Special edition

Information flow structure in large-scale product development organizational networks

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

[1]Earle P. Charlton College of Business, University of Massachusetts, North Dartmouth, MA, USA;
[2]New England Complex Systems Institute, Cambridge, MA, USA

Correspondence:
D Braha, Earle P. Charlton College of Business, University of Massachusetts, Dartmouth, MA, USA.
E-mail: braha@necsi.org, dbraha@umassd.edu

Abstract
This paper analyzes the statistical properties of real-world networks of people engaged in product development (PD) activities. We show that complex PD networks display similar statistical patterns to other real-world complex social, information, biological and technological networks. The paper lays out the foundations for understanding the properties of other intra- and inter-organizational networks that are realized by specific network architectures. The paper also provides a general framework towards characterizing the functionality, dynamics, robustness, and fragility of smart business networks.
Journal of Information Technology (2004) **19,** 244–253 doi:10.1057/palgrave.jit.2000030
Published online 30 November 2004
Keywords: large-scale design; product development; socio-technical systems; information systems; social network analysis; complex engineering systems

Introduction

On February 1, 1997, a major fire swept through one of Aisin Seiki's plants supplying brake fluid proportioning valves (or P-valves) to all Toyota vehicles manufactured by Toyota-group plants in Japan (Reitman, 1997; Nishiguchi and Beaudet, 1998). The sole reliance of Toyota on Aisin Seiki's supply and the low inventory levels of the P-valves inventory due to a just-in-time (JIT) operating environment threatened to shut down Toyota's 20 auto plants in Japan for weeks and damage local economies. Surprisingly, Toyota's car factories succeeded to recover their operations in only 5 days after the fire. The admirable Toyota's quick recovery can be attributed to the cohesive network structure of suppliers working with Toyota directly and indirectly. This enabled Toyota to reconfigure rapidly the supply chain network and pull together 36 suppliers, supported by more than 150 subcontractors, who produced small batches of P-valves on nearly 50 separate improvised tooling systems and production lines (Reitman, 1997). The above supply chain disaster recovery illustrates the importance of coordination and collaboration among supply chain partners (e.g., manufacturers, suppliers, and retailers) as a means for achieving greater strategic and operational value to the organization. Today, supply chain integration is further realized by complex business-to-business interactions via

information technology, most importantly the Internet (Kambil and van Heck, 2002). In such supply chain networks, partners are involved in an intricate web of information transfer such as demand data, inventory status, and shipment schedules.

The usefulness of understanding organizational network structure as a tool for assessing the effects of decisions on organizational performance has been illustrated in the social science and management literatures (Cross *et al.*, 2002). There it has been shown that informal networks of relationships (e.g., communication, information, and problem-solving networks) – rather than formal organizational charts – determine to a large extent the patterns of coordination and work processes embedded in the organization (Cross *et al.*, 2002). In recent years, networks have also become the foundation for the understanding of numerous and disparate complex systems outside the field of social sciences (e.g., biology, ecology, engineering, and internet technology; see Albert and Barabási, 2002; Newman, 2003).

The goal of the this paper is to examine, for the first time, the statistical properties of an important large-scale information network – new product development – and discuss their significance in providing insight into ways of improving the strategic and operational decision-making of

the organization. In general, information networks constitute the infrastructure for exchanging knowledge that is important to the achievement of work by individual agents. We believe that our results will also be relevant to other information-based networks.

Distributed product development (abbreviated as 'PD'), which often involves an intricate network of interconnected tasks carried out by hundreds of designers, is fundamental to the creation of complex manmade systems (Alexander, 1964). The interdependence between the various tasks makes system development fundamentally iterative (Braha and Maimon, 1998). This process is driven by the repetition (rework) of tasks due to the availability of new information (generated by other tasks) such as changes in input, updates of shared assumptions or the discovery of errors. In such an intricate network of interactions, iterations occur when some development tasks are attempted even though the complete predecessor information is not available or known with certainty (Yassine and Braha, 2003). As this missing or uncertain information becomes available, the tasks are repeated to either verify an initial estimate/guess or to come closer to the design specifications. This iterative process proceeds until convergence occurs (Klein *et al.*, 2003; Yassine and Braha, 2003; Yassine *et al.*, 2003; Braha and Bar-Yam, 2004a, b).

Design iterations, which are the result of the PD network structure, might slow down the PD convergence or have a destabilizing effect on the system's behavior. This will delay the time required for product development, and thus compromise the effectiveness and efficiency of the PD process. For example, it is estimated that iteration costs about one-third of the whole PD time (Osborne, 1993) while lost profits result when new products are delayed in development and shipped late (Clark, 1989). Characterizing the *real-world* structure, and eventually the dynamics of complex PD networks, may lead to the development of guidelines for coping with complexity. It would also suggest ways for improving the decision-making process, and the search for innovative design solutions.

