Email communications in a large scientific research organization

Benjamin Sims
Nikolai Sinitsyn
Stephan Eidenbenz
Los Alamos National Laboratory

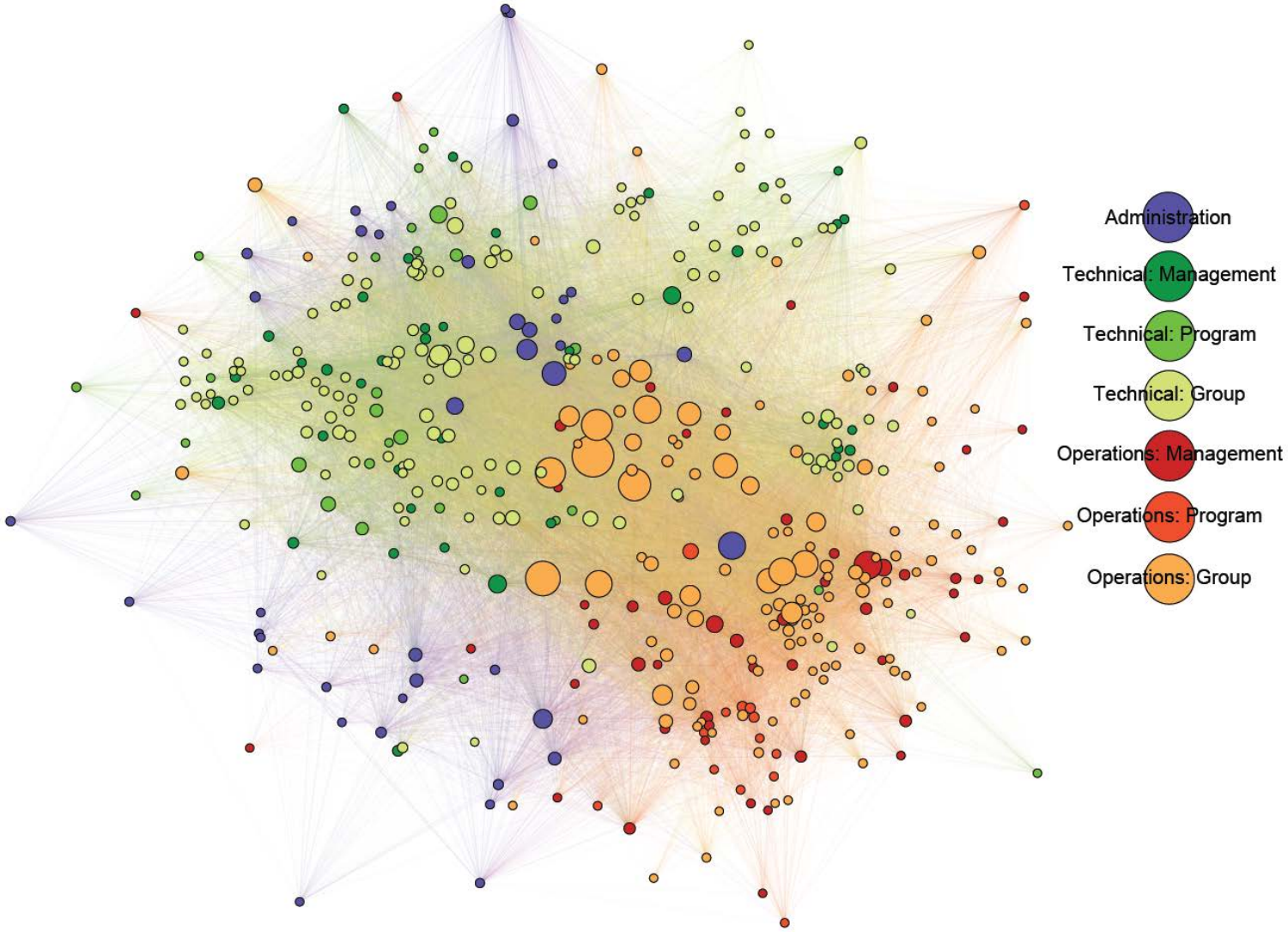
Abstract

We present findings from a large-scale study of email networks, interpreted as social networks, based on data from a large scientific research organization. We explore several methods for visualizing network structure at large and small scales, develop a model of how organizational hierarchy influences email communications, and examine temporal patterns in email traffic. Based on organization charts, we classify each organizational unit by function and organizational level. By visualizing and aggregating email traffic between organizational units using these categories, we derive several insights into how large subdivisions of the organization interact with each other and with outside organizations. In addition, we provide examples of how email data can be used to examine the internal structure of individual organizational units. We propose a power law model for predicting connectivity in organizations based on number of managers at each level and number of support staff associated with each manager, and compare it to actual email connectivity in the organization. Additional temporal analysis shows both diurnal and weekly cycles in email traffic within the organization.

