

# The Topology and Dynamics of Complex Man-Made Systems

Dan Braha

New England Complex Institute

Cambridge, MA, USA

University of Massachusetts

Dartmouth, MA, USA

[braha@necsi.edu](mailto:braha@necsi.edu)

[http://necsi.edu/affiliates/braha/dan\\_braha-Description.htm](http://necsi.edu/affiliates/braha/dan_braha-Description.htm)

**Abstract** - Understanding the structure and function of complex networks has recently become the foundation for explaining many different real-world complex biological, technological and informal social phenomena. The analysis of these networks has uncovered surprising statistical structural properties that have also been shown to have a major effect on their functionality, dynamics, robustness, and fragility. This paper examines the statistical properties of large-scale product development networks -- and discusses the significance of these properties in providing insight into ways of improving the strategic and operational decision-making of the organization. We believe that our new analysis methodology and empirical results are also relevant to other organizational information-carrying networks.

**Keywords:** Complex Product Design, Large-Scale Engineering Systems, Collective Decision Making, Social Networks, Complex Systems, Robustness, Network Dynamics, Network Effects, Tippy Dynamics

## 1 The Connectivity Syndrome

A large-scale product design and development process (PD) is a distributed problem solving activity with hundreds of designers carrying out tasks and revising their actions based on other peoples input [7-9]. If the amount of information generated by project participants is not properly controlled, project participants will be expected to act on this information creating even more information for other people to act on, etc. This chain reaction will result in unintentional delays that were not accounted for at the outset of the project. Which patterns of information connectivity lead to better project performance? What are the patterns of information connectivity observed in real-world large-scale PD organizations? Do these observed patterns share common principles? And what can the structure of connectivity teach us about PD dynamics?

The above questions are addressed by applying techniques from complex networks theory, which is reviewed in Section 2. In Section 3, we present an analysis of the PD task networks, their 'small-world' property, and node connectivity distributions. We demonstrate the

distinct roles of incoming and outgoing information flows in distributed PD processes by analyzing the corresponding in-degree and out-degree link distributions. In Sections 4 and 5, we show that the statistical structural properties of PD projects have a major effect on their functionality, dynamics, robustness, and fragility. In Section 6 we present our conclusions.

## 2 Structural Properties of Complex Networks

Complex networks can be defined formally in terms of a graph  $G = (V, E)$ , which is a pair of nodes  $V = \{1, 2, \dots, N\}$ , and a set of lines  $E = \{e_1, e_2, \dots, e_L\}$  between pairs of nodes. If the line between two nodes is non-directional, then the network is called *undirected*; otherwise, the network is called *directed*. A network is usually represented by a diagram, where nodes are drawn as points, undirected lines are drawn as edges and directed lines as arcs connecting the corresponding two nodes. Three properties have been used to characterize 'real-world' complex networks [9, 10]. The first characteristic is the average distance (geodesic) between two nodes, where the distance  $d(i, j)$  between nodes  $i$  and  $j$  is defined as the number of edges along the shortest path connecting them. The characteristic path length  $\ell$  is the average distance between any two vertices:

$$\ell = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij} \quad (1)$$

The second characteristic measures the tendency of vertices to be locally interconnected or to cluster in dense modules. The clustering coefficient  $C_i$  of a vertex  $i$  is defined as follows. Let vertex  $i$  be connected to  $k_i$  neighbors. The total number of edges between these neighbors is at most  $k_i(k_i-1)/2$ . If the actual number of edges between these  $k_i$  neighbors is  $n_i$ , then the clustering coefficient  $C_i$  of the vertex  $i$  is the ratio

$$C_i = \frac{2n_i}{k_i(k_i - 1)} \quad (2)$$

The clustering coefficient of the graph, which is a measure of the network's potential modularity, is the average over all vertices,

$$C = \frac{1}{N} \sum_{i=1}^N C_i \quad (3)$$

The third characteristic is the distribution of degrees of vertices. The degree of a vertex, denoted by  $k_i$ , is the number of nodes adjacent to it. The mean nodal degree is the average degree of the nodes in the network,

$$\langle k \rangle = \frac{\sum_{i=1}^N k_i}{N} \quad (4)$$

If the network is directed, a distinction is made between the *in-degree* of a node and its *out-degree*. The in-degree of a node,  $k_{in}(i)$ , is the number of nodes that are *adjacent to*  $i$ . The out-degree of a node,  $k_{out}(i)$ , is the number of nodes *adjacent from*  $i$ .