The last few years have witnessed substantial and dramatic new advances in understanding the large-scale structural properties of many real-world complex networks (Strogatz, 2001; Albert and Barabási, 2002; Newman, 2003). The availability of large-scale empirical data on the one hand and the advances in computing power and theoretical understanding on the other have led to a series of discoveries that have uncovered statistical properties that are common to a variety of diverse real-world social, information, biological and technological networks including the World-Wide Web (Albert *et al.*, 1999), the internet (Faloutsos *et al.*, 1999), power grids (Watts and Strogatz, 1998), metabolic and protein networks (Jeong *et al.*, 2000, 2001), food webs (Montoya and Solé, 2002), scientific collaboration networks (Amaral *et al.*, 2000; Newman, 2001a–c), citation networks (Price, 1965), electronic circuits (Ferrer *et al.*, 2001), and software architecture (Valverde *et al.*, 2002). These studies have shown that many complex networks exhibit the 'small-world' property of short average path lengths between any two nodes despite being highly clustered. They have also found that complex networks are characterized by an inhomogeneous distribution of nodal degrees (the number of nodes a particular

node is connected to) with this distribution often following a power law (termed 'scale-free' networks in Barabási and Albert, 1999). Scale-free networks have been shown to be robust to random failures of nodes, but vulnerable to failure of the highly connected nodes (Albert *et al.*, 2000). A variety of network growth processes that might occur on real networks, and that lead to scale-free and small-world networks have been proposed by Albert and Barabási (2002) and Newman (2003). The dynamics of networks can be understood to be due to processes propagating through the network of connections; the range of dynamical processes include disease spreading and diffusion, search and random walks, synchronization, games, Boolean networks and cellular automata, and rumor propagation. Indeed, the raison d'être of complex network studies might be said to be the finding that topology provides direct information about the characteristics of network dynamics. In this paper, we study network topologies in the context of large-scale product development and discuss their relationship to the functional utility of the system (a more detailed model of PD dynamics is presented elsewhere in Braha and Bar-Yam, 2004a).

Planning techniques and analytical models that view the PD process as a network of interacting components have been proposed before (Steward, 1981; Eppinger *et al.*, 1994; Yassine and Braha, 2003). However, others have not yet addressed the large-scale statistical properties of real-world PD task networks. In the research we report here, we study such networks. We show that task networks have properties (sparseness, small world, scaling regimes) that are like those of other biological, social, and technological networks. We also demonstrate a previously unreported asymmetry in the cutoffs between the distribution of incoming and outgoing links.

The paper is organized as follows: in the second section, we review the basic structural properties of real-world complex networks. In the third section, we describe the data on PD tasks. In the fourth section, we present an analysis of the PD task networks, their small-world property and node connectivity distributions, and classification of individual node prominence according to their position in the network. In particular, 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 the fifth section we present our conclusions.

Structural properties of complex networks

Complex networks can be defined formally in terms of a graph G, which is a pair $G = (V, E)$ consisting of two sets: a set of nodes $V = \{1, 2, \ldots, N\}$ and a set of lines $E = \{e_1, e_2, \ldots, e_L\}$ 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 small points, undirected lines are drawn as edges and directed lines as arcs connecting the corresponding two nodes. Three major characteristics have been identified to play a major role in the understanding of 'real-world' complex networks (Albert and Barabási, 2002; Newman, 2003). 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 ℓ is the average distance between any two vertices:

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

The second characteristic measures the tendency of vertices to cluster in densely interconnected 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

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

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

$$C = \frac{1}{N} \sum_{i=1}^{N} C_i. \tag{3}$$

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

$$\bar{k} = \frac{\sum_{i=1}^{N} k_i}{N}. \tag{4}$$

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 (Strogatz, 2001). On the other hand, many 'real-world' social, biological, and technological networks appear more random than regular (Strogatz, 2001; Albert and Barabási, 2002; Newman, 2003). With the scarcity of large-scale empirical data on one and the lack of computing power on the other scientists have been led to model real-world networks as completely random graphs using the probabilistic graph models of Erdös and Rényi (1959).

In their seminal paper on random graphs, Erdös and Rényi have 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 $\bar{k} \cong pN$, and a Poisson distribution approximates the distribution of the nodal degree. In a Poisson random graph, the probability of nodes with at least k edges decays rapidly for large values of k. Consequently, a typical Poisson graph is rather homogenous, where most of the nodal degrees are concentrated around the mean. The average distance between any pair of nodes in a Poisson random graph is the smallest among all connected graphs with the same number of nodes and edges. In particular, the average distance between any pair

of nodes ℓ_{random} scales with the number of nodes as $\ell_{random} \sim \ln(N)/\ln(\langle k \rangle)$. 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 $C_{random} = p \cong \bar{k}/N$. 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. Watts and Strogatz (1998) 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., $C_{real} \gg C_{random}$), while the characteristic path length ℓ_{real} is close to the theoretically minimum distance obtained for a random graph with the same average connectivity. Small-world networks are a class of graphs that are highly clustered like regular graphs ($C_{real} \gg C_{random}$), but with a small characteristic path length like a random graph ($\ell_{real} \approx \ell_{random}$). Many real-world complex systems have been shown to be small-world networks, including power-line grids (Watts and Strogatz, 1998), neuronal networks (Watts and Strogatz, 1998), social networks (Amaral *et al.*, 2000; Newman, 2001a–c), the World-Wide Web (Albert *et al.*, 1999), the Internet (Albert *et al.*, 2000), food webs (Montoya and Solé, 2002), and chemical-reaction networks (Jeong *et al.*, 2000).

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,

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

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* (Barabási and Albert, 1999). 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 of values that are far above the mean (as opposed to the fast decaying tail of a Poisson distribution). 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 (Albert and Barabási, 2002; Newman, 2003) 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 has a scale-free regime with an exponential cutoff, that is, $P(k) \sim k^{-\gamma} f(k/k^*)$ where k^* is the cutoff (Amaral *et al.*, 2000; Strogatz, 2001). The existence of a cutoff has been attributed to physical costs of adding links or limited capacity of a vertex (Amaral *et al.*, 2000). 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 (Amaral *et al.*, 2000; Strogatz, 2001).

