NECSI Technical Report 2002-11-01

November 2002

Asymmetric Topology of Distributed Problem Solving Networks

Dan Braha[1] and Yaneer Bar-Yam[2]

[1] Center for Innovation in Product Development, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139, USA; Department of Industrial Engineering, Ben-Gurion University

e-mail: Braha@mit.edu

[2] New England Complex Systems Institute, Cambridge, Massachusetts 02138, USA

Distributed problem solving, which often involves an intricate network of interconnected tasks carried out by hundreds of actors, is fundamental to the creation of manmade systems[1] as well as the organization of work in biological systems[2]. Here we analyze, for the first time, the statistical properties of four human large-scale distributed task networks. We find that the distribution of outgoing communication links is scale-free[3-6] (power law decay) with or without a cutoff[7-10]. The distribution of incoming information flows always has a cutoff, and when both distributions have cutoffs the incoming distribution has a cutoff that is lower by more than a factor of two. The functional significance of this asymmetric topology can be explained by considering the dynamical interactions that take place in distributed problem solving.

The study of many ‘real-world’ social, biological and technological networks[11] have been shown to be neither completely regular nor completely random. Instead, these networks are typified as “small-world” networks[11, 12], combining the large degree local clustering of connections characteristic of regular networks with the small average path length of random graphs. Empirical work shows that the total node degree distribution of a variety of real networks has a scale-free regime, possibly with a cutoff[7-10]. The existence of a cutoff has been attributed to physical costs of adding links or limited capacity of a vertex[8].

We analyzed distributed problem solving data of four different large-scale organizations in the United States and England involved in vehicle development, operating software development, pharmaceutical facility development, and a sixteen story hospital facility development. The information flow forming the directed links between the tasks has been constructed by the organizations, and was based on design documentation data, structured interviews with experienced engineers, and a survey of individual decision makers. In all cases the resulting directed networks are highly sparse (see Figure 1); indicate the "small-world" character[11,12]; and show a significant negative correlation between the node's degree and its clustering coefficient[12], which suggests an inherently hierarchical organization with embedded modularity[13].

We next compared the cumulative probability distributions $P_{in}(k)$ and $P_{out}(k)$ that a task has more than k incoming and outgoing links, respectively (see Fig. 1). For all four networks, we find scaling regimes for both $P_{in}(k)$ and $P_{out}(k)$; however, the cutoff k^* occurs lower for $P_{in}(k)$ than for $P_{out}(k)$. Our findings suggest that although the cutoff may be attributed to constraints on the information-processing capacities of the actors[8] (so-called bounded-rationality[1]), there is asymmetry between the distributions of incoming and outgoing information flows. The narrower power law regime for $P_{in}(k)$ suggests that the costs of adding incoming links and limited in-degree capacity of a task are higher than their counterpart out-degree links.

The functional significance of the asymmetric topology can be attributed to the distinct roles of incoming and outgoing links in distributed problem solving. The narrow scaling regime governing the information flowing into a task implies that tasks with large incoming connectivity are practically absent. This indicates that distributed problem solving networks strive to limit conflicts by reducing the multiplicity of interactions that affect a single task, as reflected in the incoming links. This characteristic reduces the amount and range of potential revisions that occur in the dynamic problem solving process, and thus increases the likelihood of converging to a successful solution. The scale-free nature of the outgoing communication links means that some tasks communicate their outcomes to many more tasks than others do, and may play the role of coordinators. Unlike the case of large numbers of incoming links, this may improve the integration and consistency of the problem solving process; thus reducing the number of potential conflicts.

References

1. Simon, H. A. *The Sciences of the Artificial* (MIT Press, Cambridge, MA, 1998).

2. Visscher, P.K., and Camazine, S. *Nature*. **397**, 400 (1999).

3. Barabási, A.-L.& Albert, R. *Science* **286**, 509-512 (1999).

4. Albert, R., Jeong, H. & Barabási, A.-L. *Nature* **401**, 130 (1999).

5. Faloutsos, M., Faloutsos, P. & Faloutsos, C. *Comp. Comm. Rev*. **29**, 251-262 (1999).

6. Jeong H., Tombor B., Albert R., Oltavi, Z. N., & Barabási, A.-L. *Nature* **407**, 651-654 (2000).

7. Newman, M. E. J. *Proc. Nat. Ac. Sci USA* **98**, 404 (2001).

8. Amaral, L. A. N., Scala, A., Barthélémy, M., & Stanley, H. E. *Proc. Nat. Ac. Sci USA* **97**, 11149-11152 (2000).

9. Jeong, H., Mason, S., Barabási A.-L. & Oltvai, Z. N. *Nature* **411**, 41 (2001).

10. Cancho, R. F., Janssen, C. & Solé, R. V. *Phys. Rev. E* **63** (2001).

11. Strogatz, S. H. *Nature* **410**, 268-276 (2001).

12. Watts, D. J. & Strogatz, S.H. *Nature* **393**, 440-442 (1998).

13. Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A.-L. *Science* **297**, 1551-1555 (2002).

