the estimates. Finally, in the full network model (pSTO-4), the similarity effect remains strongly significant ($\alpha^-_{STO} = -0.471$).

D. Proportion of Restricted Stocks

Refer to Table 7 for a summary of results. In the first model (pRST-1), there is no significant similarity or size effect for the restricted stock component. In the second model (pRST-2), again, there is no significant similarity or size effect even if we control for industry. In the third model (pRST-3), we control for market sizes and we obtain a similar estimates for the difference and sum effects as other models. However, none of α^{-}_{RST} and α^{+}_{RST} are significant. In the final model (pRST-4), the similarity effect remains significant ($\alpha^{-}_{RST} = -0.048$), although the magnitude is smaller than that observed in the stock option model (pSTO-4).

F. Summary

Using a hierarchy of social selection p^* models (Robins *et al.*, 2001a), we have shown evidence that there is significant similarity effect for the proportions of stock option and for the proportions of restricted stocks within the board interlocking network, even after we control for structural network dependency effects, industry effects and market size effects. The models we have used interpret these similarity effects as firms selecting their interlocking partners based on the proportion of stock options and on the proportion of restricted stocks. On the other hand, we do not find significant similarity effects for the proportion of salary and bonus.

VI. Conclusion

In this study, we have hypothesized that there is a link between board interlocks and the similarity of compensation package designs because of imitation effects (Haunschild and Beckman, 1998) and class-cohesion/homophily effects (Useem, 1982; McPherson *et al.*, 2001). While these two effects have opposite causal directions, they nevertheless lead to interlocked firms having more similar package designs. We have used a hierarchy of *social selection* p^* models (Robins *et al.*, 2001a) to model our interlocking network and the results provide evidence that interlocked firms do have more similar proportions of *stock-based* components (stock options and restricted stocks). For the cases of salary and cash bonus, however, we do not find such a relationship.

While we have used the *social selection* p^* models (Robins *et al.*, 2001a) for our analysis here, it may be interesting to also interpret the relationship between interlocks and the similarity of compensation package design using a *social influence* p^* model (Robins *et al.*, 2001b). However such an endeavour is outside the scope of this study. More importantly, we have pointed out that the p^* class of models actually have no real power in *disentangling* social selection and social influence processes within the network (i.e. whether interlocks affect design, or design affect interlocks, or both) because of its static nature. Both the social selection or social influence p^* model simply enforces a particular *a priori* causal direction (and hence we would expect the parallel social influence p^* analysis on our current dataset to give similar results as in this study). In order to provide *direct evidence* for the causal direction, one needs to use dynamic network models like Steglich *et al.* (2005), which would require much more involved analysis and takes much longer in simulation-based estimation procedures.

Nevertheless, we believe that determining the casual direction is important and is a natural follow-up to this study because of the following. We have discussed earlier that stock-based components have become the dominant compensation component in recent years. As far as we know, there have not been many satisfactory explanations for such a trend (except a discussion by Holmstom and Kaplan (2001) that relates the trend to the increasing dominance of institutional investors). Indeed, given the importance of social contexts in economic decisions (Granovetter, 1985), and given our empirical results here regarding similarity in stock-based compensation proportions between interlocked firms, an important question to ask is: can we relate the current trend with the possibility that the use of stock-based compensation is *diffused* through the interlocking network (i.e. interlock implies similarity)? We find parallels to the case of spreading the use of poison pills over the interlocking network, as demonstrated by Davis (1991). We believe that our results in this study are instrumental in warranting further analysis along these lines.

References

[1] Alchian, A.A., Demsetz, H., 1972. Production, information costs and economic organization, *American Economic Review* **62**, 777–795.

[2] Allen, M.P., 1974. The structure of interorganizational elite cooptation, *American Sociological Review* 39, 393–406.

[3] Allen, M.P., 1978. Economic interest groups and the corporate elite structure, *Social Science Quarterly* 58, 597–615.

[4] Anderson, S., Cavanagh, J., Klinger, S., Stanton, L., 2005. *Executive Excess 2005* (Institute for Policy Studies and United for a Fair Economy)

[5] Arrow, K., 1971. *Essays in the theory of risk bearing* (Markham, Chicago).

[6] Bebchuk, L.A., Fried, J.M., 2003. Executive compensation as an agency problem, *Journal Economic Perspectives* 17, 71–92.

[7] Bebchuk, L.A., 1990. A framework for analyzing legal policy towards proxy contests, *California Law Review* 78, 1071–1135.

