

Measuring Collaboration in Modern Organizations[†]

By STEPHEN MICHAEL IMPINK, ANDREA PRAT, AND RAFFAELLA SADUN*

Economic theory has long postulated that internal communication plays a central role in firms. Simon (1947) devotes a whole chapter of *Administrative Behavior* to internal communication. The main goal of an organization is to set up processes and structures to overcome the cognitive limits of individuals. As communication bandwidth is one of our most obvious limits, the organization shapes communication channels as a function of its priorities. Simon also highlights the coexistence of formal and informal communication channels and the role that internal communication plays in building organizational memory. The point about the endogeneity of internal communication is echoed with great clarity in Arrow (1974): “An organization can acquire more information than any one individual, for it can have each member perform different experiments. Thus, the limitations on an individual’s capacity are overcome. But as always there is a price to be paid. . . . The information has to be coordinated if it is to be of any use to the organization. Formally stated, communication channels have to be created within the organization.” Arrow sees internal communication as a key factor in the *agenda* of an organization: decisions require communication, and communication leads to decision. Arrow hypothesizes that organizations will also create specialized *codes* to communicate effectively among employees.

Theories such as Arrow’s and Simon’s yield a powerful testable implication: internal communication patterns reflect the agenda of the organization. A change in the agenda should

determine a change in communication, and conversely, we should be able to infer a lot about an organization if we can observe its internal communication patterns.

For decades, empirical organizational economists were aware of these predictions but unable to test them with large-scale observational studies because of the lack of data. However, over the past decade, researchers with access to electronic communications data have started to quantify the internal structural organization and collaboration of firms. These data have the potential to provide new insights on the wide and persistent heterogeneity in firm performance documented in the prior literature (Syverson 2011). Similar to practices and routines (Bloom et al. 2012; Bloom, Sadun, and Van Reenen 2016), communications provide insight into some of the “softer” and less quantifiable aspects of the firm, such as its underlying culture and managers’ behavior, and how these aspects affect firm performance (Gibbons and Henderson 2012). As electronic communications data are now often captured by many firms and email providers, researchers are able to measure the dynamics of internal communication interactions reliably and at scale, both within and across firms.

In this paper, we first briefly discuss the potential value of data on employee communications and interactions within firms in the context of prior literature in organizational economics. We then discuss the potential benefits and risks associated with the use of electronic communications data (email and meetings) as empirical proxies for these internal patterns of virtual and physical interactions, as well as the way in which prior research has employed electronic communications and meeting data within individual organizations. Finally, we discuss the research possibilities opened by the emergence of comparable communications data across larger sets of firms and some concrete examples of proxies for the intensity and type of internal communications that can be built by using these data.

* Impink: Stern School of Business, New York University (email: simpink@stern.nyu.edu); Prat: Columbia Business School, Columbia University (email: andrea.prat@columbia.edu); Sadun: Harvard Business School, Harvard University (email: rsadun@harvard.edu).

[†] Go to <https://doi.org/10.1257/pandp.20201068> to visit the article page for additional materials and author disclosure statement(s).

I. Why Measure Internal Communications?

Internal communication has been a central theme in organizational economics. The seminal contributions of Simon (1947) and Arrow (1974) discussed above have inspired a number of formal models. Dessein and Santos (2006) studies the trade-off between decentralized decision-making, which allows units within a firm to adapt to local circumstances, and centralized coordination, which ensures that the firm realizes synergies between units. The trade-off is modified by investments in internal communication. An improvement in communication technology leads to both better adaptation and better coordination. A strand of literature initiated by Radner (1993) and Bolton and Dewatripont (1994) models firms as hierarchies of processors that communicate with one another and engage in parallel processing. Crémer, Garicano, and Prat (2007) formalizes Arrow's (1974) idea of *coding*: firms develop optimal language tailored to the set of problems they face. A "narrow" language achieves great communication efficiency within a small set of problems, while a "broader" language allows more units to communicate with one another, albeit in a less efficient manner.