The goal of the present paper is to investigate the statistical properties of large-scale distributed PD networks. We show that large-scale PD networks, although of a

different nature, have general properties that are shared by other social, information, technological, and biological networks.

Data

We analyzed distributed product development data of different large-scale organizations in the United States and England involved in vehicle design (Cividanes, 2002), operating software design (Denker),[1] pharmaceutical facility design, and a 16-story hospital facility design (Newton and Austin).[2] A PD distributed network can be considered as a directed graph with N nodes and L arcs, where there is an arc from task v_i to task v_j if task v_i *feeds information to* task v_j. The information flow forming the directed links between the tasks has been based on structured interviews with experienced engineers and design documentation data (design process models). In all cases, the repeated nature of the product development projects and the knowledgeable people involved in eliciting the information flow dependencies reduce the risk of error in the construction of the product development networks. More specifically, Cividanes (2002) obtained the vehicle development network by *directly* questioning at least one engineer from each task 'where do the inputs for the task come from (e.g., another task)?' and 'where do the outputs generated by the task go to (e.g., another task)?'[3] The answers to these questions were used by him to construct the network of information flows (Cividanes, 2002). The operating software development network was obtained from module/subsystems dependency diagrams compiled by Denker; and both the pharmaceutical facility development and the hospital

facility development networks were compiled by Newton and Austin[4] from data flow diagrams and design-process model diagrams Austin and Baldwin (1999) deployed by the organizations. An example of a diagram from the pharmaceutical facility and 16-story hospital facility process models is shown in Figure 1.

Results

### Small world properties

An example of one of these distributed PD networks (operating software development) is shown in Figure 2. Here we consider the undirected version of the network, where there is an edge between two tasks if they exchange information between them (not necessarily reciprocal). We see that this network is sparse ($2L/N(N-1) = 0.0114911$) with the average total degree of each node only 5.34, which is small compared to the number of possible edges $N-1 = 456$. A clear deviation from a purely random graph is observed.

We see that most of the nodes have low degree while a few nodes have a very large degree. This is in contrast to the nodal degree homogeneity of purely random graphs, where most of the nodal degrees are concentrated around the mean. The software development network also illustrates the 'small-world' property (see second section), which 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 ($C_{real} \gg C_{random}$), but with small characteristic
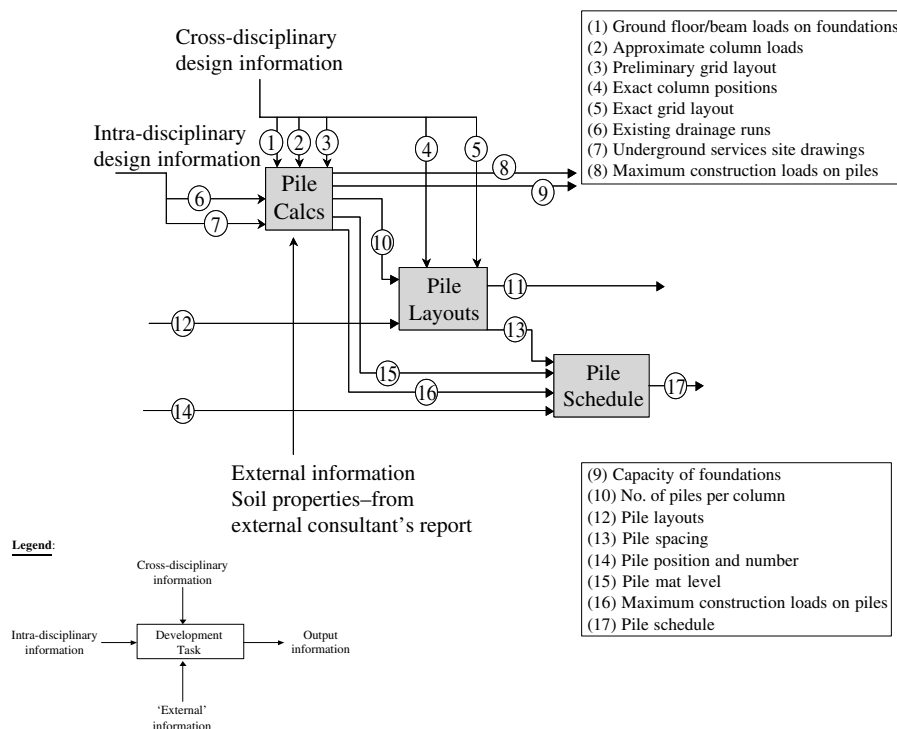
Figure 1 Example of a diagram from a design process model used to construct the pharmaceutical facility and the 16-story hospital facility networks (adapted from Austin *et al.*, 2000).
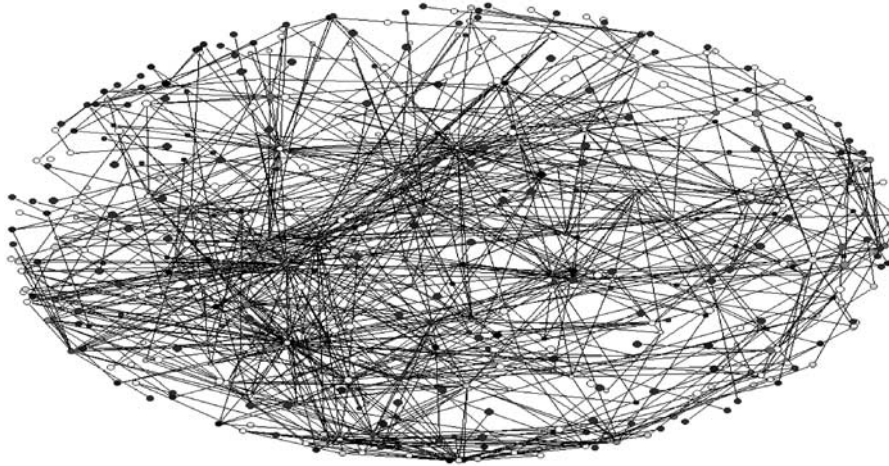
Figure 2 Network of information flows between tasks of an operating system development process. This PD task network consists of 1245 directed information flows between 466 development tasks. Each task is assigned to one or more actors ('design teams' or 'engineers') who are responsible for it. Nodes with the same degree are colored the same.

