VirginiaTech
Institute for Critical Technology and Applied Science

Ted and Karyn
Hume Center for National Security and Technology

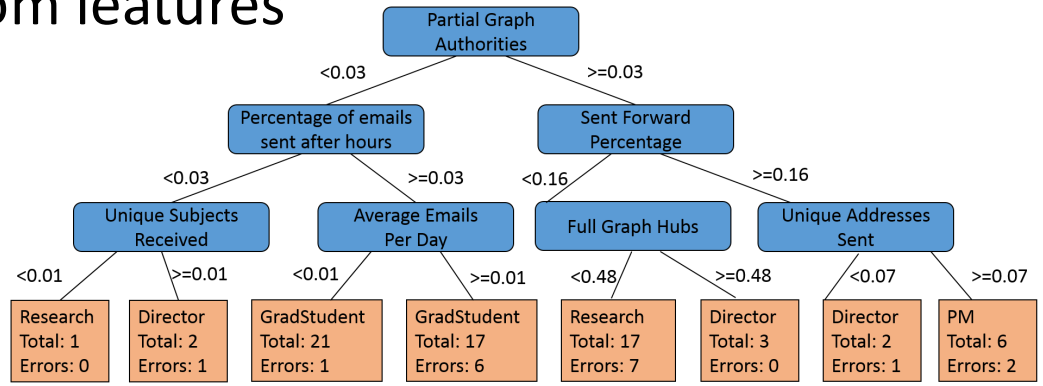
Data Mining Academic Emails to Model Employee Behaviors and Analyze Organizational Structure