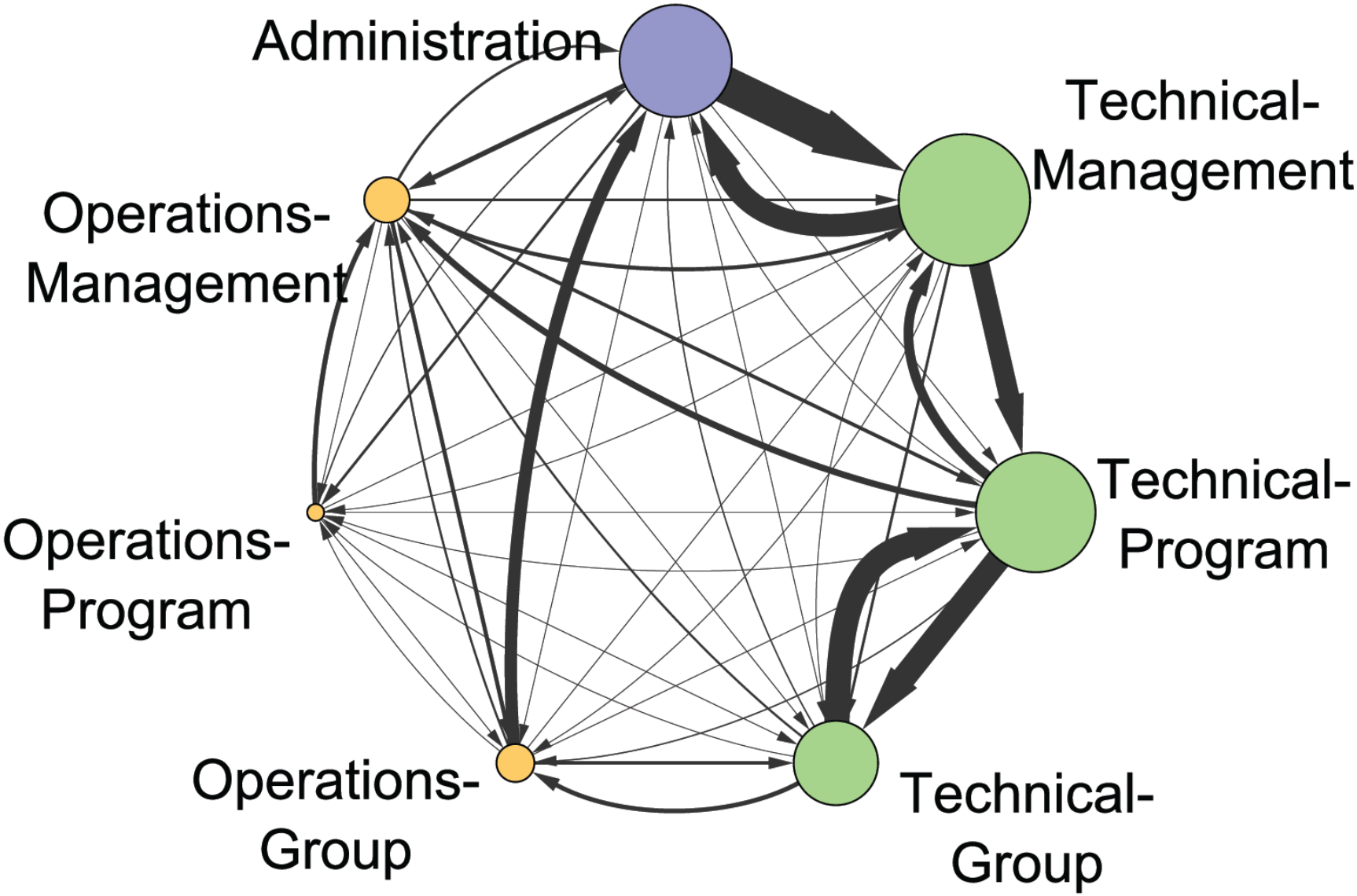
Overview

- Analyzed complete email traffic for Los Alamos National Laboratory over a one month period
- Used a variety of analytic tools to analyze data in relation to known organizational features
- Results demonstrate a variety of potentially useful approaches to understanding organizational communication patterns
- Useful for understanding organizational issues, as well as identification of potentially anomalous communications