Regular networks, where all the degrees of all the nodes are equal (such as circles, grids, and fully connected graphs) have been traditionally employed in modeling physical systems of atoms [4]. On the other hand, many 'real-world' social, biological and technological networks appear more random than regular [4, 6]. With the scarcity of large-scale empirical data on one hand and the lack of computing power on the other hand scientists have been led to model real-world networks as completely random graphs using the probabilistic graph models of Erdős and Rényi [10].

In their seminal paper on random graphs, Erdős and Rényi considered a model where  $N$  nodes are randomly connected with probability  $p$ . In this model, the average degree of the nodes in the network is  $\langle k \rangle \equiv pN$ , and a Poisson distribution approximates the distribution of the nodal degree. In a Poisson random network, the probability of nodes with at least  $k$  edges decays rapidly for large values of  $k$ . Consequently, a typical Poisson random network is rather homogenous, where most of the nodal degrees are concentrated around the mean. In particular, the average distance between any pair of nodes  $\ell_{\text{random}}$  scales with the number of nodes as  $\ell_{\text{random}} \sim \ln(N)/\ln(\langle k \rangle)$ . This feature of having a relatively short path between any two nodes, despite the often large graph size, is known as the *small-world* effect. In a Poisson random graph, the clustering coefficient is  $C_{\text{random}} = p \equiv \langle k \rangle / N$ . Thus, while

the average distance between any pair of nodes grows only logarithmically with  $N$  the Poisson random graph is poorly clustered.

Regular networks and random graphs serve as useful models for complex systems; yet, many real networks are neither completely ordered nor completely random. It has been found that social, technological, and biological networks are much more highly clustered than a random graph with the same number of nodes and edges (i.e.,  $C_{\text{real}} \gg C_{\text{random}}$ ), while the characteristic path length  $\ell_{\text{real}}$  is close to the theoretically minimum distance obtained for a random graph with the same average connectivity [6]. Small-World Networks are a class of graphs that are highly clustered like regular graphs ( $C_{\text{real}} \gg C_{\text{random}}$ ), but with a small characteristic path length like a random graph ( $\ell_{\text{real}} \approx \ell_{\text{random}}$ ). Many real-world complex systems have been shown to be small-world networks, including power-line grids, neuronal networks, social networks, the World-Wide Web, the Internet, food webs, and chemical-reaction networks.

Another important characteristic of real-world networks is related to their nodal degree distribution. Unlike the bell-shaped Poisson distribution of random graphs, the degree distribution of many real-world networks have been documented to have power-law degree distribution,

$$p(k) \sim k^{-\gamma} \quad (5)$$

where  $p(k)$  is the probability that a node has  $k$  edges. Networks with power-law distributions are often referred to as *scale-free networks* [6]. The power-law distribution implies that there are a few nodes with many edges; in other words, the distribution of nodal degrees has a long right tail (resulting in an extremely large variance) of values that are far above the mean (as opposed to the fast decaying tail of a Poisson distribution, which results in a small variance). Power-law distributions of both the in-degree and out-degree of a node have been also observed in a variety of directed real-world networks [6] including the World-Wide Web, metabolic networks, networks of citations of scientific papers, and telephone call graphs. Although scale-free networks are prevalent, the power-law distribution is not universal. Empirical work shows that the total node degree distribution of a variety of real networks often has a scale-free regime with an exponential cutoff, i.e.  $p(k) \sim k^{-\gamma} f(k/k^*)$  where  $k^*$  is the cutoff [11]. The existence of a cutoff has been attributed to physical costs of adding links or limited capacity of a vertex [11]. In some networks, the power-law regime is not even present and the nodal degree distribution is characterized by a distribution with a fast decaying tail. It is also not clear that a scale-free network optimizes properties of network behavior, and alternatives have been proposed [12]. Moreover, a recent

study of the dynamics of link utilization in complex networks offers a radical alternative to the static-based view of complex networks [14].

The goal of the present paper is to investigate the statistical properties of large-scale distributed product development networks. We show that large-scale PD networks, although of a different nature, have general properties that are shared by other social, technological, and biological networks.

