

The Social Network of the Planetary Data System: A Comparative Analysis of Network Representations

Mark Avnet
Kate Martin

Supervisor:
Daniel Whitney

May 16, 2006
ESD.342 Advanced System Architecture

1. Introduction

Since the beginning of the space program, the National Aeronautics and Space Administration (NASA) has collected several terabytes of data about our solar system. In fact, the amount of data available far exceeds the combined analytical capacity of the entire planetary science community. As part of an effort to mitigate this problem, NASA has created a data management tool called the Planetary Data System (PDS).¹ The PDS, established in the early 1990s, is an arrangement of eight nodes located at various NASA Centers, universities, and research institutions across the country. Although the use of the PDS is currently somewhat limited, the intent is that it eventually will be the primary repository of all planetary data. Therefore, analyzing scientists' usage of the PDS can provide some important insights about the planetary science community. For example, answering a few key questions could help the PDS system operators understand why the system is not used to its full potential. One such question is: does the categorization of the data match the research communities that actually form around those data?

One purpose of this project is to develop a better understanding of the usage of the Planetary Data System. The system is used in two distinct but complementary ways. The first is that the collectors of data can upload their data to the system to, in a sense, "immortalize" it (a few people associated with PDS indicated in informal interviews that it was important to have a system like PDS because the datasets can last forever, but people do not). The second type of system usage occurs when scientists or others

¹ All background information about the Planetary Data System in this section from: National Aeronautics and Space Administration, "About PDS," *Planetary Data System*, <<http://pds.jpl.nasa.gov/aboutpds.html>>, accessed on May 11, 2006.

download the data for research or other purposes. This study focuses on the first type of usage because data for the second type were sparse.

To study patterns of collaboration on the usage of the PDS, several representations of the affiliation network are used. In the representation most typical to the social network literature, a network was created with authors (of datasets) as nodes and datasets as the edges. Other information about the PDS and the datasets are used to compare several representations of essentially the same network of author collaborations and planetary data. This information is also used to study the community structure and to compare community algorithm results and centrality measures to contextual understanding of the author roles and interactions. The use of multiple network representations allows us to fulfill a second purpose – to examine some of the limitations of popular network analysis techniques and metrics. We examine the basic statistics of each representation of the network and consider the effects of representation and weighting on the results. The paper concludes with some suggestions for future work in both of these areas.

1.1. PDS Structure

The conceptual structure of the PDS involves a central Project Management office and eight separate nodes on which the data are contained. The term “PDS node” used here is quite distinct from the network nodes discussed in this paper. Furthermore, the PDS nodes are not just the servers on which the data are contained. The term actually refers to “teams” of people that span research centers and universities. The five science nodes, named for five major sub-disciplines of planetary science, are Atmospheres, Geosciences, Planetary Plasma Interactions (PPI), Rings, and Small Bodies. The three remaining nodes provide support. These nodes are Engineering, which “provides systems engineering support to the entire project;” Navigational Ancillary Information Facility (NAIF), which “supplies calibration and ephemeris information;” and Imaging, which “offers expertise in sophisticated image processing.” Figure 1 shows the conceptual layout of the PDS.