2/3/16

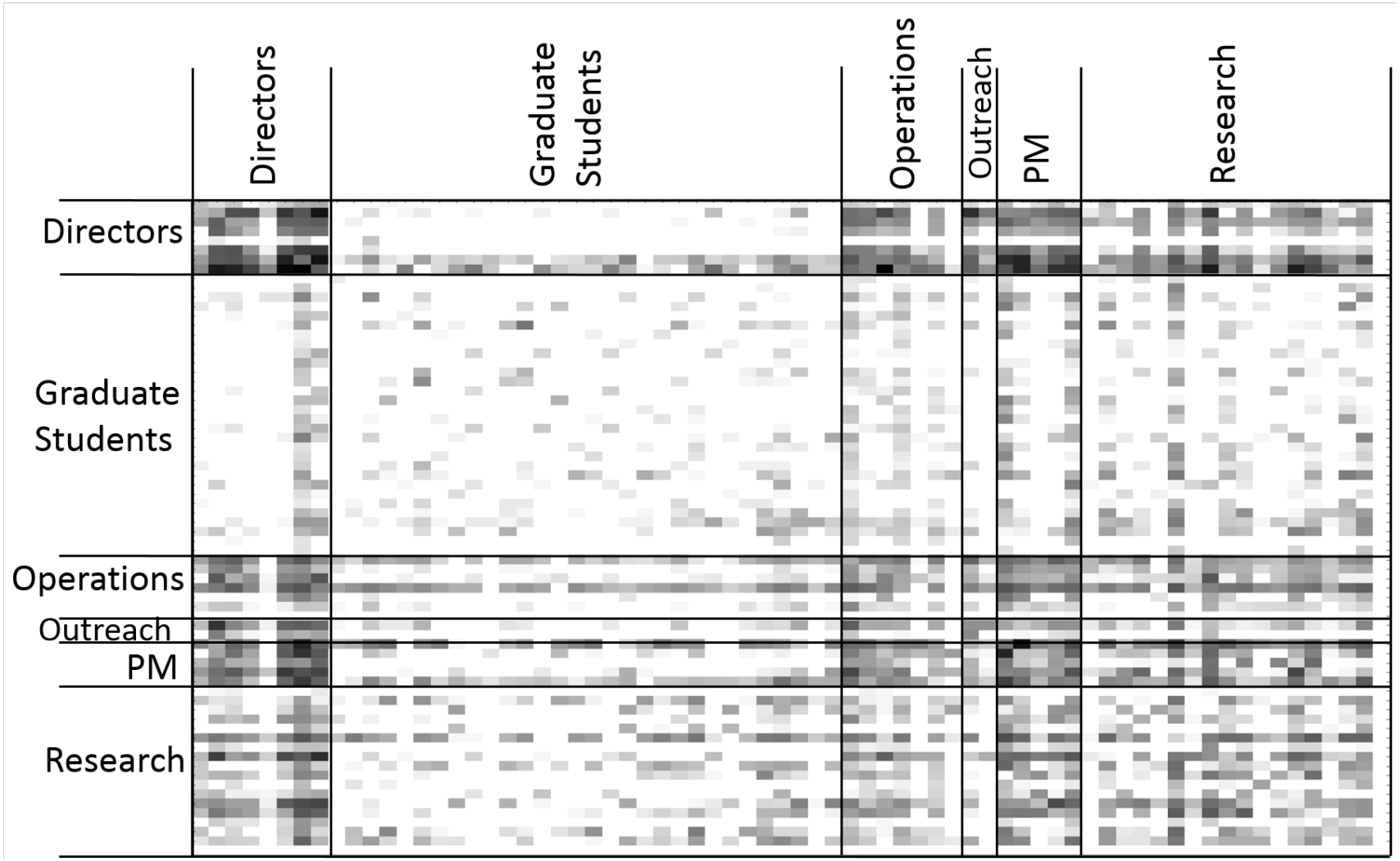
hume@vt.edu
www.hume.vt.edu

Summary

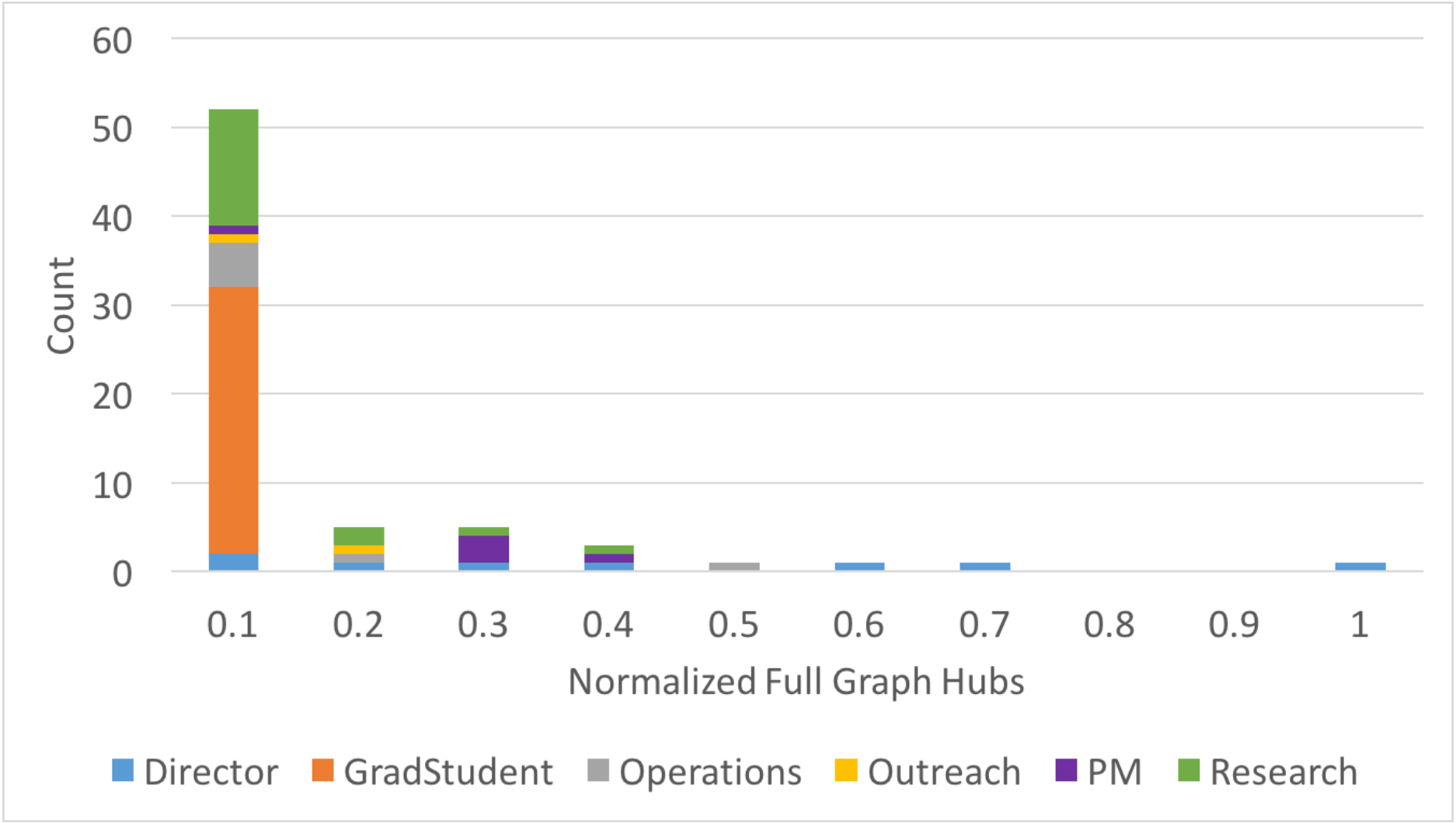
- Problem
  - Email is everywhere!
  - Can a better understanding of an organization be gleaned from internal emails?
  - Application: business reorganization
- Approach
  - Enron dataset has issues, so collect anonymous email metadata from volunteers
  - Calculate 98 different features, including traffic-based and social-based
    - Unique subjects received, number of forwards received, hubs
  - Use random forest algorithm to predict job title from features
  - Compare trends in data with organizational chart