### 3 Structural Properties of Complex Product Development Networks

Recent studies [1-3] provide the first empirical and theoretical explanation for the patterning of information flows in large-scale project environments. Information flow data of four large-scale projects was reconstructed by detailed field study in several major companies in the US and UK [1-3]. Each project includes hundreds of tasks connected to each other by an intricate web of information forming a highly complex network of activities forming the 'information highway' of the project. In these networks, a task might feed information to another task but not necessarily in the other direction (see Figure 1).

Strikingly, the statistical patterns of information flow embedded in these networks are found to be similar to those discovered in other information, technological or biological networks [6].

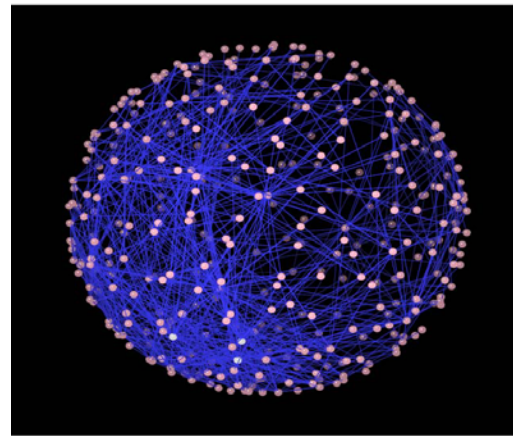
#### 3.1 Small-World Properties

It is found that PD complex networks exhibit the "small-world" property. As mentioned in Section 2, the 'small-world' property can be detected by measuring two basic statistical characteristics: 1) the average distance (geodesic) between two nodes; and 2) the clustering coefficient of the graph. Small-world networks are a class of graphs that are highly clustered like regular graphs ( $C_{\text{real}} \gg C_{\text{random}}$ ), but with small characteristic path length like a random graph ( $\ell_{\text{real}} \approx \ell_{\text{random}}$ ). For the software development network shown in Figure 1, the network is highly clustered as measured by the clustering coefficient of the graph ( $C_{\text{software}} = 0.327$ ) compared to a random graph with the same number of nodes and edges ( $C_{\text{random}} = 0.021$ ) but with small characteristic path length like a random graph ( $\ell_{\text{software}} = 3.700 \approx \ell_{\text{random}} = 3.448$ ).

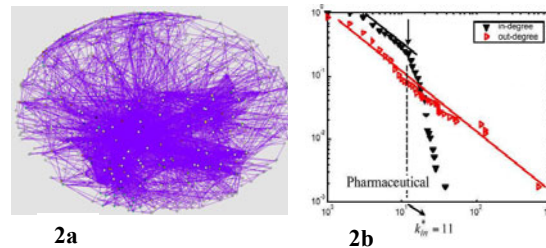
#### 3.2 In-degree and Out-degree Distributions

Two typical network images that come to mind are road maps, with their regular pattern of information flows (where the nodal in-degree is often like a bell-shaped Poisson distribution); and airport systems with their 'hub and spokes' patterns of connectivity (where the nodal out-degree is often like a power-law distribution). Careful

analysis of the data reveals a dual, asymmetric pattern of connectivity associated with the information flowing into tasks and information flowing out of tasks. More specifically, the incoming information flow is found to behave like road maps (see Figure 2b), while the outgoing information flow is found to behave like airport systems (see Figure 2b). The noticeable asymmetry between the distributions of incoming and outgoing information flows exhibited by large-scale design networks suggests that the incoming capacities of tasks are much more limited than their counterpart outgoing capacities. The cut-offs observed in the in-degree and out-degree distributions might reflect Simon's notion of bounded rationality [13], and its extension to group-level information processing. It is also found that tasks that generate a lot of information to other tasks do not consume, in general, a lot of information from others.



**Figure 1.** Network of information flows between tasks of an operating system large-scale design. These task networks consist of enormous number of directed information flows between hundreds of development tasks. Each task is assigned to one or more actors ("design teams," "engineers," or "scientists") who are responsible for it. Here, the information links are directed – each task consumes information *from* others and generates information *to* others.