NODES/SUBNODES/DATA NODES

Function

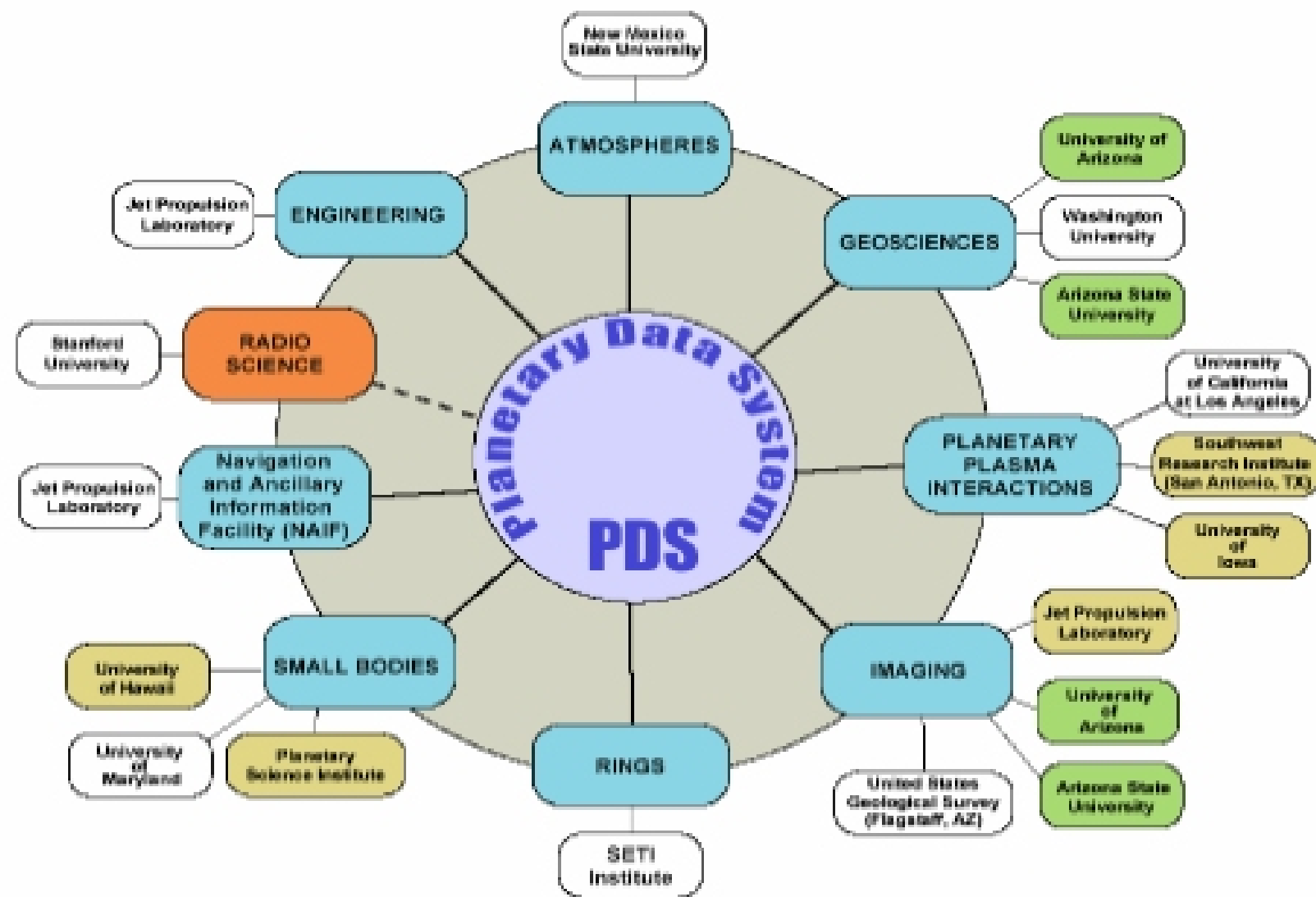


Figure 1. The conceptual structure of the Planetary Data System.
Image courtesy of NASA. *Source:* <http://pds.jpl.nasa.gov/files/node.pdf>.

The Project Management function of the PDS is located at NASA Goddard Space Flight Center (GSFC) in the Solar System Exploration Data Services Office. This, however, implies a level of centralized control that does not exist in the actual implementation of the system. Each of the nodes run independently of the others, and the use of each node is dependent on the people managing it. The leadership of the nodes changes every five years (the third selection was completed in June 2004) via a competitive grant solicitation process run by NASA. These changes in PDS node leadership can affect activity on the node. In some cases, the universities desire management responsibility for a node not because of the scientific importance of the PDS but because of political pressure and promise of funding for related programs. In many ways, the PDS is as much a public relations tool as a data management tool. Therefore, the actual structure of the PDS that has arisen from the loose-knit usage of the system is rather different from the prescribed structure shown in Figure 1.