[8] Berle, A., Means, G., 1932. *The modern corporation and private property* (Macmillan, New York).

[9] Besag, J.E., 1974. Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society, Series B* 36 96–127.

[10] Binney, J.J., Dowrick, N.J., Fisher, A.J., Newman, M.E.J., 1992. *The theory of critical phenomena* (Clarendon Press, Oxford).

[11] Bizjak, J., Lemmon, M., Whitby, R., 2009. Option backdating and board interlocks, *Review of Financial Studies* 22(11), 4821-4847.

[12] Bollabás, B., 1998. *Modern graph theory* (Springer, New York).

[13] Bollabás, B., 1985. *Random graphs*, 2nd Ed. (Cambridge University Press)

[14] Burt, R.S., 1983. *Corporate Profits and Cooptation: Networks of Market Constraints and Directorate Ties in the American Economy* (Academic Press, New York)

[15] Bryan, S., Hwang, L., Lilien, S., 2000. CEO Stock-Based Compensation: An Empirical Analysis of Incentive-Intensity, Relative Mix, and Economic Determinants, *Journal of Business* 73, 661–693.

[16] Carter, M.E., Lynch, L.J., Tuna, I., 2006. *The Role of Accounting in the Design of CEO Equity Compensation*, Knowledge@Wharton.

[17] Cohen, L., Frazzini, A., 2008. Economic links and predictable returns, *Journal of Finance* 63(4), 1977-2011.

[18] Cohen, L., Frazzini, A., Malloy C.J., 2008. The small world of investing: Board connections and mutual fund returns, *Journal of Political Economy* 116(5), 951-979.

[19] Core, J.E., Holthausen, R.W., Larcker D.F., 1999. Corporate governance, chief executive officer compensation, and firm performance, *Journal of Financial Economics* 51, 371–406.

[20] Davis, G.F., 1991. Agents without principles? The spread of the poison pill through the intercoporate network, *Administrative Science Quarterly* 36, 583–613.

[21] Davis, G.F., Yoo, M., Baker, W.E., 2003. The small world of the American corporate elite, 1982–2001, *Strategic Organization* 1, 301–326.

[22] DiMaggio, P.J., Powell, W.W., 1983. The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields, *American Sociological Review* 48 147–160.

[23] Dolley, P., 1969. The interlocking directorate, *American Economic Review* 59, 314–323.

[24] Eisenhardt, K.M., 1989. Agency theory: An assessment and review, *Academy of Management Review* 14, 57–74.

[25] Erdös, P., Rényi, A., 1959. *On random graphs*. I., Publicationes Mathematicae (Debrecen) 6, 290–297.

[26] Farkas, I., Derényi, I., Palla, G., Vicsek, T., 2004. *Equilibrium statistical mechanics of network structures* (Springer, Berlin).

[27] Fama, E.F., 1980. Agency problems and the theory of the firm, *Journal of Political Economy* 88, 288–307.

[28] Fama, E.F., Jensen, M.C., 1983. Separation of ownership and control, *Journal of Law and Economics* 26, 301–325.

[29] Frank, O., Strauss, D., 1986. Markov graphs, *Journal of the American Statistical Association* 81, 832–842.

[30] Friedman, M., Savage, L.J., 1948. The Utility Analysis of Choices Involving Risk, *Journal of Political Economy* 56, 279–304.

[31] Gabaix, X., Landier, A., 2008. Why has CEO pay increased so much? *Quarterly Journal of Economics* 123(1), 49-100.

[32] Galaskiewicz, J., Wasserman, S., 1989. Mimetic processes within an interorganization field: An empirical test, *Administrative Science Quarterly* 34, 454–479.

[33] Gale, D.M., Kariv, S., 2007. Financial networks, *American Economic Review* 97(2), 99-103.

[34] Garen, J.E., 1997. Executive compensation and principal-agent theory, *Journal of Political Economy* 102, 1175–1199

[35] Geyer, C.J., Thompson, E.A., 1992. Constrained Monte Carlo Maximum Likelihood for Dependent Data, *Journal of the Royal Statistical Society. Series B* 54, 657–699.

[36] Grossman, S.J., Hart, O.D., 1983. An analysis of the principal-agent problem, *Econometrica* 51, 7–45.

[37] Hall, B.J., Liebman, J.B., 1998. Are CEOs really paid like bureaucrats?, *Quarterly Journal of Economics* 113, 653–691.