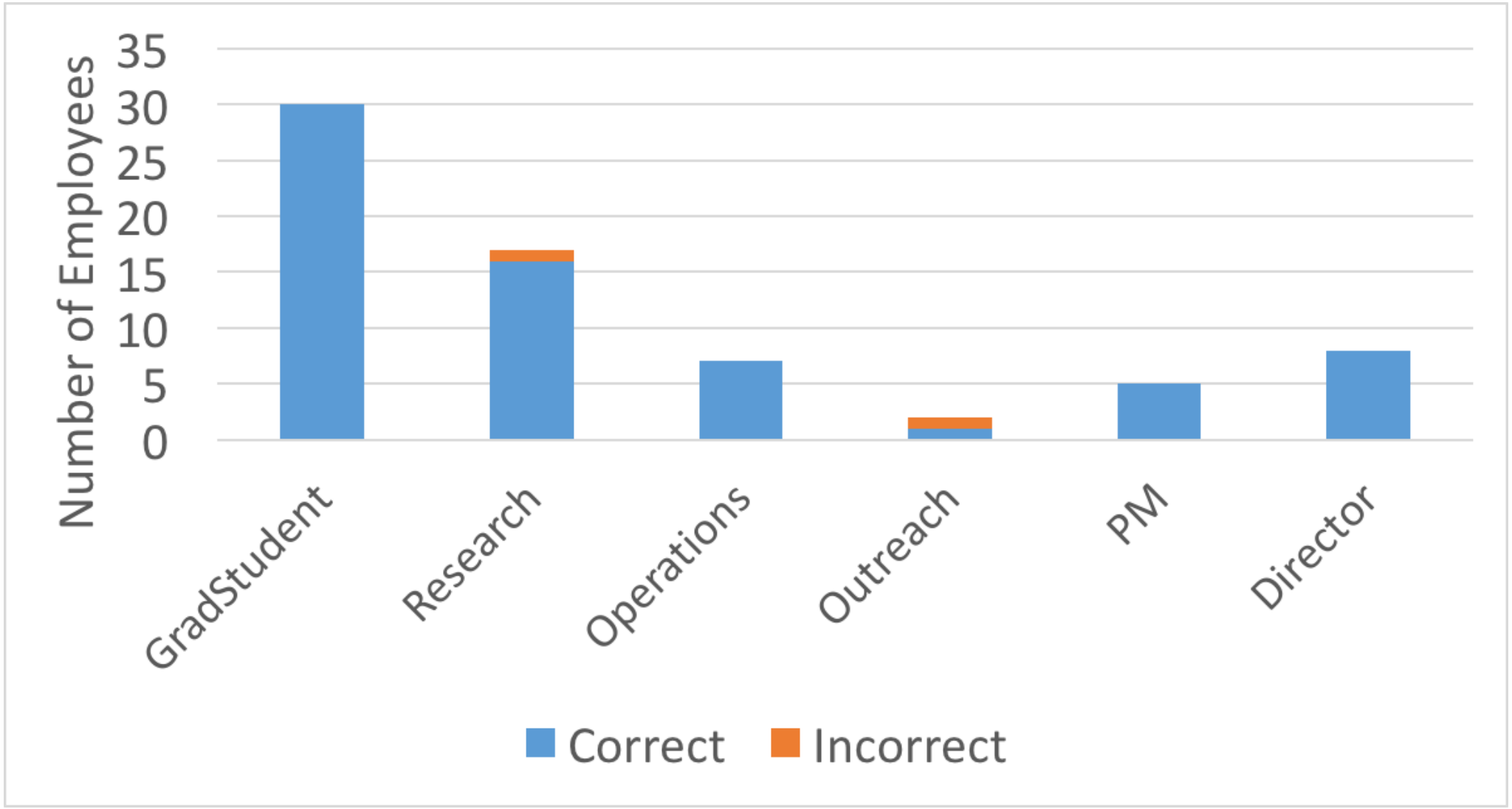
Results


Directors
Graduate Students
Operations
Outreach
PM
Research
Directors
Graduate Students
Operations
Outreach
PM
Research

Results


60
50
40
30
20
10
0
Count
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Normalized Full Graph Hubs
Director
GradStudent
Operations
Outreach
PM
Research

Results


Number of Employees
35
30
25
20
15
10
5
0
GradStudent
Research
Operations
Outreach
PM
Director
Correct
Incorrect

Questions?