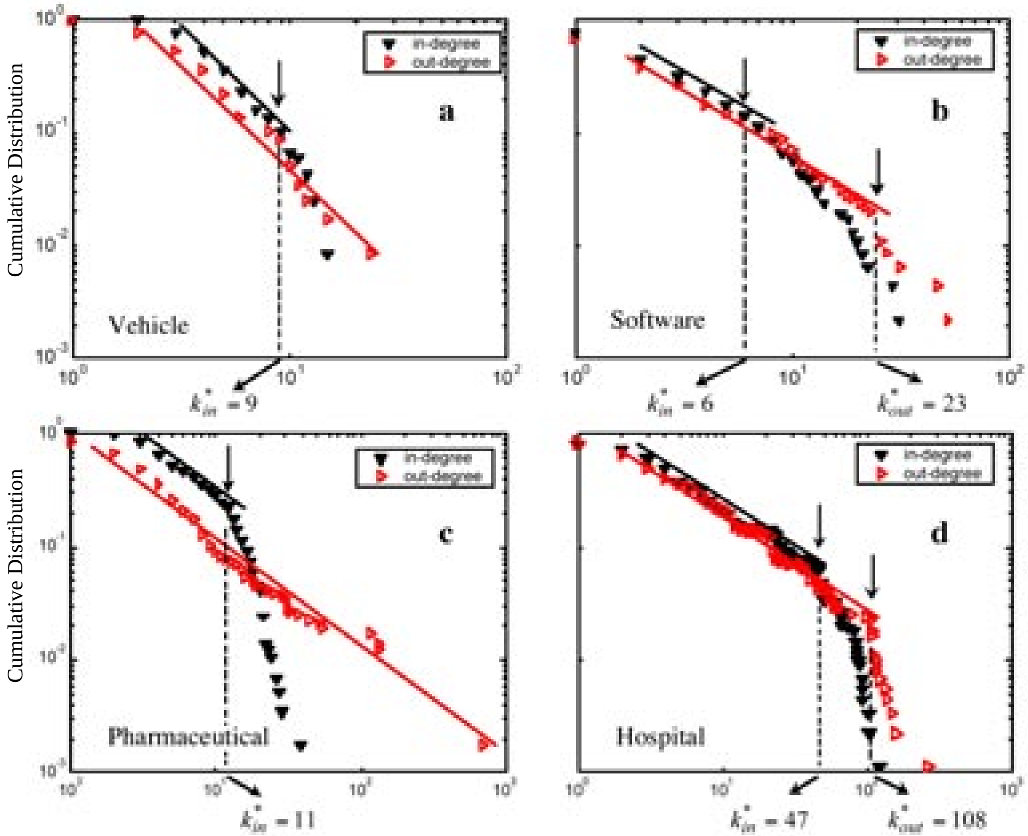

Figure 1 Degree distributions for four distributed problem solving networks. The log-log plots of the cumulative distributions of incoming and outgoing links show a power law regime (Pearson coefficient $R > 0.98$, $p < 0.001$) with or without a fast decaying tail in all cases. The in-degree distribution has a lower best-fit cutoff k_{in}^{*} in each case. **a**, Vehicle development with 120 tasks and 417 arcs. The exponents of the cumulative distributions are $\gamma_{vehicle}^{in} - 1$ and $\gamma_{vehicle}^{out} - 1$, where $\gamma_{vehicle}^{in} \approx 2.91$ and $\gamma_{vehicle}^{out} = 2.97$ denote the exponents of the associated probability density functions. **b**, Software development with 466 tasks and 1245 arcs, where $\gamma_{software}^{in} \approx 1.97$ and $\gamma_{software}^{out} \approx 2.17$. **c**, Pharmaceutical facility development with 582 tasks and 4123 arcs, where $\gamma_{pharmaceutical}^{in} \approx 1.8$ and $\gamma_{pharmaceutical}^{out} \approx 1.96$. **d**, Hospital facility development with 889 tasks and 8178 arcs, where $\gamma_{hospital}^{in} \approx 1.76$ and $\gamma_{hospital}^{out} \approx 1.89$.