[38] Hallen, B.L., 2008. The causes and consequences of the initial network position of new organizations: From whom do entrepreneurs receive investments? *Administrative Science Quarterly* 53(4), 685-718.

[39] Haunschild, P.R., 1993. Interorganizational imitation: The impact of interlocks on corporate acquisition activity, *Administrative Science Quarterly* 39, 564–592.

[40] Haunschild, P.R., Miner, A.S., 1997. Modes of interorganizational imitation: The effects of outcome salience and uncertainty, *Administrative Science Quarterly* 42, 472–500.

[41] Haunschild, P.R., Beckman, C.M., 1998. When do interlocks matter? Alternate sources of information and interlock influence, *Adminstrative Science Quarterly* 43, 815–844.

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

[43] Holmstrom, B., 1979. Moral hazard and observability, *Bell Journal of Economics* 10, 74–91.

[44] Holmstrom, B., Kaplan, S., 2001. Corporate Governance and Merger Activity in the United States: Making Sense of the 1980s and 1990s, *Journal of Economic Perspectives* 15, 121–144.

[45] Huisman, M., Snijders, T.A.B., 2003, Statistical analysis of longitudinal network data with changing composition, *Sociological Methods & Research*, 32(2), 253-287.

[46] Janson, S., Luczak, T., Ruciński, A., 2000. *Random graphs* (John Wiley, New York).

[47] Jensen, M., Meckling, W., 1976. Theory of the firm: Managerial behavior, agency costs, and ownership structure, *Journal of Financial Economics* 3, 305–360.

[48] Jensen, M., Murphy, K.J., 1990. It's not how much you pay, but how, *Harvard Business Review* 69, 138–153.

[49] Bayesian inference for dynamic social network data, *Journal of Statistical Planning and Inference* 137(12), 3930-3938.

[50] Kotz, D.M., 1978. *Bank control of large corporations in the United States.*

[51] Kuhnen, C.M., 2009, Business networks, corporate governance, and contracting in the mutual fund industry, *Journal of Finance* 64(5), 2185-2220.

[52] Leitner, Y., 2005. Financial networks: contagion, commitment, and private sector bailouts, *Journal of Finance* 60(6), 2925-2953.

[53] Lorsch, J.L., MacIver, E.M., 1989. *Pawns or potentates? The reality of America's corporate boards* (Hardward Business School Press, Boston).

[54] Mace, M.L., 1971. *Directors: Myth and Reality* (Harvard Business School Press, Boston).

[55] Mariolois, P., Jones, M.J., 1982. Centrality in corporate interlock networks: Reliability and stability, *Adminstrative Science Quarterly* 27, 571–584.

[56] McPherson, M., Smith-Lovin, L., Cook, J.M., 2001. Birds of a Feather: Homophily in Social Networks, *Annual Review of Sociology* 27, 415–444.

[57] Mirrlees, J.A., 1976. The optimal structure of incentives and authority within an organization, *Bell Journal of Economics* 7, 105–131.

[58] Mizruchi, M.S., 1982. *The American Corporate Network: 1904–1974* (Sage, Beverly Hills).

[59] Mizruchi, M.S., 1996. What do interlocks do? An Analysis, Critique, and Assessment of Research on Interlocking Directorates, *Annual Review of Sociology* 22, 271–298.

[60] Mizruchi, M.S., Stearns L., 1988. A longitudinal study of the formation of interlocking directorates, *Administrative Science Quarterly* 33, 194–210.

[61] Murphy, K.J., 1998. Executive compensation in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics* (North-Holland, Amsterdam).

[62] Murphy, K.J., Oyer, P., 2003. Discretion in Executive Incentive Contracts, Working Paper, University of Southern California.

[63] Murphy, K.J., Zábojník, J., 2003. Managerial capital and the market for CEOs, unpublished, University of Southern California.

[64] Murphy, K.J., Zábojník, J., 2004. CEO pay and appointments: A market-based explanation for recent trends, *American Economic Review* 94, 192–196.

[65] Ofek, E., Yermack, D., 2000. Taking stock: Equity-based compensation and the evolution of managerial ownership, *Journal of Finance* 55, 1367–1384.

[66] Palmer, D., 1983. Broken ties: Interlocking Directorates and intercoporate coordination, *Administrative Science Quarterly* 28, 40-55.