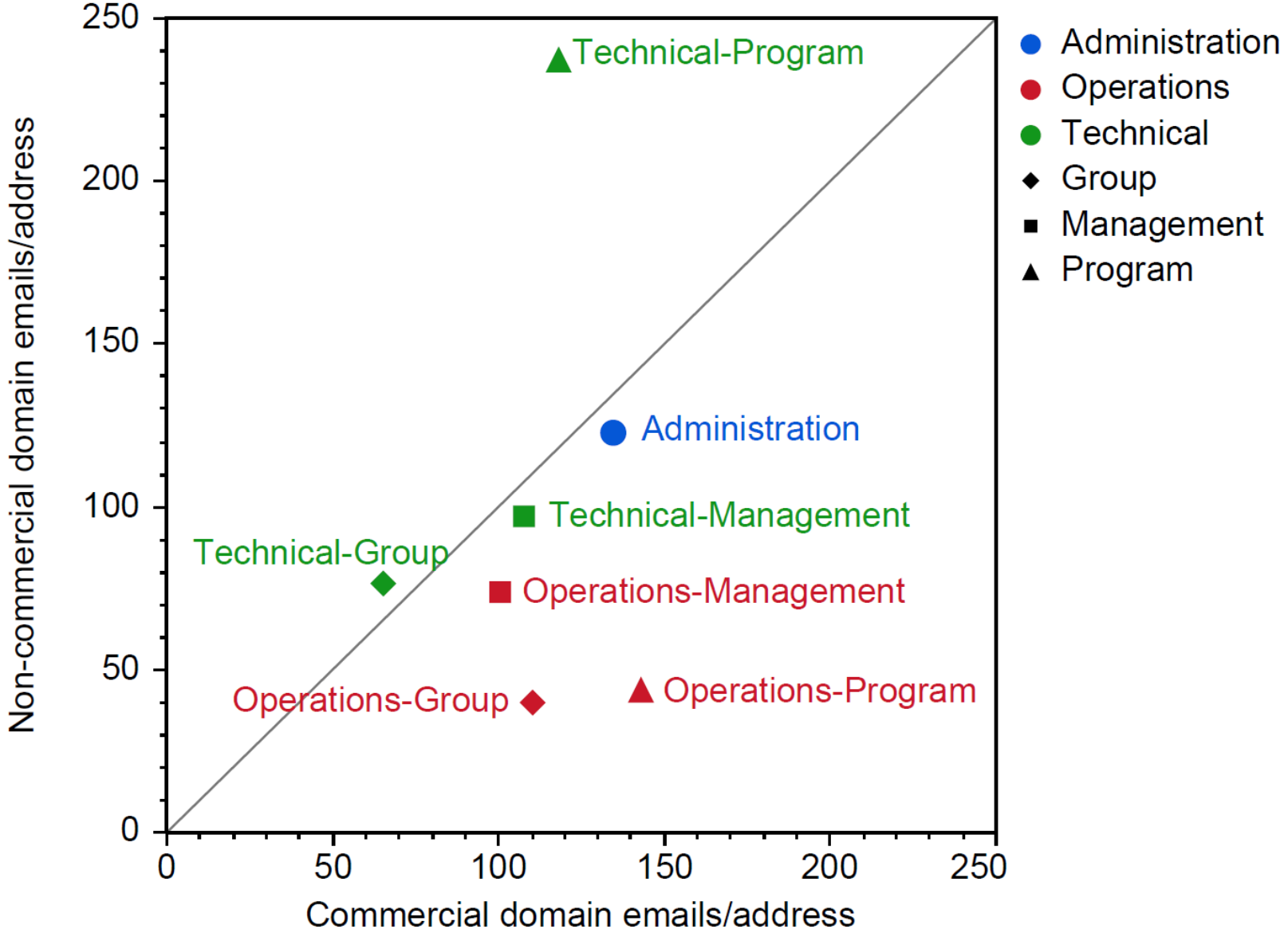
Email traffic between organizational units