Organizational economists have combined tools from network economics and team theory to model some of Arrow's ideas about the agenda of organizations. Calvó-Armengol, De Martí, and Prat (2015) studies an organization where agents are engaged in bilateral communication. Communication is costly, with each agent investing in active communication (speaking/writing) and passive communication (listening/reading). The main result is a characterization of communication and influence patterns among agents. Dessein, Galeotti, and Santos (2016) models the firm as a network of rationally inattentive agents whose communication cost is proportional to the entropy reduction generated by the signal, as in Sims (2003). Optimal networks focus attention on a small set of tasks and individuals, thus providing a microfoundation to a central idea in management that firms should focus on core competencies.

Communication patterns—both inside and across organizations—have also been a topic of interest in the network literature. This literature provides a broader motivation on the value of measuring and studying human interactions

relative to the organizational economics literature. "Humans are a fundamentally social species" (Jackson 2010), and as such, better understanding their interactions can provide insight into economic behavior and firm performance. For example, several intrafirm studies have focused on how access to and control over information flow can lead to beneficial outcomes for individuals or groups in certain network positions (Burt 2004). Being more central in a network enables increased access to information, and the act of brokerage, defined as bridging two disparate groups, increases access to more novel information (Granovetter 1973). Others argue that there is value to redundant information paths (Uzzi and Spiro 2005, Obstfeld 2005). Using a mix of social network and electronic communications data, Aral and Van Alstyne (2011) argues that the value of brokerage is dependent on the firm's underlying information environment. In certain environments, brokerage may not lead to novel information that can enhance individual creativity and generate "good ideas."

II. The Value of Electronic Communications Data

Social relations of employees are no less structurally embedded (Granovetter 1985, Zukin and DiMaggio 1990) than ever before, but new methods are needed to adequately capture these interactions in modern organizations.

Early studies of collaboration rely heavily on surveys to assemble how individuals (i.e., "nodes") map to one another (Borgatti, Everett, and Johnson 2018) in a network. The collection of survey-based network data, however, presents several methodological challenges. First, network surveys do not easily scale across large organizations or communities and, due to high costs and the need for administrator approval, are usually able to capture data only within a single entity or across several smaller entities (Banerjee et al. 2019). Second, surveys are subject to the cognitive biases of respondents and may lead to an underrepresentation of reported connections that are less salient, such as connections with individuals who work remotely.

Alternative proxies of social proximity, such as geographically based measures (i.e., distances between offices or desks), can potentially overcome the limitations of survey-based measures (Battiston, Blanes i Vidal, and Kirchmaier

2017). However, even these metrics may be progressively undermined by technological advancements in communication methods that enable workers to collaborate with others in different locations (and, in some instances, even to work remotely from home (Bloom, Kretschmer, and Van Reenen 2009)) and by the increasing interoffice mobility allowing for open floor plan seating and unassigned “hot desks.” Overall, these trends can potentially reduce the value of individuals’ office locations as a proxy for underlying communication patterns.

Electronic communications data have the potential to overcome the measurement challenges outlined above. Knowledge workers rely on email, meetings, and, in many cases, instant messenger software to collaborate with others. As a consequence, electronic communications are able to capture valuable (though certainly imperfect, as we discuss below) information on the underlying interactions among employees within a team, across functional groups (internally), and with partners and customers (externally).

These data present several advantages relative to survey-based measures of recalled interactions or office maps. First, the data are collected passively (i.e., email exchange and calendar data are typically saved in firms’ servers), thus overcoming the costs and risks associated with survey-based measures necessitating active responses from employees. Second, the data have the potential to reach near-population samples for knowledge workers regardless of their physical location, thus capturing both physical and remote interactions, which typically elude survey measures. Third, firms typically retain logs of emails and calendar information over long periods of time for compliance reasons, which potentially enables the creation of longitudinally comparable datasets.

The intrinsic value of employees’ electronic communication can be gleaned from the increasing number of startups and high-technology firms dedicated to the analysis and interpretation of these communications. Early entrants, such as VoloMetrix, created a business model around analyzing anonymous communications data from email, meetings, social platforms, and line-of-business applications. Before being purchased by Microsoft Corporation, VoloMetrix patented measures of organizations’ performance, including measures of organizational

load, organizational fragmentation, and network efficiency. Firms can now license software that enables them to analyze their employees’ electronic communications to better understand collaboration patterns, identify high-performance employees, and prevent certain work groups from being overloaded (Nielsen 2016).