[67] Palmer, D., Jennings, P.D., Zhou, X., 1993. Late adoption of the multidivisional form by large U.S. corporations: Institutional, political and economic accounts, *Administrative Science Quarterly* 17, 382–394.

[68] Pattison, P., Wasserman, S., 1999. Logit models and logistic regressions for social networks: II. Multivariate relations., *British Journal of Mathematical and Statistical Psychology* 52, 169–193.

[69] Park, J., Newman, M.E.J., 2004. Statistical mechanics of networks, *Physical Review E* 70, 066117.

[70] Pennings, J.M., 1980. *Interlocking Directorates* (Jossey-Bass Publishers, San Francsico and London).

[71] Perlo, V., 1957. *The Empire of High Finance* (International Publisher, New York).

[72] Perry, T., 2000. Incentive compensation for outside directors and CEO turnover. Presented at Tuck-JFE Contemporary Corporate Governance Conference. Available at SSRN: http://ssrn.com/abstract=236033.

[73] Pfeffer, J., 1972. Size and composition of corporate boards of directors, *Administrative Science Quarterly* 17, 218–228.

[74] Pfeffer, J., Salancik, G., 1978. *The external control of organizations: A resource dependence perspective* (Harper and Row, New York).

[75] Rapoport A., 1953. Spread of information through a population with socio-structural bias: I. assumption of transitivity, *Bulletin of Mathematical Biology*. 15, 523–533.

[76] Robins, G., Alexander, M., 2004. Small worlds among interlocking directors: Network structure and distance in bipartite graphs, *Computational and Mathematical Organization Theory* 10, 69–94.

[77] Robins, G., Elliot, P., Pattison, P., 2001a. Network models for social selection processes, *Social Networks* 23, 1–30.

[78] Robins, G., Pattison, P., Elliot, P., 2001b. Network models for social influence processes, *Psychometrika* 66, 161–190.

[79] Robins, G., Snijders, T.A.B., Wang, P., Handcock, M., Pattison, P., 2007. Recent developments in exponential random graph (p^*) models for social networks, *Social Networks* 29(2), 192-215.

[80] Ross, S., 1973. The economic theory of agency: The principal's problem, *American Economic Review* 63, 134–139.

[81] Schweitzer, F., Fagiolo, G., Sornette, D., Vega-Redondo, F., White, D.R., 2009, Economic networks: What do we know and what do we need to know? *Advances in Complex Systems* 12(4-5), 407-422.

[82] Shivdasani, A., Yermack, D., 1999. CEO involvement in the selection of new board members: an empirical analysis, *Journal of Finance* 54, 1829–1853.

[83] Snijders, T.A.B, 2002. Markov chain Monte Carlo estimation of exponential random graph models, *Journal of Social Structure* 3(2).

[84] Snijders, T.A.B., Pattison, P.E., Robins, G.L., Handcock, M.S., 2006. New specifications for exponential random graph models, *Sociological Methodology* 36, 99–153.

[85] Sorenson, O., Stuart, T.E., 2008. Bringing the context back in: Settings and the search for syndicate partners in venture capital investment networks, *Administrative Science Quarterly* 53(2), 266-294.

[86] Steglich, C., Snijders, T.A.B., Pearson, M., 2004. Dynamic Networks and Behavior: Separating Selection from Influence, submitted (available at http://stat.gamma.rug.nl/snijders/).

[87] Stokman, F.N., Ziegler R., Scott, J., 1985. *Networks of corporate power : a comparative analysis of ten countries* (Polity Press, Oxford)

[88] Strauss, D., Ikdeda, M., 1990. Pseudolikelihood Estimation for Social Networks, *Journal of the American Statistical Association* 85, 204–212.

[89] Useem, M., 1982. Classwide rationality in the politics of managers and directors of large corporations in the United States and Great Britain, *Administrative Science Quarterly* 27, 199-226.

[90] Useem, M., 1984. *The inner circle* (Oxford University Press, New York).

[91] Wasserman, S., Faust, K., 1994. *Social Network Analysis: methods and applications* (Cambridge University Press).