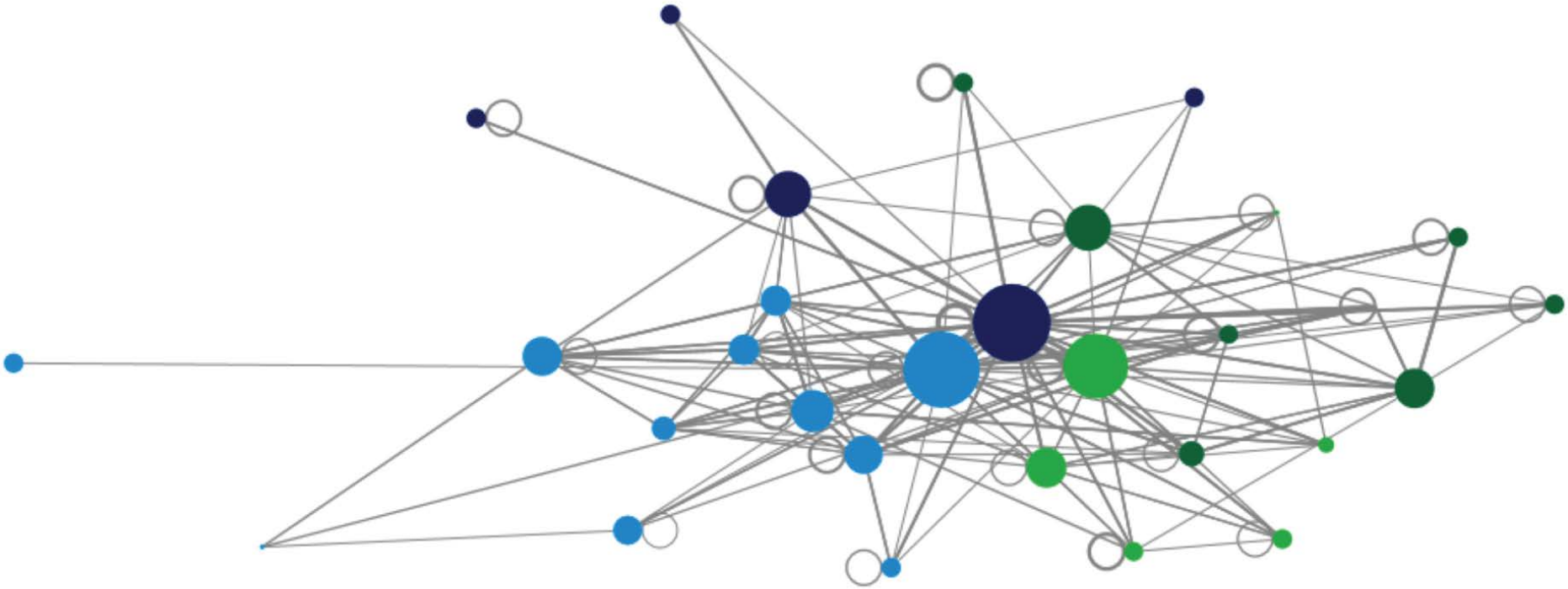
Email traffic between types of organizational units