The potential value of these types of data has also been recognized by researchers across numerous academic disciplines outside economics. Kleinbaum and coauthors (Kleinbaum and Tushman 2008, Kleinbaum 2012, Kleinbaum and Stuart 2014), for example, use email data on a large global firm to model relations among employees and to study resulting employee and group-level outcomes within that firm. Other studies use publicly available email data from Enron to analyze firm communication in times of organizational crisis (Diesner, Frantz, and Carley 2005) and social network identification (McCallum, Wang, and Corrada-Emmanuel 2007).

Electronic communications data, however, also present considerable limitations. First, when available, these data may miss a considerable fraction of actual emails and interactions. For instance, employees could use their calendar inconsistently or may not actually attend meetings that appear in their calendars. Equivalently, employees who deal with highly sensitive information may opt for an ad hoc phone call or in-person meeting to avoid sharing the information in a legally discoverable manner. Second, even more importantly, having access to these data for research purposes has traditionally been extremely difficult due to privacy considerations, ranging from the worry that emails and meeting logs may provide strategic or legally relevant information to the concern that the mere act of studying the data may negatively affect employee morale by fostering a sense of being “micromanaged” or “spied on” by the company. These obstacles to access have created a situation in which the electronic communications data that have been made available to researchers typically cover only very small sets of firms at a time (in most cases only one) and have often been made available under very unique circumstances (e.g., preexisting relationships with management). As a consequence, it has so far been impossible to build consistent and comparable metrics of email and meeting usage across many firms, and it is unclear

whether the interesting findings emerging from existing studies generalize to organizations different from the ones that self-selected into the sharing agreement.

III. Recent Developments

As individual firms realize the value of their electronic communications data, high-technology companies have started to build software products that analyze email and meeting data at scale in a manner that preserves the employees' rights to privacy—for example, deidentifying employee-level information. Additionally, the consolidation of email providers has created the preconditions to build standardized and comparable firm-level collaboration measures across large sets of organizations, thus overcoming the selection concerns outlined above. Taken together, these trends provide unprecedented opportunities for researchers to partner with email providers to generate new insights on internal collaborations or organizational structure across an array of firms. For example, Polzer, DeFilippis, and Tobio (2018) derives standardized measures of collaboration from electronic communications data (email and meetings) on a large sample of firms obtained from an email provider to study the variation in communication patterns across countries.

In our own work in partnership with an email provider, we have started investigating how different variables based on firm-wide email and meeting data across a large sample of firms can be used to build proxies for internal communication patterns. We outline some of the variables that we are currently exploring in this ongoing work below.

Intensity.—For most knowledge workers, intensity of time and effort spent in interactions can be measured through the number of emails sent, emails read, hours logged into one's computer, or number of meetings planned and/or attended. These crude intensity measures can be augmented and refined in several ways. For example, the intensity of the effort expended in any given interaction (virtual or personal) is likely to vary according to the number of people involved in the interaction (email recipients or attendees): an all-hands meeting with 50 colleagues may require less attention (one may even decide to multitask and reply to emails)

than a one-on-one meeting with a manager or close colleagues, and mass emails are less likely to be read than personalized communications sent to a working team or individual. In addition, the duration of meetings could provide insight into the nature of interaction. All-day or half-day meetings are more likely to be trainings or full-team events than time spent working with a single employee or smaller group.

Collaboration.—For a given intensity of an exchange, electronic communications and meeting data can reveal information on the nature of collaboration among employees. For example, the nature of a meeting can be inferred by the type of participants involved in the interactions (same or different departments, internal or external participants) or their hierarchical level in the organization. As an example, proxy measures could be built for the proportion of meetings or emails including employees from multiple functional groups, or for internal versus external interactions.

Aggregations.—In some cases, email and meeting data can be paired with salient aspects of a firm's organizational structure, such as the hierarchical level of employees, their tenure, and their location (e.g., at headquarters or in a subsidiary). An employee's management level could be derived from the firm's listing of formal reporting relationships, which is usually recorded when an email provider onboards a firm. This allows researchers to distinguish between employees and managers and, within managers, among individuals with different spans of control. These distinctions are valuable for studying the extent to which managerial status is associated with a different intensity or modality of interactions.

