# MEASURING MAIL: NEW ANALYSES OF E-MAIL DATA FOR THE STUDY OF CROSS-DIVISIONAL INNOVATION

ADAM M. KLEINBAUM Harvard Business School 328 Morgan Hall Soldiers Field Road Boston, MA 02163 akleinbaum@hbs.edu

## INTRODUCTION

Most large firms fail to take advantage of the opportunity to create new businesses that combine resources from disparate parts of the firm. In seeking to understand the sources of this failure, organizational scholarship has focused primarily in two areas of inquiry: formal organizational structure (e.g., Galbraith, 1973, 1994) and programs of incentives (e.g., Kaplan & Henderson, 2005; Prendergast, 1999). Both of these areas of study are well-developed, mature fields (Roberts, 2004), but even together, they offer only a limited understanding of the challenge of cross-divisional innovation. Recently, scholars have begun to call for systematic research into the role of intra-organizational social networks (e.g., Galbraith, 2006; Kleinbaum & Tushman, 2005) to complement the research on formal structure and incentives, but little such research has been done to date.

In this paper, I highlight two methodological challenges to the use of social network analysis to understand cross-divisional innovation. First, the kind and quality of data that have typically been collected to conduct network analysis are inadequate. Most studies rely on network data collected using questionnaires (Marsden, 1990, 2005), but such data is most likely to be both available and accurate when collected from relatively small groups or organizations; in contrast, in the large organizations engaged in multiple lines of business that provide the context for cross-divisional innovations, it is nearly impossible to collect complete, unbiased data using questionnaires. The first major goal of this paper is to argue for one solution to this problem: conducting network analysis on the basis of data drawn from electronic communication archives, including e-mail, instant messaging and electronic calendars. Second, there is a paucity of research that accounts for the embeddedness of the informal structure in the formal. The traditional organization design literature focuses on the formal while the social network analysis literature focuses on the informal, but rare has been the attempt to join the two. And yet, the creation of cross-divisional innovation requires both informal structure, which facilitates the flow of information across boundaries to uncover opportunities to collaborate, and formal structure, needed to officially recognize and implement cross-divisional innovation. To date, the literature has not explicitly developed analyses that combine formal and informal structure. Thus, the second major goal of this paper is to develop measures of several important inter-unit constructs - network density, tie strength, and brokerage - based on individual-level data.

# THE KIND AND QUALITY OF DATA IN SOCIAL NETWORK ANALYSIS

Research on electronic communication in organizations originated largely with communications scholars, and not surprisingly, focused primarily on understanding

communication patterns in organizations *per se*. The earliest studies that used e-mail data examined the effect of the introduction of a new medium of communication on underlying communication patterns (e.g., Eveland & Bikson, 1986). Sproull and Kiesler (1986) found that different media are conducive to the exchange of different information, rather than just the redistribution of the same information. Rice summarizes this stream of research: "Potential changes in communication may reinforce traditional network patterns but may also foster new kinds of interaction, data and processes, perhaps institutionalizing new organizational structures or changing the nature of interpersonal relations," (Rice, 1994b: 168).

In the basic model of organizational communication, communication medium is assumed to affect the flow of information in an organization. But the choice of medium is endogenous and some scholars have taken a contingency perspective to examine what factors determine which medium an actor is likely to choose to convey a particular message. This stream of research, known as media richness theory, has argued and found empirical support that organizational effectiveness results from congruence between the information processing requirements of a message and the choice of media (Daft, Lengel, & Trevino, 1987). Empirical evidence has generally supported the theory that richer media, such as face-to-face communication, are more suitable for subtle or sensitive messages because they better enable the exchange of social, as well as substantive, information (Finholt, Sproull, & Kiesler, 1990; Hinds & Kiesler, 1995; McKenney, Zack, & Doherty, 1992).