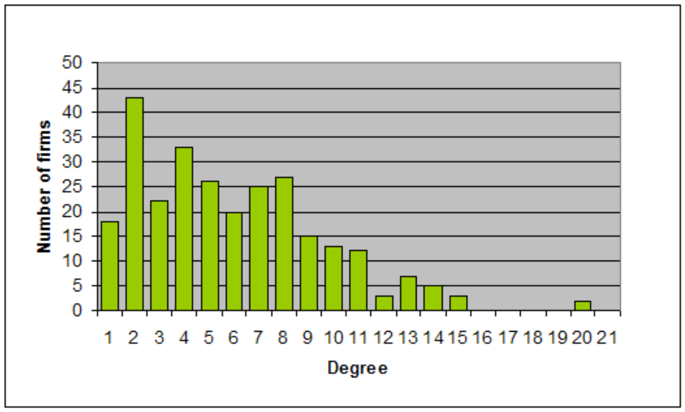
[92] Wasserman, S., Pattison, P., 1996. Logit models and logistic regression for social networks: I. An introduction to Markov graphs and p^*, *Psychometrika* 61, 401–425.

[93] Watts, D.J., 1999. *Small Worlds: The dynamics of networks between order and randomness* (Princeton University Press, New Jersey).

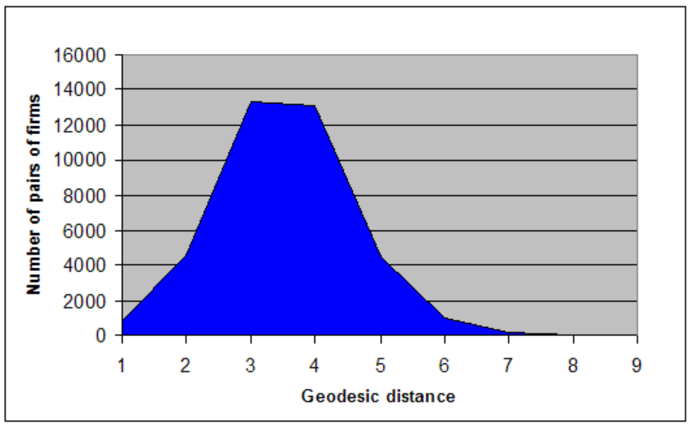
[94] Yermack, D., 1997. Good Timing: CEO Stock Option Awards and Company News Announcements, *Journal of Finance* 52, 449–476.

[95] Zajac, E.J., 1988. Interlocking directorates as an interorganizational strategy: A test of critical assumptions, *Academy of Management Journal* 31, 428–438.

[96] Zeitlin, M., 1974. Corporate ownership and control: The large corporation and the capitalist class, *American Journal of Sociology* 79, 1073–1119.