Longitudinal Variation.—Electronic communications data can be used to develop measures of intensity and collaboration for aggregations of employees longitudinally, at the daily, monthly, and (depending on storage capabilities) yearly level. As such, these data can be used within research designs that are aimed at estimating the effects of exogenous "shocks" on internal communication patterns—that is, simple before-and-after or difference-in-difference research designs. Augmenting the study of electronic communications data with more

sophisticated research designs would increase the value and credibility of studies exploiting this type of information for inference.

Textual Analysis.—Until recently, data could have only a numerical structure. However, advances in text processing have produced numerous powerful techniques that allow economists to treat text as data (Gentzkow, Kelly, and Taddy 2019). Within organizational economics, Hansen, McMahon, and Prat (2018) uses textual analysis to understand the effect of a change in the degree of outside scrutiny on the effectiveness of committee decision-making. The setting is the Federal Open Market Committee of the US Federal Reserve, and the policy change occurred in 1993, when it was suddenly announced that meeting transcripts would be published verbatim five years after each meeting. The career concerns literature identifies the possible effect of more intense scrutiny: better discipline and conformism. In both cases, the effect is predicted to be stronger for newer agents. The authors use a Bayesian topic modeling technique called latent Dirichlet allocation (Blei, Ng, and Jordan 2003) to transform members' statements into vectors of topic weights that can be used in a standard difference-in-difference analysis. Both effects are present, but the discipline effect seems to dominate. Keeping in mind the privacy concerns outlined above, a systematic analysis of the text of emails sent within one or multiple organizations has the potential to provide invaluable information on the "soft" aspects of company culture and values that are embodied in language and that have so far largely escaped economics research.

IV. Conclusion

Electronic communications data can be transformational for organizational economics. They enable the measurement and study of internal and external information exchange and collaboration patterns, and they have the potential to lead the way to the study of whether communication varies across firms and the extent to which this variation may be helpful in understanding the observed variation in performance across organizations. So far, the study of these data has been mainly confined to disciplines outside economics. The emergence of detailed and comparable data on electronic communications

and meetings across large sets of organizations affords economists the opportunity to advance the research program outlined in Simon (1947) and Arrow (1974) into a fully fledged empirical endeavor.

REFERENCES

- Aral, Sinan, and Marshall Van Alstyne.** 2011. "The Diversity-Bandwidth Trade-Off." *American Journal of Sociology* 117 (1): 90–171.
- Arrow, Kenneth J.** 1974. *The Limits of Organization*. New York: Norton.
- Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson.** 2019. "Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials." *Review of Economic Studies* 86 (6): 2453–90.
- Battiston, Diego, Jordi Blanes i Vidal, and Tom Kirchmaier.** 2017. "Is Distance Dead? Face-to-Face Communication and Productivity in Teams." CEPR Discussion Paper DP1473.
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan.** 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3: 993–1022.
- Bloom, Nicholas, Christos Genakos, Raffaella Sadun, and John Van Reenen.** 2012. "Management Practices across Firms and Countries." *Academy of Management Perspective* 26 (1): 12–33.
- Bloom, Nicholas, Raffaella Sadun, and John Van Reenen.** 2016. "Management as a Technology?" NBER Working Paper 22327.
- Bloom, Nick, Tobias Kretschmer, and John Van Reenen.** 2009. "Work-Life Balance, Management Practices and Productivity." In *International Differences in the Business Practices and Productivity of Firms*, edited by Richard B. Freeman and Kathryn L. Shaw, 15–54. Chicago: University of Chicago Press.
- Bolton, Patrick, and Mathias Dewatripont.** 1994. "The Firm as a Communication Network." *Quarterly Journal of Economics* 109 (4): 809–39.
- Borgatti, Stephen P., Martin G. Everett, and Jeffrey C. Johnson.** 2018. *Analyzing Social Networks*. London: Sage.
- Burt, Ronald S.** 2004. "Structural Holes and Good Ideas." *American Journal of Sociology* 110 (2): 349–99.