Table 1 Empirical statistics of the four large-scale PD networks

Network	N	L	C	ℓ	C_{random}	ℓ_{random}
Vehicle	120	417	0.205	2.878	0.070	2.698
Operating Software[a]	466	1245	0.327	3.700	0.021	3.448
Pharmaceutical Facility	582	4123	0.449	2.628	0.023	2.771
Sixteen-story Hospital Facility[a]	889	8178	0.274	3.118	0.024	2.583

[a]We restrict attention to the largest connected component of the graphs, which includes ~82% of all tasks for the Operating Software network, and ~92% of all tasks for the 16-story Hospital Facility network.

path length like a random graph ($\ell_{real} \approx \ell_{random}$). For the software development network, the network is highly clustered as measured by the clustering coefficient of the graph ($C_{software} = 0.327$) compared to a random graph with the same number of nodes and edges ($C_{random} = 0.021$) but with small characteristic path length like a random graph ($\ell_{software} = 3.700 \approx \ell_{random} = 3.448$).

In Table 1, we present the characteristic path length and clustering coefficient for the four distributed PD networks examined in this paper, and compare their values with random graphs having the same number of nodes and edges. In all cases, the empirical results display the small-world property ($C_{real} \gg C_{random}$ and $\ell_{real} \approx \ell_{random}$).

We restrict attention to the largest connected component of the graphs, which includes ~82% of all tasks for the Operating Software network, and ~92% of all tasks for the 16-story Hospital Facility network.

An interpretation of the functional significance of the architecture of PD networks must be based upon a recognition of the factors that such systems are optimizing. Shorter development times, improved product quality, and lower development costs are the key factors for successful complex PD processes. The existence of cycles in the PD networks, readily noted in the network architectures investigated, points to the seemingly undeniable truth that there is an inherent, iterative nature to the design process (Braha and Maimon, 1998). Each iteration results in changes that must propagate through the PD network

requiring the rework of other reachable tasks. Consequently, late feedback and excessive rework should be minimized if shorter development time is required.

The functional significance of the small-world property can be attributed to the fast information transfer throughout the network, which results in immediate response to the rework created by other tasks in the network. The high clustering coefficient of PD networks suggests an inherently modular organization of PD processes;[5] that is, the organization of the PD process in clusters that contain most, if not all, of the interactions internally and the interactions or links between separate clusters is eliminated or minimized, see Alexander (1964), Braha and Maimon (1998), Yassine and Braha (2003). The dynamic models developed in Braha and Bar-Yam (2004a) show that a speed up of the PD convergence to the design solution is obtained by reducing or 'ignoring' some of the task dependencies (e.g. eliminating some of the arcs in the corresponding PD network). A modular architecture of the PD process is aligned with this strategy.

In-degree and out-degree distributions

We 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 Figure 3).[6] For all four networks, we find that the in-degree and out-degree distributions can be described by power-laws ('scale-free'
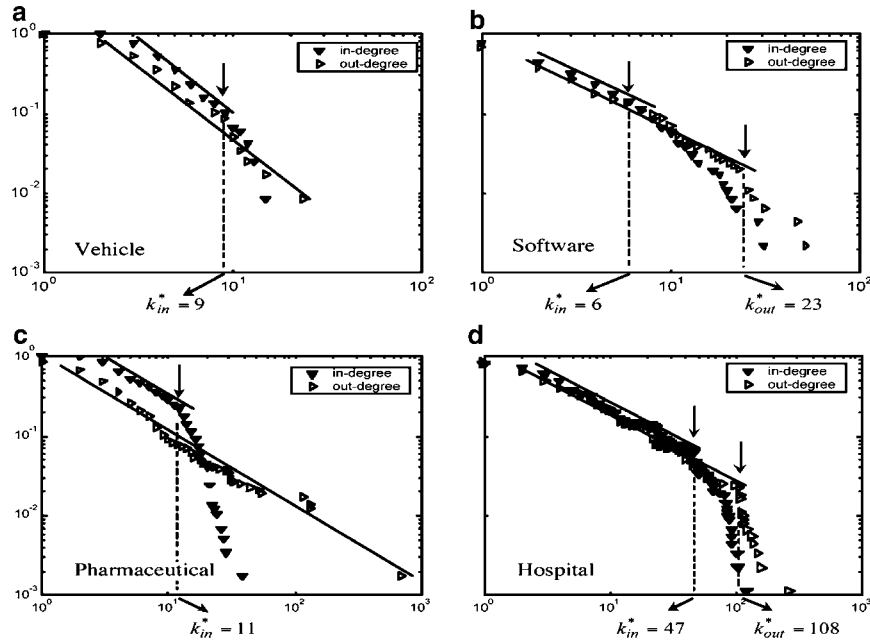
Figure 3 Degree distributions for four distributed problem-solving networks. The log–log plots of the cumulative distributions of incoming and outgoing links show a power-law regime (Pearson coefficient $R > 0.98$, $P > 0.001$) with or without a fast decaying tail in all cases. The in-degree distribution has a lower best visual 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 $\gamma_{vehicle}^{in} \approx 2.82 \pm 0.25$ and $\gamma_{vehicle}^{out} = 2.97 \pm 0.24$ denote the exponents of the associated probability density functions. (b) Software development with 466 tasks and 1245 arcs, where $\gamma_{software}^{in} \approx 2.08 \pm 0.13$ and $\gamma_{software}^{out} \approx 2.25 \pm 0.15$. (c) Pharmaceutical facility development with 582 tasks and 4123 arcs, where $\gamma_{pharmaceutical}^{in} \approx 1.92 \pm 0.07$ and $\gamma_{pharmaceutical}^{out} \approx 1.96 \pm 0.07$. (d) Hospital facility development with 889 tasks and 8178 arcs, where $\gamma_{hospital}^{in} \approx 1.8 \pm 0.03$ and $\gamma_{hospital}^{out} \approx 1.95 \pm 0.03$.