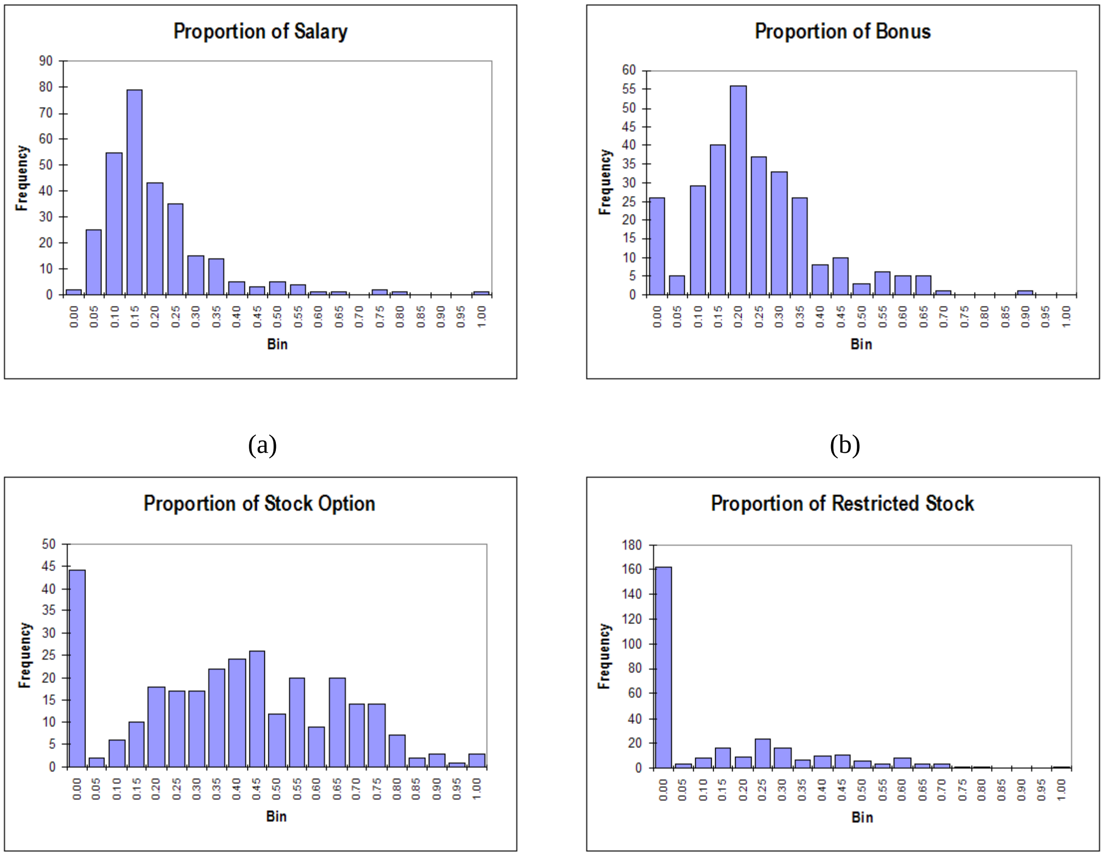


(a)



(b)

Figure 1: Degree distribution (a) and geodesic distance distribution (b) of the sample interlocking network.



(a) (b) (c) (d)

Figure 2: Histograms fo proportions of four most common forms of compensation within the toal compensation package: (a) salary, (b) bonus, (c) stock option, and (d) restricted stocks.

Table 1

Correlations

Panel A shows the correlations among the proportions of salary (pSAL), bonus (pBON), stock options (pSTO), and restricted stocks (pRST). Panel shows the correlations of above variables with size as measured by the market value (natural logarithm of market capitalization) of the firm.

Panel A: Compensation components				
	pSAL	**pBON**	**pSTO**	**pRST**
pSAL	-	0.14	-0.38	-0.25
pBON	-	-	-0.45	-0.22
pSTO	-	-	-	-0.43
pRST	-	-	-	-
Panel B: Compensation components and firm size				
ln(MV)	-0.31	0.11	0.15	0.06

Table 2

Salary Component

Structural and attribute parameter estimates using various attributed p^* models. These estimates are quoated as "Mean $\pm$ Std. dev.". Estimates with "$\star$" symbols are estimates whose confidence intervals (Mean $\pm$ 2 $\times$ Std. dev.) are bounded away from 0.000 and the corresponding network structure has t-statistic of less than 0.1.

	[St-1] **Erdös-Rényi model**	[St-2] Markov **Model 2/3-stars;1-△**	[St-3] **New specifications 2/3-stars; 1/k-△**	[St-4] **New specifications 2/3/k-star;1/k-△**
θ	$-3.932 \pm 0.036\star$	$-5.785 \pm 0.216\star$	$-5.343 \pm 0.157\star$	
	$(t = 0.033)$	$(t = 0.037)$	$(t = -0.083)$	
σ_2	-	0.237 ± 0.038	$0.102 \pm 0.026\star$	unable
		$(t = 0.509)$	$(t = 0.066)$	
σ_3	-	-0.030 ± 0.007	$-0.012 \pm 0.004\star$	to
		$(t = 1.116)$	$(t = 0.054)$	
τ	-	0.944 ± 0.099	0.177 ± 0.259	obtain
		$(t = 3.971)$	$(t = -0.056)$	
κ_S	-	-	-	convergence
κ_T	-	-	$0.750 \pm 0.132\star$	
			$(t = 0.060)$	

Table 3

Simulation Specification

This table shows a summary of all specifications for all the models in this study.

Model	Consecutive run(s)	Number of Stage 3 Steps
pSO-1	1	5,000
pSO-2	1	10,000
pSO-3	2	25,000
pSO-4	1	1,000
pSAL-1	1	25,000
pSAL-2	1	20,000
pSAL-3	1	20,000
pSAL-4	2	2,000
pBON-1	2	20,000
pBON-2	1	20,000
pBON-3	2	25,000
pBON-4	1	1,000
pRST-1	1	20,000
pRST-2	2	25,000
pRST-3	2	25,000
pRST-4	1	5,000

Table 4

Salary Component

Structural and attribute parameter estimates using various attributed p^* models. These estimates are quoated as "Mean ± Std. dev.". Estimates with "⋆" symbols are estimates whose confidence intervals (Mean ± 2 × Std. dev.) are bounded away from 0.000 and the corresponding network structure has t-statistic of less than 0.1.