**Figure 2.** (a) Network of information flows between tasks of pharmaceutical facility design with 582 tasks and 4123 arcs. (b) The log-log plot of the cumulative distribution of outgoing links shows a power-law regime. The very low cutoff of the in-degree distribution suggests a single scale (characteristic of exponential decay) for the in-degree connectivities.

## 4 Dynamics of Complex Product Development Networks

To explain the structural patterns reported above, one has to find a bridge between structural properties of product development networks and their resulting performance measures. To put things in focus, the following question is posed: Which network topologies would make the PD process converge faster to a state where all ‘open issues’ or ‘open tasks’ have been resolved? The key insight to addressing this question is to apply a metaphor from epidemiological theory. According to this metaphor, ‘open tasks’ diffuse to other connected tasks in the network, which might require additional rework and unnecessary elongation of the process. Like a filled bucket of water with a hole at the bottom, if the rate of fixing problems is slower than the rate of creating problems, the process is expected to approach the vicious mode of ‘solve-and rework’. How is the network topology related to this behavior? Surprisingly, it is found that the larger the percentage of nodes in the network that are both ‘heavy’ information consumers and generators, the more extreme the ‘solve-and rework’ mode becomes. This theory might account for the high degree of negative correlation between ‘heavy’ information consumers and ‘heavy’ information generators found in real-world PD projects.

## 5 Robustness and Sensitivity of Product Development

But why are outgoing information flows like the ‘hub-and-spoke’ airport transportation system and the incoming information flows more like the regular road transportation system? Analysis of the model showed that this asymmetry renders two important properties of complex PD networks: *robustness* and *sensitivity*. Robustness refers to the ability of a network to absorb shocks (e.g., delays) taking place randomly at specific tasks. The ‘hub-and-spoke’ structure of outgoing information flow shields the system keeping it intact even when a significant number of randomly selected tasks become degraded. This is like assembling the perfect car from a collection of used car components. However, there is no free lunch – the Achilles’ heel of the PD network are those (few) hubs that generate and consume a lot of information. Errors associated with these hubs will shift the network dramatically towards the endless ‘solve-and rework’ mode. Although this sounds like ‘bad news,’ informed managers can protect these troublesome hubs ahead of time, immunizing the PD network from external noise. Also, focusing development efforts on these hubs can boost the PD performance significantly. Thus, if wisely exploited, the sensitivity of PD complex networks to variations that are targeted at specific tasks can yield great benefits with minimal effort, yet the sensitivity characteristic may also result in detrimental effects if not properly controlled. The surprising fact here is that the hub-and-spoke structure was not designed purposefully, and yet

without it the PD network would not perform satisfactorily. Evolutionary forces shaping PD projects must have been the driving force leading to the hub-and-spoke structures.

## 6 Conclusion

In the last few years, the study of complex network topologies has become a rapidly advancing area of research across many fields of science and technology [6]. One of the key areas of research is understanding the network properties that are optimized by specific network architectures. Here we have analyzed the statistical properties of real-world networks of people engaged in product development activities. We have shown that complex PD networks display similar statistical patterns to other real-world networks of different origins, and have shown how the underlying network topologies provide direct information about the characteristics of PD dynamics. In particular:

- PD complex networks exhibit the “small-world” property, which means that they react rapidly to changes in design status;
- PD complex networks are characterized by inhomogeneous distributions of incoming and outgoing information flows of tasks. Consequently, PD task networks are dominated by a few highly central ‘information-consuming’ and ‘information-generating’ tasks;
- PD networks exhibit a noticeable *asymmetry* (related to the cut-offs) between the distributions of incoming and outgoing information flows, suggesting that the incoming capacities of tasks are much more limited than their counterpart outgoing capacities. The cut-offs observed in the in-degree and out-degree distributions might reflect Herbert Simon’s notion of bounded rationality [13], and its extension to *group-level* information processing.
- Focusing engineering and management efforts on central ‘information-consuming’ and ‘information-generating’ PD tasks will likely improve the performance of the overall PD process;
- ‘Failure’ of central PD tasks affects the vulnerability of the overall PD process; Positive correlation between the in-degree and out-degree of a task tends to limit the range of the parameters’ values for which the system converges to the uniformly resolved state.