property) with cutoffs introduced at some characteristic scale k^*; $k^{-\gamma} f(k/k^*)$ (typically the function f corresponds to exponential or Gaussian distributions). More specifically, we find scaling regimes (i.e., straight-line regimes) for both $P_{in}(k)$ and $P_{out}(k)$, however, the cutoff k^* occurs lower (by more than a factor of two) for $P_{in}(k)$ than for $P_{out}(k)$. The 'scale-free' property suggests that complex PD task networks are dominated by a few *highly central* tasks. This is in contrast to the bell-shaped Poisson distribution of random graphs, where each node has approximately the same number of links (in such a homogeneous network each node *equally* affects the network behavior). The 'failure' (e.g., excessive rework, lack of integration ability, or delays) of central PD tasks will likely affect the *vulnerability* of the *overall* PD process. Focusing engineering efforts and resources (e.g., funding and technology support) as well as developing appropriate control and management strategies for central PD tasks will likely maintain the sustainability and improve the performance of the PD process.

The edge directedness of the task networks enables us to study the relationships between the in-degrees and out-degrees of tasks. Thus, for example, we are interested in questions such as 'Do tasks with high in-degree also have high out-degree?' or 'Do tasks with high in-degree have small out-degree?'. We address these questions by analyzing the functional relationship between the in-degree and out-degree of tasks.

Interestingly enough it turns out that to a large extent, when considering product development networks, the results reveal almost no correlation between the in-degrees

of tasks and their out-degrees; that is, there are tasks that have a small in-degree but yet have a large out-degree, and *vice versa*. To illustrate this finding, we listed the top 10 tasks of the vehicle development network at General Motors' Research & Development Center ranked according to their in-degree and out-degree centrality measures. We have noticed that only two out of the 10 tasks appear both in the in-degree ranking and in the out-degree ranking. This finding implies that, generally, there is a clear distinction between large-scale generators of information ('Information-Generators' with high out-degree) and large-scale consumers ('Information-Receivers' with high in-degree); a high generator of information could be a low consumer and *vice versa*. This further suggests that a distinction has to be made between in- and out-centrality as far as control and management strategies are concerned. Moreover, those tasks that have both high in- and out-centrality (e.g., 'track total vehicle issues' at General Motors' vehicle design) should be carefully protected and maintained against uncertain disturbances during the PD process.

The presence of cutoffs in node degree distributions has been attributed to physical costs of adding links and limited capacity of a node (Amaral *et al.*, 2000). Such networks may also arise if network formation occurs under conditions of preferential attachment with limited information (Mossa *et al.*, 2002). As previously noted (Amaral *et al.*, 2000; Mossa *et al.*, 2002), the limited capacity of a node, or limited information-processing capability of a node are similar to the so-called 'bounded rationality' concept of Simon (1998).

We find that there is an *asymmetry* between the distributions of incoming and outgoing information flows. The narrower power-law regime suggests that the costs of adding incoming links and limited in-degree capacity of a task are higher than their counterpart out-degree links. We note that this is consistent with the realization that bounded rationality applies to incoming information, and to outgoing information only when it is different for each recipient, not when it is duplicated. This naturally leads to a weaker restriction on the out-degree distribution.

An additional functional significance of the asymmetric topology can be attributed to the distinct roles of incoming and outgoing links in distributed PD processes. 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 PD 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 PD process, and thus increases the likelihood of converging to a successful solution. This empirical observation is found to be consistent with the dynamic PD model developed in Braha and Bar-Yam (2004a). There it was shown that additional rework might slow down the PD convergence or have a destabilizing effect on the system's behavior. As a general rule, the rate of problem solving has to be measured and controlled such that the total number of design problems being created is smaller than the total number of design problems being solved.

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 (or product integrators). 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. Product integrators put the separate development tasks together to ensure fit and functionality. Since late changes in product design are highly expensive, product integrators continuously check unfinished component designs and provide feedback to a large number of tasks accordingly.

Measures of PD centrality using social network analysis

In order to analyze and understand the roles of nodes in networks, many network-analytic studies in the social sciences have relied on the evaluation of centrality measures defined for each node in the network (Wasserman and Faust, 1999). These measures have been given to rank the individual node's prominence according to its position in the network.