	[pSTO-1] **Model St-1 +pBLK**	[pSTO-2] **Model St-1 +pBLK+ ind**	[pSTO-3] **Model St-1 +pBLK+ln(MB)**	[pSTO-4] **Model St-3 +pBLK+ln(MV)**
α^{-}_{STO}	-0.152 ± 0.403	-0.179 ± 0.400	$-0.887 \pm 0.404\star$	-0.605 ± 0.335
	$(t = -0.924)$	$(t = -0.019)$	$(t = -0.023)$	$(t = -0.014)$
α^{+}_{STO}	$-0.858 \pm 0.287\star$	$-0.847 \pm 0.288\star$	-0.241 ± 0.293	0.233 ± 0.226
	$(t = -0.040)$	$(t = 0.012)$	$(t = 0.010)$	$(t = -0.017)$
$i^{=}$	-	0.127 ± 0.0913	0.119 ± 0.090	0.112 ± 0.090
		$(t = 0.058)$	$(t = 0.033)$	$(t = -0.024)$
β^{-}_{MV}	-	-	-0.072 ± 0.038	-0.038 ± 0.033
			$(t = -0.024)$	$(t = 0.009)$
β^{+}_{MV}	-	-	$0.263 \pm 0.023\star$	$0.123 \pm 0.018\star$
			$(t = -0.024)$	$(t = 0.021)$
θ	$-3.628 \pm -0.079\star$	$-3.651 \pm 0.081\star$	$-5.279 \pm 0.173\star$	$-5.830 \pm 0.202\star$
	$(t = 0.011)$	$(t = 0.044)$	$(t = -0.018)$	$(t = 0.024)$
α_2	-	-	-	$0.084 \pm 0.027\star$
				$(t = -0.010)$
α_3	-	-	-	$-0.011 \pm 0.004\star$
				$(t = -0.000)$
κ_T	-	-	-	$0.824 \pm 0.046\star$
				$(t = -0.002)$

Table 5

Bonus Component

Structural and attribute parameter estimates using various attributed p^* models. These estimates are quoated as "Mean ± Std. dev.". Estimates with "$\star$" symbols are estimates whose confidence intervals (Mean ± 2 × Std. dev.) are bounded away from 0.000 and the corresponding network structure has t-statistic of less than 0.1.

	[pSTO-1] **Model St-1 +pBLK**	[pSTO-2] **Model St-1 +pBLK+ ind**	[pSTO-3] **Model St-1 +pBLK+ln(MB)**	[pSTO-4] **Model St-3 +pBLK+ln(MV)**
α^-_{STO}	$-0.593 \pm 0.295\star$	$-0.604 \pm 0.296\star$	$-0.592 \pm 0.296\star$	-0.388 ± 0.236
	$(t = 0.020)$	$(t = 0.043)$	$(t = -0.026)$	$(t = 0.023)$
α^+_{STO}	$0.575 \pm 0.184\star$	$0.556 \pm 0.186\star$	$0.387 \pm 0.187\star$	0.221 ± 0.132
	$(t = 0.034)$	$(t = 0.063)$	$(t = -0.035)$	$(t = 0.012)$
$i^=$	-	0.329 ± 0.503	0.121 ± 0.091	0.116 ± 0.089
		$(t = -0.002)$	$(t = 0.000)$	$(t = 0.030)$
β^-_{MV}	-	-	$-0.082 \pm 0.038\star$	-0.045 ± 0.033
			$(t = -0.032)$	$(t = 0.005)$
β^+_{MV}	-	-	$0.259 \pm 0.022\star$	$0.118 \pm 0.017\star$
			$(t = -0.025)$	$(t = 0.035)$
θ	$-4.090 \pm 0.083\star$	$-4.405 \pm 0.506\star$	$-5.348 \pm 0.145\star$	$-5.824 \pm 0.186\star$
	$(t = -0.030)$	$(t = -0.004)$	$(t = -0.024)$	$(t = 0.020)$
α_2	-	-	-	$0.085 \pm 0.027\star$
				$(t = 0.030)$
α_3	-	-	-	$-0.011 \pm 0.004\star$
				$(t = 0.036)$
κ_T	-	-	-	$0.823 \pm 0.045\star$
				$(t = 0.026)$

Table 6

Stock Option Component

Structural and attribute parameter estimates using various attributed p^* models. These estimates are quoated as "Mean $\pm$ Std. dev.". Estimates with "$\star$" symbols are estimates whose confidence intervals (Mean $\pm$ 2 $\times$ Std. dev.) are bounded away from 0.000 and the corresponding network structure has t-statistic of less than 0.1.