PD dynamics is highly error tolerant, yet highly responsive to perturbations that are targeted at specific tasks. In the context of product development, what is the meaning of these patterns? How do they come to be what they are? We propose several explanations for these patterns. Successful PD processes in competitive environments are often

characterized by short time-to-market, high product performance, and low development costs [7, 8]. In many high technology industries, an important tradeoff exists between minimizing time-to-market and development costs and maximizing the product performance. In PD task networks, accelerating the PD process can be achieved by “cutting out” some of the links between the tasks [8, 9]. Although the elimination of some arcs should result in more rapid PD convergence, this might worsen the performance of the end system. Consequently, a tradeoff exists between the elimination of task dependencies (speeding up the process) and the desire to improve the system’s performance through the incorporation of additional task dependencies. PD networks are likely to be highly optimized when both PD completion time and product performance are accounted for. Recent studies have shown that an evolutionary algorithm involving *minimization of link density and average distance* between any pair of nodes can lead to non-trivial types of networks including truncated scale-free networks; i.e.  $p(k) = k^{-\gamma} f(k/k^*)$ . This might suggest that an evolutionary process that incorporates similar generic optimization mechanisms (e.g., minimizing a weighted sum of development time and product quality losses) might lead to the formation of a PD network structure with the small-world and truncated scale-free properties.

Another explanation for the characteristic patterns of PD networks might be related to the close interplay between the design structure (product architecture) and the related organization of tasks involved in the design process. It has been observed that in many technical systems design tasks are commonly organized around the architecture of the product. Consequently, there is a strong association between the information flows underlying the PD task network and the design network composed of the physical (or logical) components of the product and the interfaces between them. If the task network is a “mirror image” of the related design network, it is reasonable that their large-scale statistical properties might be similar. Evidence for this can be found in recent empirical studies that show some design networks (electronic circuits and software architectures) exhibit small-world and scaling properties. The scale-free structure of design networks, in turn, might reflect the strategy adopted by many firms of reusing existing modules together with newly developed modules in future product architectures. Thus, the highly connected nodes of the scale-free design network tend to be the most reusable modules. Reusing modules at the product architecture level has also a direct effect on the task level of product development; it allows firms to reduce the complexity and scope of the product development project by exploiting the knowledge embedded in reused modules, and thus significantly reduce the product development time.

Projects can fail due to network effects that result in unstable internal information processes that drive the

project out of control into the ‘solve-and rework’ mode of operation. The emerging discipline of complex networks research offers a new and potentially powerful perspective on managing large-scale projects. Much like successful navigation depends on detailed and precise geomaps, PD managers could promote informed decision making by first mapping the information flow highway of their PD environment. Managers can then identify the ‘information bottlenecks’ – those tasks and activities that consume and generate a significant amount of information. Luckily, the hub-and-spoke structure suggests that there are not many of those. This ‘law of the vital few’ is a manifestation of the Pareto principle (also known as the 80/20 rule) expressed in the language of PD networks. Targeted improvement efforts can then be allocated to these vital few. If a large percentage of ‘heavy’ information consumers and generators in the network are present, rewiring the information links in the network might be needed. This restructuring procedure has to take place in the context of existing resources and engineering constraints.

Of greatest significance for the analysis of generic network architectures, we demonstrated a previously unreported difference between the distribution of incoming and outgoing links in a complex network. Specifically, we find that the distribution of incoming communication links always has a cutoff, while outgoing communication links is scale-free with or without a cutoff. In the cases studied, when both distributions have cutoffs, the incoming distribution has a cutoff that is significantly lower by more than a factor of two. From a product development viewpoint, the functional significance of this asymmetric topology has been explained by considering a bounded-rationality argument originally put forward by Simon in the context of human interactions [1-3]. Accordingly, this asymmetry could be interpreted as indicating a limitation on the actor's capacity to process information provided by others rather than the ability to transmit information over the network. In the latter case, boundedness is less apparent since the capacity required to transmit information over a network is often less constrained, especially when it is replicated (e.g., many actors can receive the same information from a single actor by broadcast). In light of this observation, we expect a distinct cut-off distribution for in-degree as opposed to out-degree distributions when the network reflects communication of information between human beings as a natural and direct outcome of Simon's bounded rationality argument. It seems reasonable to propose that the asymmetric link distribution is likely to hold for such networks when nodes represent information processing elements. Indeed, it is shown in [1] that this property can be found more generally in other directed human or non-human networks (see also **Appendix**).