Social network analysis (SNA, Wasserman and Faust, 1999) has been utilized for the first time by Braha (2003) to analyze PD networks by incorporating a variety of centrality measures beyond the in- and out-degree centrality measures presented above. In SNA, measures of centrality have been developed at the level of individual nodes ('node centrality') as well as at the level of the whole network ('network centrality'). Also, when analyzing

directed networks (as in PD), a distinction is made between centrality for incoming arcs and centrality for outgoing arcs.

In Braha (2003), the following centrality measures have been found to be most important in the context of PD:

- *In- and out-degree centrality*: 'Degree centrality' measures the activity of a node in terms of the number of arcs coming into, or going out of a node. 'In-degree centrality' measures the extent of influence that one node has on other nodes; while 'out-degree centrality' measures the extent of support needed from others.
- *In- and out-closeness centrality*: 'Closeness centrality' measures how easily accessible the node is to all other nodes. In other words, a node has high closeness centrality if it has very short communication paths to the others. 'In-closeness centrality' is measured as a function of the minimum geodesic distance from all other nodes to the selected node; while 'out-closeness centrality' is measured as a function of the minimum geodesic distance linking that node to the other nodes. While degree centrality measures use only direct and local connectivity information, closeness centrality measures are more global in the sense of taking into account also indirect connectivity information.
- *Betweenness centrality*: 'Betweenness centrality' measures the extent to which a node lies on the paths between others. The betweenness centrality of a node is defined to be the fraction of shortest paths between pairs of nodes in a network in which the node takes part. High betweennees centrality scores indicate a high ratio of shortest paths a node lies on. Betweenness centrality can also be regarded as a measure of the extent to which a node has control over information flowing between others. In a network in which flow is entirely or at least mostly along geodesic paths, the betweenness of a node measures how much flow will pass through that particular node indicating the node's prominence according to its position in the network.

The above centrality measures have been applied for each of the complex PD networks presented above (Braha, 2003). More specifically, for each large-scale PD network, the various centrality measures have been determined for each node. Braha (2003) has computed the Pearson correlation coefficients between the various measures, as illustrated in Tables 2 and 3, for the Vehicle and Pharmaceutical Facility development networks.

Tables 2 and 3 reveal three important aspects of complex PD networks: (1) Generally, there is a clear distinction between large-scale generators of information (i.e. with high out-degree) and large-scale consumers (i.e. with high in-degree); a high generator of information could be a low consumer and *vice versa*. (2) The high correlation between the degree centrality measures and the closeness centrality measures suggests that the direct connections of a task provides useful information regarding the indirect connections of that task. (3) While there is a significant correlation between the degree and closeness centrality measures, the betweenness centrality measure seems to be less correlated with either measure. This suggests that complex PD networks involve special nodes, called 'Information-Brokers,'

Table 2 Pearson correlation coefficients between various centrality measures for the Vehicle development network

	In-degree	*Out-degree*	*In-closeness*	*Out-closeness*	*Betweeness*
In-degree	1				
Out-degree	0.171982668	1			
In-closeness	0.633377641	−0.100389605	1		
Out-closeness	0.106048961	0.770368778	−0.240491453	1	
Betweeness	0.491792093	0.253367214	0.401053689	0.250013872	1

Table 3 Pearson correlation coefficients between various centrality measures for the Pharmaceutical Facility development network

	In-degree	*Out-degree*	*In-closeness*	*Out-closeness*	*Betweeness*
In-degree	1				
Out-degree	0.101454218	1			
In-closeness	0.668625258	−0.011807343	1		
Out-closeness	0.242899356	0.548082559	0.087212956	1	
Betweeness	0.327432976	0.321065057	0.222732467	0.377755663	1

which potentially might have some control over the interactions between any two non-adjacent tasks in the network. These tasks may not be necessarily highly information-generators nor highly information-receivers. Still, 'failures' (e.g., excessive rework, lack of integration ability, or delays) of 'Information-Brokers' tasks will likely affect the *vulnerability* of the *overall* PD process.

Conclusions

The study of complex network topologies across many fields of science and technology has become a rapidly advancing area of research in the last few years, see Strogatz (2001), Albert and Barabási (2002), Newman (2003). One of the key areas of research is understanding the network properties that are optimized by specific network architectures (Amaral *et al.*, 2000; Cancho and Solé, 2001; Mossa *et al.*, 2002; Valverde *et al.*, 2002; Shargel *et al.*, 2003). Here we analyzed the statistical properties of real-world networks of people engaged in product development activities. We show that,

- Complex PD networks display similar statistical patterns to other real-world networks of different origins;
- 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 an *inhomogeneous distribution* of nodal centrality measures (e.g., degree, closeness, or betweenness centrality measures);
- PD task networks are *dominated* by a *few highly central* tasks;
- focusing engineering and management efforts on central PD tasks will likely improve the *performance* of the overall PD process;
- PD tasks can be classified into three major categories: 'Information-Receivers,' 'Information-Generators,' and 'Information-Brokers.'
- 'failure' of central PD tasks affects the *vulnerability* of the overall PD process;

- 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 cutoffs observed in the in-degree and out-degree distributions might reflect Herbert Simon's notion of bounded rationality, and its extension to *group-level* information processing.