	[pSTO-1] **Model St-1 +pBLK**	[pSTO-2] **Model St-1 +pBLK+ ind**	[pSTO-3] **Model St-1 +pBLK+ln(MB)**	[pSTO-4] **Model St-3 +pBLK+ln(MV)**
α^-_{STO}	$-0.540 \pm 0.183\star$	$-0.544 \pm 0.179\star$	$-0.639 \pm 0.178\star$	$-0.471 \pm 0.160\star$
	$(t = 0.002)$	$(t = 0.055)$	$(t = -0.003)$	$(t = -0.011)$
α^+_{STO}	0.071 ± 0.182	-0.069 ± 0.100	-0.106 ± 0.099	-0.044 ± 0.068
	$(t = -0.014)$	$(t = 0.039)$	$(t = 0.001)$	$(t = -0.005)$
$i^=$	-	-0.134 ± 0.092	-0.119 ± 0.090	0.113 ± 0.088
		$(t = 0.028)$	$(t = 0.009)$	$(t = -0.009)$
β^-_{MV}	-	-	$-0.079 \pm 0.037\star$	-0.043 ± 0.032
			$(t = -0.002)$	$(t = -0.079)$
β^+_{MV}	-	-	$0.270 \pm 0.022\star$	$0.128 \pm 0.018\star$
			$(t = -0.007)$	$(t = -0.008)$
θ	$-3.836 \pm 0.093\star$	$-3.859 \pm 0.094\star$	$-5.043 \pm 0.145\star$	$-5.661 \pm 0.195\star$
	$(t = 0.006)$	$(t = -0.059)$	$(t = -0.088)$	$(t = 0.012)$
α_2	-	-	-	$0.083 \pm 0.027\star$
				$(t = 0.022)$
α_3	-	-	-	$-0.011 \pm 0.004\star$
				$(t = 0.026)$
κ_T	-	-	-	$0.822 \pm 0.046\star$
				$(t = 0.030)$

Table 7

Restricted Stock Component

Structural and attribute parameter estimates using various attributed p^* models. These estimates are quoated as "Mean $\pm$ Std. dev.". Estimates with "$\star$" symbols are estimates whose confidence intervals (Mean $\pm$ 2 $\times$ Std. dev.) are bounded away from 0.000 and the corresponding network structure has t-statistic of less than 0.1.

	[pSTO-1] **Model St-1 +pBLK**	[pSTO-2] **Model St-1 +pBLK+ ind**	[pSTO-3] **Model St-1 +pBLK+ln(MB)**	[pSTO-4] **Model St-3 +pBLK+ln(MV)**
α^{-}_{STO}	$0.540 \pm 0.250\star$	$0.067 \pm 0.252\star$	$0.007 \pm 0.250\star$	$-0.048 \pm 0.016\star$
	$(t = -0.019)$	$(t = -0.015)$	$(t = 0.013)$	$(t = 0.087)$
α^{+}_{STO}	-0.010 ± 0.182	-0.019 ± 0.100	-0.072 ± 0.181	-0.043 ± 0.067
	$(t = -0.004)$	$(t = 0.021)$	$(t = 0.016)$	$(t = 0.054)$
$i^{=}$	-	0.130 ± 0.092	0.119 ± 0.092	0.112 ± 0.088
		$(t = 0.028)$	$(t = 0.030)$	$(t = 0.058)$
β^{-}_{MV}	-	-	$-0.081 \pm 0.037\star$	-0.040 ± 0.032
			$(t = -0.061)$	$(t = 0.052)$
β^{+}_{MV}	-	-	$0.262 \pm 0.022\star$	$0.127 \pm 0.018\star$
			$(t = -0.038)$	$(t = 0.073)$
θ	$-3.938 \pm -0.051\star$	$-3.963 \pm 0.054\star$	$-5.269 \pm 0.136\star$	$-5.669 \pm 0.191\star$
	$(t = -0.022)$	$(t = 0.025)$	$(t = -0.042)$	$(t = 0.068)$
α_2	-	-	-	$0.083 \pm 0.027\star$
				$(t = 0.067)$
α_3	-	-	-	$-0.011 \pm 0.004\star$
				$(t = 0.062)$
κ_T	-	-	-	$0.823 \pm 0.046\star$
				$(t = 0.0070)$