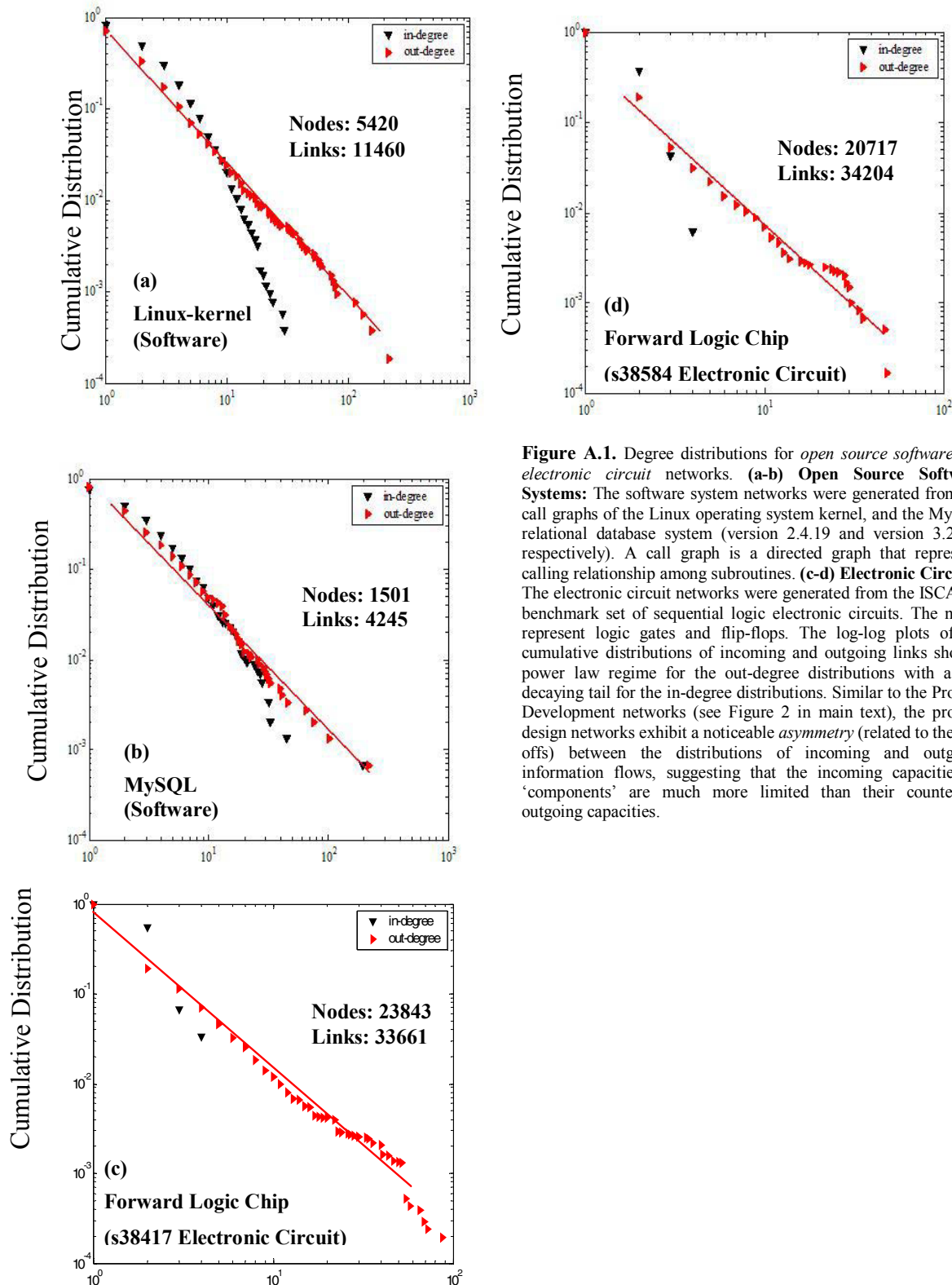
The paper analyzes an *intra-organizational* network where PD tasks are nodes. It would be interesting to see if the statistical patterns uncovered for intra-organizational networks remain invariant when moving to the *inter-*

*organizational* level where enterprises form the nodes. We conjecture that the level of abstraction will not significantly change the qualitative structure of the network's topology; but may change the embedded parameters underlying the network's characteristics (e.g., coefficients and cut-offs of the power-law distributions). We have identified two generic categories of network nodes: "information-consuming" and "information-generating." We believe that this categorization could be expanded by at least three methods: 1) considering other unit centrality measures (e.g., closeness and betweenness centrality, see [3] for initial analysis); 2) analyzing the structure of sub-graphs ("building blocks") embedded in the networks; 3) assigning richer data structures that more naturally describe a product development; e.g., adding characteristics to each task or adding information bandwidth (weights) to links.