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 (Clark, 1989). An important tradeoff exists in many high-technology industries between minimizing time-to-market and development costs and maximizing the product performance. Considering the PD task network, accelerating the PD process can be achieved by 'cutting out' some of the links between the tasks. Although the elimination of some arcs should result in a speed up of the 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, that is, $p(k) = k^{-\gamma} f(k/k^*)$ (Cancho and Solé, 2001; Valverde *et al.*, 2002). 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 (Eppinger *et al.*, 1994). 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 by Ferrer *et al.* (2001) and software architectures by Valverde *et al.* (2002)) 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 (Braha and Maimon, 1998). 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 PD project by exploiting the knowledge embedded in reused modules, and thus significantly reduce the PD time.

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. When both distributions have cutoffs, the incoming distribution has a cutoff that is significantly lower, in the cases studied by more than a factor of two. From a PD viewpoint, the functional significance of this asymmetric topology has been explained by considering a bounded-rationality argument originally put forward by Simon (1998) in the context of human interactions. 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 cutoff 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 would be interesting to see whether this property can be found more generally in other directed human or non-human networks. It seems reasonable to propose that the asymmetric link distribution is likely to hold for such networks when nodes represent information processing/using elements.

Notes

1 Denker, S., private communication; Available at http://necsi.org/projects/braha/largescaleengineering.html.

2 Newton, A. and Austin, S. private communication; Available at http://necsi.org/projects/braha/largescaleengineering.html.

3 Cividanes, A., private communication. See also Cividanes, A. MSc Thesis, Mechanical Engineering Department, Massachusetts Institute of Technology, 2002. A complete description of the tasks, the list of interviewees, and the result of the survey are available at http://necsi.org/projects/braha/largescaleengineering.html. For further details regarding the data collection process at GM's Research & Development Center see Cividanes's thesis.

4 For a detailed description of data flow and design-process model diagrams see Austin, S., Baldwin, A., Li, B. and Waskett, P. (1999). 'Analytical Design Planning Technique: A Model of the Detailed Building Design Process,' *Design Studies* **20**(3): 279–296.

5 Actually, in Braha (2003) and Braha and Bar-Yam (2004a) it has been shown that the system-level structure of complex organizational networks is best approximated by a *hierarchical network organization* with *seamlessly nested modularity*. In contrast to current intuitive views of modularity, which assume the coexistence of relatively independent groups of nodes, real-world networks have an inherent self-similar property: There are many highly integrated small modules, which group into a few larger modules, that in turn can be integrated into even larger modules.

6 Note that a power-law distribution of the in-degree distribution (respectively, the out-degree distribution) $p_{in}(k) \sim k^{-\gamma_{in}}$ with exponent γ_{in} translates into a power-law distribution of the cumulative probability distribution $P_{in}(k) \sum_{k\prime=k}^{\infty} k\prime^{-\gamma_{in}} \sim k^{-(\gamma_{in}-1)}$ with exponent $\gamma_{in}-1$.

References

Albert, R. and Barabási, A.-L. (2002). Statistical Mechanics of Complex Networks, *Reviews of Modern Physics* **74**: 47–97.

Albert, R., Jeong, H. and Barabási, A.-L. (1999). Diameter of the World Wide Web, *Nature* **401**: 130–131.

Albert, R., Jeong, H. and Barabási, A.-L. (2000). Error and Attack Tolerance in Complex Networks, *Nature* **406**: 378–382.

Alexander, C. (1964). *Notes on the Synthesis of Form*, Cambridge, MA: Harvard University Press.

Amaral, L.A.N., Scala, A., Barthélémy, M. and Stanley, H.E. (2000). Classes of Small-World Networks, *Proceedings of National Academy of Sciences of the United States of America* **97**: 11149–11152.

Austin, S., Baldwin, A., Li, B. and Waskett, P. (2000). Integrating Design in the Project Process, *Proceedings of the Institution of Civil Engineers* **138**(4): 177–182.

Barabási, A.-L. and Albert, R. (1999). Emergence of Scaling in Random Networks, *Science* **286**: 509–512.

Braha, D. (2003). Socio-Technical Complex Networks, Lecture Notes, Tutorial presented at the 2003 ASME International Design Engineering Technical Conferences, Chicago, Illinois.

Braha, D. and Bar-Yam, Y. (2004a). The Statistical Mechanics of Product Development, Technical Report, NECSI Technical Report 2004-09-01.

Braha, D. and Bar-Yam, Y. (2004b). Topology of Large-Scale Engineering Problem-Solving Networks, *Physical Review E* **69**: 016113.

Braha, D. and Maimon, O. (1998). *A Mathematical Theory of Design: Foundations, Algorithms, and Applications*, Boston, MA: Kluwer Academic Publishers.

Cancho, R.F. and Solé, R.V. (2001). SFI Working Paper 01-11-068.

Clark, K.B. (1989). Project Scope and Project Performance: The Effect of Parts Strategy and Supplier Involvement on Product Development, *Management Science* **35**(10): 1247–1263.

Cross, R., Borgatti, S.P. and Parker, A. (2002). Making Invisible Work Visible: Using Social Network Analysis to Support Strategic Collaboration, *California Management Review* **44**(2): 25–46.

Eppinger, S.D., Whitney, D.E., Smith, R.P. and Gebala, D.A. (1994). A Model-Based Method for Organizing Tasks in Product Development, *Research in Engineering Design* **6**(1): 1–13.

Erdös, P. and Rényi, A. (1959). On random graphs, *Publicationes Mathematicae* **6**: 290–297.

Faloutsos, M., Faloutsos, P. and Faloutsos, C. (1999). On Power-Law Relationships of the Internet Topology, *Computer Communication Review* **29**: 251–262.