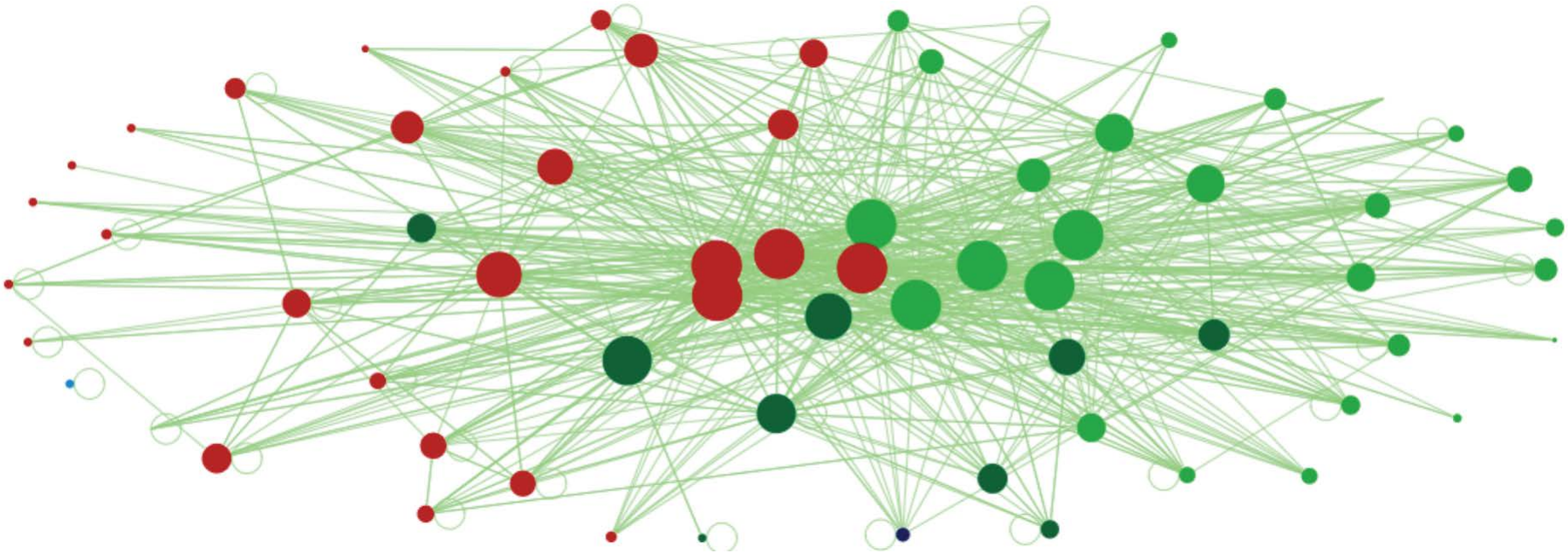
Email traffic with outside domains



Internal email traffic, small organizational unit



Internal email traffic, larger organizational unit



Node connectivity distribution as a function of organizational hierarchy

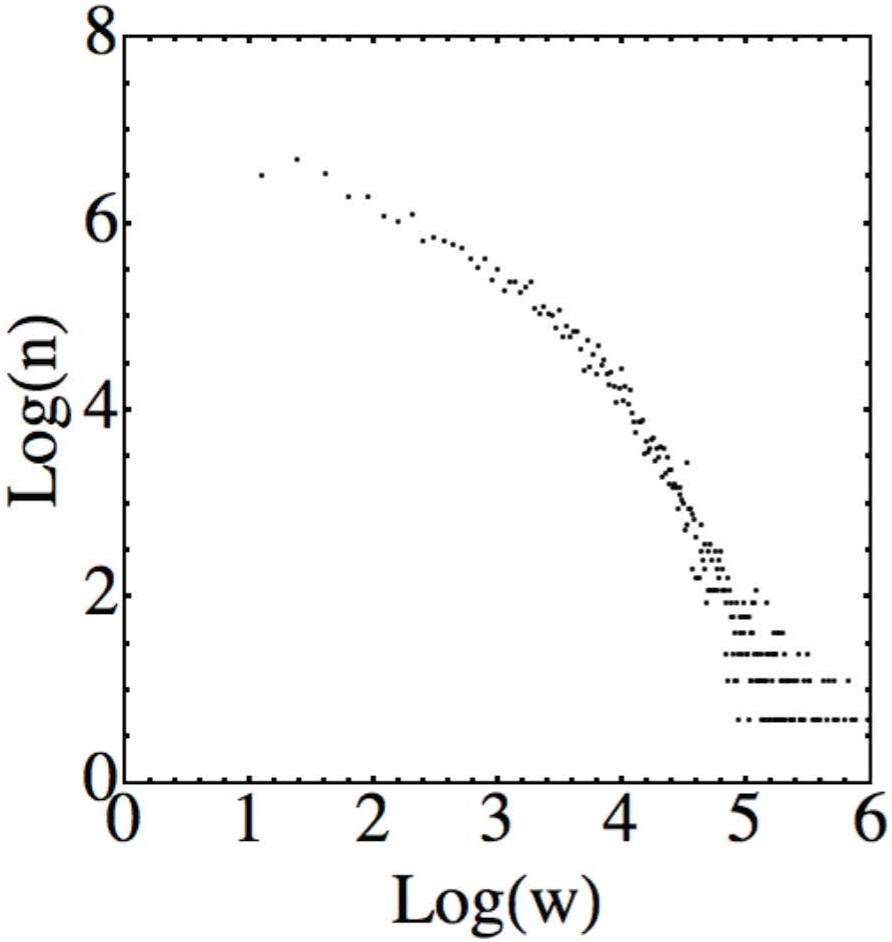
- Top manager has highest connectivity N
- l lower-level managers report to each manager, each with connectivity N/l
- a administrators assist each manager, also with connectivity N/l
- Therefore, at level x, $(al)^x$ managers (or their proxies) send emails to $N/(l^x)$ people under their supervision
- Consider a plot of number of nodes n vs. outdegree w. Excluding the variable x, we find:

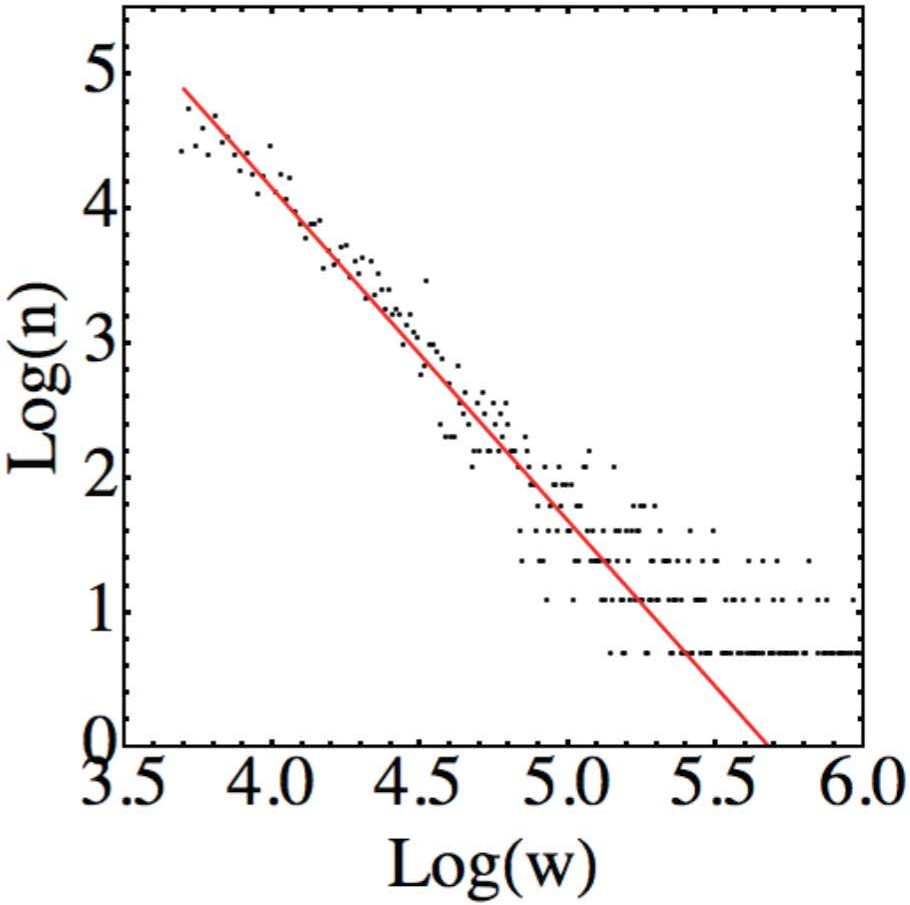
$$\log(n) = \frac{\log(N)}{\log(l)} - \frac{\log(al)}{\log(l)} \log(w)$$

Node connectivity distribution as a function of organizational hierarchy (cont.)

- Outdegree distribution is then, generally, a power law with exponent $\frac{\log(al)}{\log(l)} > 1$
- At some level *x,* the hierarchy terminates around the point where $(al)^x = \mathrm{N}/(l^x)$, so number of managers does not exceed number of employees
- Therefore, power law is expected to hold only for nodes with large outdegree (e.g. >40)

Out-degree distribution, with linear fit for >40





Fitting the model to organizational data

Fitting model to our data, we find:

- $l \approx 4$ (4x greater managers at each level than level above)
- $a \approx 7$ (each manager has 6 support staff)
- Power law terminates at $x \approx 3$, indicating about four levels of management above each employee
- Based on organizational observation, *l* appears to be too small, *a* somewhat large, and levels of management approximately correct
- Further work needed to refine this model

Email traffic and number of senders, one week