#### The Argument for Network Analysis based on Electronic Communication Data

We begin with the assumption that the network of communication is a valid measure of the social structure of an organization (Ahuja, Galletta, & Carley, 2003; Orlikowski & Yates, 1994; Rogers & Kincaid, 1981; Wasserman & Faust, 1994) and ask whether the e-mail network is a valid, unbiased measure of the intraorganizational communication structure. In addition to its substantive importance to communications scholars, media richness theory also provides a way to begin to answer this important question. Embedded in the theory is the important assumption that organizational actors systematically choose to use different media for different types of communications. If the communication networks within organizations differ according to the medium of communication - and specifically, if the e-mail network is not representative of the overall communications network – empirical findings based on studies of the e-mail network will be of limited and questionable value. In this section, I present three arguments in support of the use of e-mail data as a proxy for communication networks in substantive organizational network analysis. First, I argue that the extant literature supports, albeit equivocally, the use of e-mail data for network analysis. Second, the phenomenon of substantive interest is particularly conducive to using e-mail data. And third, data from electronic archives, while admittedly not perfect, offer certain advantages relative to other methods of collecting network data that makes them particularly valuable.

Conceptually, Monge and Contractor (2003) argue that the formal and informal networks of communication and authority in organizations are increasingly converging. They attribute this convergence to shifts in organizational structure and management philosophy from bureaucratic hierarchy (Weber, 1990 [1924]) toward lateral organization (Galbraith, 1994), enabled by the increasing ubiquity of information technology. As technology, and not hierarchy, increasingly mediates information flow, they suggest that network analysis of e-mail data is a reasonable proxy for intraorganizational social structure. In short, empirical research that explicitly

compares the structures of communication networks from different media suggests, albeit somewhat equivocally, that network data from e-mail communication is a valid measure of the social structure of an organization (e.g., Feldman, 1987; Rice, 1994a).

Second, the phenomenological and empirical contexts of the proposed study are particularly conducive to using e-mail data. Cross-divisional collaboration is a phenomenon that occurs in large, multi-divisional firms; questionnaires are poorly suited to provide data in such a context. First, a complete questionnaire data set requires an extremely high response rate; participation is easier to enforce and high response rates easier to achieve in small groups or organizations. Additionally, in the study of cross-divisional collaboration, ties that span divisional boundaries will be particularly important to capture accurately. Because an individual's strongest, most salient ties are likely to exist within her own division (Granovetter, 1973; Krackhardt & Stern, 1988; Nelson, 1989), it will be particularly difficult to accurately measure cross-divisional ties. Thus, the study of cross-divisional innovation is an area in which electronic communication data is particularly advantageous over survey-based approaches.

Finally, I appeal to the age-old, but oft-ignored, claim that network analysis has suffered from "an underemphasis upon data-gathering and measurement" (Rogers, 1987: 285). One benefit of e-mail data is that its strengths coincide with the weaknesses of the more traditional form of data collection, the network survey. And as many scholars have pointed about – about both surveys in general (Fowler, 2002) and network surveys in particular (Rogers, 1987) – survey data is subject to numerous sources of error and bias. Social networks constructed from electronic communication will avoid all of these problems (but will, admittedly, be subject to others). Additionally, the size of networks that can be studied using e-mail data is much larger than networks that can be studied using traditional survey-based data collection approaches.

# MEASURES OF INTER-UNIT SOCIAL STRUCTURE FROM INDIVIDUAL-LEVEL DATA

Data collection is not the only challenge for researchers attempting to explain the role of social structure in enabling cross-divisional innovation; more appropriate analytical measures are needed as well. The challenge is one of bringing together formal and informal structure and, in doing so, joining the inter-personal with the inter-unit levels of analysis. By definition, informal structure is the domain of social network analysis, and since the field of organizational network analysis was born within sociology, network analysts have sought to make sense of informal structure. But in large organizational settings, the informal structure is embedded within the formal; formal organizational structure has been shown empirically to both seed and constrain the formation of social ties by organizational members (Han, 1996; Henderson & Clark, 1990; Ibarra, 1995). As a result, the social networks in many organizations are strongly correlated with the formal organizational structure (Krackhardt et al., 1988), a phenomenon which many researchers have noted in passing (e.g., Burt, 2004; Nelson, 1989). In spite of these observations, few researchers have systematically examined the interplay between formal and informal structure; existing network measures generally do not accommodate the role of formal structure and this fact presents a serious obstacle to future research. In this section, I develop and empirically test a set of measures for network analysis that simultaneously account for formal organizational structure and informal social structure.