- Calvó-Armengol, Antoni, Joan de Martí, and Andrea Prat.** 2015. "Communication and Influence." *Theoretical Economics* 10 (2): 649–90.
- Crémer, Jacques, Luis Garicano, and Andrea Prat.** 2007. "Language and the Theory of the Firm." *Quarterly Journal of Economics* 122 (1): 373–407.
- Dessein, Wouter, and Tano Santos.** 2006. "Adaptive Organizations." *Journal of Political Economy* 114 (5): 956–95.
- Dessein, Wouter, Andrea Galeotti, and Tano Santos.** 2016. "Rational Inattention and Organizational Focus." *American Economic Review* 106 (6): 1522–36.
- Diesner, Jana, Terrill L. Frantz, and Kathleen M. Carley.** 2005. "Communication Networks from the Enron Email Corpus: 'It's Always about the People. Enron Is No Different.'" *Computational and Mathematical Organizational Theory* 11 (3): 201–28.
- Hansen, Stephen, Michael McMahon, and Andrea Prat.** 2018. "Transparency and Deliberation within the FOMC: A Computational Linguistics Approach." *Quarterly Journal of Economics* 133 (2): 801–70.
- Gentzkow, Matthew, Bryan Kelly, and Matt Taddy.** 2019. "Text as Data." *Journal of Economic Literature* 57 (3): 535–74.
- Gibbons, Robert, and Rebecca Henderson.** 2012. "Relational Contracts and Organizational Capabilities." *Organization Science* 23 (5): 1350–64.
- Granovetter, Mark.** 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78 (6): 1360–80.
- Granovetter, Mark.** 1985. "Economic Action and Social Structure: The Problem of Embeddedness." *American Journal of Sociology* 91 (3): 481–510.
- Jackson, Matthew O.** 2010. *Social and Economic Networks*. Princeton, NJ: Princeton University Press.
- Kleinbaum, Adam M.** 2012. "Organizational Misfits and the Origins of Brokerage in Intrafirm Networks." *Administrative Science Quarterly* 57 (3): 407–52.
- Kleinbaum, Adam M., and Michael L. Tushman.** 2008. "Managing Corporate Social Networks." *Harvard Business Review*, July–August 2008, 8–14. <https://hbr.org/2008/07/managing-corporate-social-networks>.
- Kleinbaum, Adam M., and Toby E. Stuart.** 2014. "Inside the Black Box of the Corporate Staff: Social Networks and the Implementation of Corporate Strategy." *Strategic Management Journal* 35 (1): 24–47.
- McCallum, Andrew, Xuerui Wang, and Andrés Corrada-Emmanuel.** 2007. "Topic and Role Discovery in Social Networks with Experiments on Enron and Academic Email." *Journal of Artificial Intelligence Research* 30: 249–72.
- Nielsen, Chantrelle.** 2016. "What Work Email Can Reveal about Performance and Potential." *Harvard Business Review*, February 10, 2016. <https://hbr.org/2016/02/what-work-email-can-reveal-about-performance-and-potential>.
- Obstfeld, David.** 2005. "Social Networks, the *Terminus Iungens* Orientation, and Involvement in Innovation." *Administrative Science Quarterly* 50 (1): 100–30.
- Polzer, Jeffrey T., Evan DeFilippis, and Kristina Tobio.** 2018. "Countries, Culture, and Collaboration." *Academy of Management Proceedings* 2018 (1): 17645.
- Radner, Roy.** 1993. "The Organization of Decentralized Information Processing." *Econometrica* 61 (5): 1109–46.
- Simon, Herbert A.** 1947. *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization*. New York: Macmillan.
- Sims, Christopher A.** 2003. "Implications of Rational Inattention." *Journal of Monetary Economics* 50 (3): 665–90.
- Syversen, Chad.** 2011. "What Determines Productivity?" *Journal of Economic Literature* 49 (2): 326–65.
- Uzzi, Brian, and Jarrett Spiro.** 2005. "Collaboration and Creativity: The Small World Problem." *American Journal of Sociology* 111 (2): 447–504.
- Zukin, Sharon, and Paul DiMaggio, eds.** 1990. *Structures of Capital: The Social Organization of the Economy*. Cambridge: Cambridge University Press.