Ferrer, R., Janssen, C. and Solé, R.V. (2001). Topology of Technology Graphs: Small World Patterns in Electronic Circuits, *Physical Review E* **63**: 32767.

Jeong, H., Mason, S., Barabási, A.-L. and Oltvai, Z.N. (2001). Lethality and Centrality in Protein Networks, *Nature* **411**: 41–42.

Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N. and Barabási, A.-L. (2000). The Large-Scale Organization of Metabolic Networks, *Nature* **407**: 651–654.

Kambil, A. and van Heck, E. (2002). *Making Markets: How Firms Can Design and Profit from Online Auctions and Exchanges*, Cambridge, MA: Harvard University Press.

Klein, M., Sayama, H., Faratin, P. and Bar-Yam, Y. (2003). The Dynamics of Collaborative Design: Insights from Complex Systems and Negotiation Research, *Concurrent Engineering* **11**(3): 201–209.

Montoya, J.M. and Solé, R.V. (2002). Small World patterns in Food Webs, *Journal of Theoritical Biology* **214**: 405–412.

Mossa, S., Barthélémy, M., Stanley, H.E. and Amaral, L.A.N. (2002). Truncation of Power Law Behavior in 'Scale-Free' Network Models due to Information Filtering, *Physical Review Letters* **88**: 138701.

Newman, M.E.J. (2001a). Scientific Collaboration Networks. I. Network Construction and Fundamental Results, *Physical Review E* **64**: 016131.

Newman, M.E.J. (2001b). Scientific Collaboration Networks. II. Shortest Paths, Weighted Networks, and Centrality, *Physical Review E* **64**: 016132.

Newman, M.E.J. (2001c). The Structure of Scientific Collaboration Networks, *Proceedings of National Academy of Sciences of the United States of America* **98**: 404–409.

Newman, M.E.J. (2003). The Structure and Function of Complex Networks, *SIAM Review* **45**: 167–256.

Nishiguchi, T. and Beaudet, A. (1998). The Toyota Group and the Aisin Fire, *Sloan Management Review* **40**(1): 49–59.

Osborne, S.M. (1993). Product Development Cycle Time Characterization Through Modeling of Process Iteration, MSc Thesis, Massachusetts Institute of Technology.

Price, S. (1965). Networks of Scientific Papers, *Science* **149**: 510–515.

Reitman, V. (1997). Toyota's Fast Rebound, *Wall Street Journal*, May 8, 1997.

Shargel, B., Sayama, H., Epstein, I.R. and Bar-Yam, Y. (2003). Optimization of Robustness and Connectivity in Complex Networks, *Physical Review Letters* **90**(6): 068701.

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

Steward, D.V. (1981). The Design Structure System: A Method for Managing the Design of Complex Systems, *IEEE Transactions on Engineering Management* **28**: 71–74.

Strogatz, S.H. (2001). Exploring Complex Networks, *Nature* **410**: 268–276.

Valverde, S., Cancho, R.F. and Solé, R.V. (2002). Scale Free Networks from Optimal Design, *Europhysics Letters* **60**: 512–517.

Watts, D.J. and Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks, *Nature* **393**: 440–442.

Yassine, A. and Braha, D. (2003). Complex Concurrent Engineering and the Design Structure Matrix Method, *Concurrent Engineering* **11**(3): 165–176.

Yassine, A., Joglekar, N., Braha, D., Eppinger, S. and Whitney, D. (2003). Information Hiding in Product Development: The Design Churn Effect, *Research in Engineering Design* **14**(3): 131–144.

Wasserman, S. and Faust, K. (1999). *Social Network Analysis*, Cambridge: Cambridge University Press.

About the authors

Dan Braha is an Associate Professor at the Earle P Charlton College of Business of the University of Massachusetts, Dartmouth. He is also an affiliate of the New England Complex Systems Institute (NECSI) in Cambridge, MA. Prior to that he was a Visiting Professor at the Massachusetts Institute of Technology (MIT), and a researcher at Boston University. Dan Braha has published extensively in various prestigious journals, including a book on the foundations of design with Kluwer, a book on data mining in design and manufacturing with Kluwer, and an edited book on the emerging science of Complex Engineered Systems with Perseus Publishers (forthcoming in 2004). He serves as an Associate Editor for *IEEE Transactions on Systems, Man, and Cybernetics*, as well as a member of the editorial board of *AI EDAM* (Cambridge University Press). Professor Braha was the editor of several special journal issues, and has served on executive committees and as chair at a number of international conferences. Currently, he aims to advance the understanding of Complex Engineered Systems (CES), arrive at their formal analysis, as well as facilitate their application.

Yaneer Bar-Yam received the PhD degree in physics from Massachusetts Institute of Technology (MIT). He is the founder and President of the New England Complex Systems Institute (NECSI), and is the Chairman of the annual International Conference on Complex Systems. He is the author of over 100 published articles (including in Nature and Science magazines) and several books including the textbook Dynamics of Complex Systems (Addison-Wesley, 1997). He has held positions at the IBM Research Center, Weizmann Institute of Science, and Associate Professor at Boston University. He left Boston University in 1997 to become President of the New England Complex Systems Institute. Professor Bar-Yam uses complex systems concepts to understand how organizations and patterns of behavior arise, evolve, adapt, and how we can use multi-scale representations to relate fine and large-scale, short- and long-term perspectives. He has studied the role of complexity in corporate organization and effectiveness, in health-care and health-care management, in the education system, in military conflict, and in third world development. Applications are to the relationship of structure and function and meeting complex challenges at all scales.