Much of the foundational social networks research neglects the role of formal structure altogether, instead conceptualizing properties of an individual's structure within a monolithic

social structure (e.g., Bonacich, 1987; Burt, 1992, 2000; Freeman, 1979). Since then, although there have been numerous adaptations and extensions of the original measures there are still relatively few network measures that bring insight to the question of how divisions – and the individuals that comprise them – relate to one another to jointly develop innovation (cf. Doreian, Batagelj, & Ferligoj, 2005; Everett & Borgatti, 2005).

In the remainder of this section, I develop empirical approaches to measure the density of inter-unit ties, their average strength and the existence of inter-unit brokerage ties. To date, a large number of constructs have been proposed and used to measure relations in a social system at the inter-personal level; the over-arching goal of this section is to bring in the formal structure of the organization and, using formal structural arrangements, to partition well-established social structural constructs into intra-unit and inter-unit components. As such, the measures defined in this section will be derived from existing, well-established network measures.

#### Data

The present study provides a "pilot test" of the analyses using data drawn from a publicly available data set on the network of global trade in miscellaneous manufactures of metal in 1994 (de Nooy, Mrvar, & Batagelj, 2005). This data, admittedly not ideal, features several important similarities with the e-mail data. First, as in e-mail communication, ties in the world trade data set are directed and valued. Second, each nation is a member of exactly one continent, just as each person is a member of one division of a firm. Third, trade is, to some degree, concentrated within continents, just as e-mail communication is concentrated within organizational units, but substantial inter-continental trade also occurs. All measures were implemented in MATLAB (MathWorks Inc., 2005); tables and figures, visualized using Pajek (Batagelj & Mrvar, 2006), are available from the author.

In order to facilitate the development of these measures, I begin by defining some terminology. In the context of an organization that is formally structurally differentiated, the firm may consist of many divisions and the unit of analysis for these measures is a dyad of focal divisions; let us call them A and B. Let us further suppose that unit A consists of *m* individuals (indexed by *i*) and that unit B consists of *n* individuals (indexed by *j*). Let  $x_{ij}$  be a binary variable equal to 0 if there is no relation from actor *i* to actor *j* and equal to 1 if there is a relation from *i* to *j*.  $z_{ij}$  is a continuous measure of the strength of the directed tie from actor *i* to actor *j*.

#### **Inter-Unit Tie Density**

Perhaps the most commonly used way to measure the number of ties in a social system is a density measure (Marsden, 2005). When considering social interactions in a dyad of divisions, it makes sense to partition the traditional measure of network density into several components: the intra-unit network density for each unit (calculated trivially using the traditional density formula on the appropriate subset of the data) and the inter-unit network density: the number of existing ties from individuals in unit A to individuals in unit B as a proportion of the total such number of potential ties. Formally, we define the inter-unit network density of A to B as:

$$IUND_{AB} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij}}{m \times n}$$
(1)

This measure provides substantial insight that is not available from previous measures. By formally identifying the different divisions, we are able to both tease apart intra-unit ties from inter-unit ties and look explicitly at the direction of ties from one division to the other. To illustrate the benefits of this measure, consider the empirical case of international trade in Africa and North America. Relatively few ties join Africa to North America and all of them flow from the United States to various African countries. The traditional density measure cannot quantify the effect, but the proposed inter-unit density measure offers substantially more insight. In addition to replicating the results of the traditional measure, it also informs us that the density of ties flowing from North America to Africa is extremely low (0.0864) – lower even than the tie density within Africa. Furthermore, the density of ties flowing from Africa to North America is 0. By recognizing the formal affiliations of units within divisions, we are able to develop a more general measure that replicates the results of the traditional measure, but also provides greater insight into the structure of network ties among actors and the divisions to which they belong.