## References

- [1] D. Braha and Y. Bar-Yam, "The Statistical Mechanics of Complex Product Development: Empirical and Analytical Results," *Management Science*, Vol. 53, No. 7, pp. 1127-1145, July 2007.
- [2] D. Braha and Y. Bar-Yam, "The Topology of Large-Scale Engineering Problem-Solving Networks," *Physical Review E*, Vol. 69, No. 1, 2004.
- [3] D. Braha and Y. Bar-Yam, "Information Flow Structure in Large-Scale Product Development Organizational Networks," *Journal of Information Technology*, Vol. 19, No. 4, pp. 234-244, 2004.
- [4] S. H. Strogatz, "Exploring Complex Networks," *Nature*, Vol. 410, pp. 268-276, 2001.
- [5] H. A. Simon, *The Sciences of the Artificial* (MIT Press, Cambridge, MA, 1998). [9] R. Albert and Barabási, A.-L., "Statistical Mechanics of Complex Networks," *Reviews of Modern Physics* 74, 47-97 (2002).
- [6] R. Albert and Barabási, A.-L., "Statistical Mechanics of Complex Networks," *Reviews of Modern Physics* 74, 47-97 (2002).
- [7] D. Braha and O. Maimon, *A Mathematical Theory of Design: Foundations, Algorithms, and Applications* (Kluwer Academic Publishers, Boston, MA, 1998).
- [8] A. Yassine, and D. Braha, "Complex Concurrent Engineering and the Design Structure Matrix Method," *Concurrent Engineering*, September 2003, vol. 11, no. 3, pp. 165-176.
- [9] A. Yassine, N. Joglekar, D. Braha, S. Eppinger and D. Whitney, Information Hiding in Product Development: The Design Churn Effect. *Research in Engineering Design*. Vol. 14 (3). 131-144.
- [10] P. Erdős, and A. Rényi, "On random graphs," *Publicationes Mathematicae* 6, 290-297 (1959).
- [11] L. A. N. Amaral, A. Scala, M. Barthélémy and H. E. Stanley, "Classes of Small-World Networks," *Proc. Nat. Ac. Sci USA* 97, 11149-11152 (2000).
- [12] Y. Bar-Yam, and I. R. Epstein, "Response of complex networks to stimuli," *Proceedings of the National Academy of Sciences* 101, pp. 4341-4345 (2004).
- [13] H. A. Simon, *The Sciences of the Artificial* (MIT Press, Cambridge, MA, 1998).
- [14] D. Braha and Y. Bar-Yam, "From Centrality to Temporary Fame: Dynamic Centrality in Complex Networks," *Complexity*, Vol. 12, No. 2, pp. 59-114, November/December 2006.

**Appendix.** Network Topology of Several Man-Made Systems



**Figure A.1.** Degree distributions for *open source software* and *electronic circuit* networks. **(a-b) Open Source Software Systems:** The software system networks were generated from the call graphs of the Linux operating system kernel, and the MySQL relational database system (version 2.4.19 and version 3.23.32 respectively). A call graph is a directed graph that represents calling relationship among subroutines. **(c-d) Electronic Circuits:** The electronic circuit networks were generated from the ISCAS89 benchmark set of sequential logic electronic circuits. The nodes represent logic gates and flip-flops. The log-log plots of the cumulative distributions of incoming and outgoing links show a power law regime for the out-degree distributions with a fast decaying tail for the in-degree distributions. Similar to the Product Development networks (see Figure 2 in main text), the product design networks exhibit a noticeable *asymmetry* (related to the cut-offs) between the distributions of incoming and outgoing information flows, suggesting that the incoming capacities of ‘components’ are much more limited than their counterpart outgoing capacities.