## **Average Inter-Unit Tie Strength**

Tie strength has been associated with such important outcomes as knowledge transfer (Hansen, 1999) and cooperative problem-solving (Uzzi, 1997). By bringing formal structure into the picture, we can begin to imagine some uses for an aggregate measure of the strength of the interpersonal ties that link two organizational divisions. For example, we might expect that inter-divisional ties will tend to be stronger if the divisions have collaborated on product development projects in the past or if they frequently exchange employees; we might further expect that these stronger ties might help them to collaborate together again in the future. To fill this gap in the literature, I propose a formal measure of average inter-unit tie strength:

$$AIUTS_{AB} = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} z_{ij}}{\sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij}}$$

(2)

The development of this measure will provide important new insights into the nature of cross-divisional communication. At the individual level, strong ties have been shown to have a variety of consequences, both positive and negative, but research to date has been unable to measure tie strength at the aggregate, inter-unit level. Instead, scholars have either focused at the individual level of analysis (e.g., Burt, 1992), found clever proxies for inter-unit tie strength (e.g., Gulati, 1995), or assumed inter-unit ties to be a unit-level phenomenon (e.g., Hansen, 1999). Clearly, none of these approaches to measuring inter-unit tie strength is ideal. The present measure of average inter-unit tie strength will enable future research to tackle problems of inter-unit social structure more fully, accurately, and directly.

#### **Inter-Unit Brokerage**

The literature on structural holes and their inverse, constraint, dates back to Burt's (1980; 1982; 1992) seminal work, in which he developed a measure of aggregate constraint. Since then, Burt's notion of constraint has inspired numerous conceptual (e.g., Gargiulo, 1993) and methodological (e.g., Casciaro & Piskorski, 2005) advances. However, to my knowledge, all such applications measure constraint with respect to the entire social system. Partitioning out

constraint in the communication network according to division can offer great insight into where in the organization cross-divisional innovations are likely to arise.

The proposed measure builds upon the concepts and terminology in Burt's (1992) aggregate constraint measure, but instead of summing across all the individuals j who constrain i's autonomy, we sum according to the unit affiliations of both:

$$C_{AB} = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij}$$
(3)

Having defined aggregate inter-unit constraint ( $C_{AB}$ ), we turn to the empirical data and make several observations. Africa does not constrain North or South America or Oceana at all; not only do African nations not export any manufactured goods to those countries directly, African nations export manufactured goods to only two countries (Israel and Jordan), both of whom do not export goods to North or South America or Oceana. If results such as these were obtained from intra-organizational communications data, we would learn that there are no brokers who can provide North or South America or Oceana with information about the capabilities of Africa. This implies that information that originates within Africa reaches almost no other divisions; without information about what capabilities reside within Africa, it would be exceedingly unlikely that other divisions would seek to collaborate with Africa.

## CONCLUSION

Large, multidivisional firms face a critical growth imperative. In order to meet the demands of Wall Street analysts, the business press, employees and shareholders, they must continually find new ways to grow their businesses. Some firms have attempted to grow by developing new businesses that combine resources from different divisions, with varying degrees of success. Yet organization theory offers us little insight regarding why some firms attempt cross-divisional innovation whereas others do not and even less insight regarding why so few such attempts are met with success. The premise of this paper is that the extant literature – mostly in the areas of formal organizational structure and incentives – can be meaningfully complemented by a social network perspective, but that network analysis into this phenomenon has been hindered by a lack of sufficiently informative data and by an absence of appropriate measures. In order for the substantive literature to continue to progress, a set of new measures based on a new source of data is needed. In this paper, I have argued for a set of novel measures that take individual-level data from the exchange of electronic communications within the organization and aggregate it to develop meaningful insights about the social structure of inter-unit relations.

#### **REFERENCES AVAILABLE FROM THE AUTHOR**