Version 4/3/07, 1:30 PM

The Architecture of Complexity

From network structure to human dynamics

Albert-László Barabási

We are surrounded by complex systems, from cells, made of thousands of different molecules that seamlessly work together, to society, a collection of billions of interacting individuals that displays signatures of order and self-organization. Understanding and quantifying this complexity is a grand challenge for science in general. Kinetic theory set the stage at the end of the nineteenth century, demonstrating that the measurable properties of gases, from pressure to temperature, can be reduced to the random motion of atoms and molecules. In the 1960s and 70s, researchers developed systematic approaches to quantifying the transition from disorder to order in material systems, such as magnets and liquids. Chaos theory, with its message that complex and unpredictable behavior can emerge from the nonlinear interactions of a few components, dominated the quest to understand complex behavior in the 1980s. The 1990s were the decade of fractals, quantifying the geometry of patterns emerging in self-organized systems, from leaves to snowflakes.

Despite these conceptual advances, a complete *theory of complexity* does not yet exist. When trying to characterize complex systems, the available tools fail for increasingly obvious reasons. First, most complex systems are not made of identical indistinguishable components, as gases or magnets are. Rather, each gene in a cell or each individual in society has its own characteristic behavior. Second, while the interactions among the components are manifestly nonlinear, truly chaotic behavior is more the exception than the rule. Third, and most important, molecules and people do not obey either the extreme disorder of gases, where any molecule can collide with any other molecule, or the extreme order of magnets, where spins interact only with their immediate neighbors in a periodic lattice. Rather, in complex systems, the interactions form exquisite *networks*, where each node interacts only with a small number of selected partners, but whose presence and effects may be felt by nodes far away.

Networks are everywhere and at every scale. The brain is a network of nerve cells connected by axons, while cells are networks of molecules connected by biochemical reactions. Societies, too, are networks of people linked by friendship, family relationships, and professional ties. On a larger scale, food webs and ecosystems can be represented as networks of species. Furthermore, networks pervade technology; examples include the Internet, power grids, and transportation systems. Even the language used to convey thoughts is a network of words connected by syntactic relationships.

Despite the pervasiveness of networks, however, their structure and properties are not yet fully understood. How do the interactions of several malfunctioning genes in a complex genetic network lead to cancer? How does diffusion occur so rapidly through certain social and communications networks, leading to epidemics of diseases and computer viruses such as the *Love Bug*? And how do some networks continue to function despite the failure of the vast majority of their nodes?

Recent research is beginning to answer such questions [1]-[9]. Over the past few years, scientists have discovered that complex networks have an underlying architecture guided by universal principles. For instance, many networks, from the World Wide Web to the cell's metabolic system, to the actors of Hollywood, are dominated by a small number of nodes that are highly connected to other nodes. These important nodes, called *hubs*, greatly affect a network's overall behavior. As described in this arcticle, hubs make the network robust against accidental failures, but vulnerable to coordinated attacks.

The purpose of this article is to illustrate, through the example of human dynamics, that a thorough understanding of complex systems requires an understanding of network dynamics as well as network topology and architecture. After an overview of the topology of complex networks such as the Internet and the World Wide Web, data-driven models for human dynamics are given. These models motivate the study of network dynamics and postulate that a complete theory of complexity must incorporate the interactions between dynamics and structure. The article also advances the notion that understanding network dynamics is being made possible with the availability of large data sets, and the analysis tools gained from the study of networks structure.

The Random Network Paradigm

Complex networks were originally thought of as being completely random. This paradigm has its roots in the work of Paul Erdős and Alfréd Rényi, who, in 1959, aiming to describe networks in communications and life sciences, suggested that networks be modeled as random graphs [10],[11]. Their approach was to take N nodes and connect them by L randomly placed links. The simplicity of the model and the elegance of the theory revitalized graph theory, leading to the emergence of random networks as a mathematical field of study [10]-[12].

A key prediction of random network theory is that despite the random placement of links, the resulting network is democratic since most nodes are assigned approximately the same number of links. Indeed, in a random network the nodes follow a bell-shaped Poisson distribution. It is therefore rare to find nodes that have significantly greater or fewer number of links than a randomly chosen node. Random networks are also called exponential networks because the probability that a node is connected to k other nodes decreases exponentially for large k (Figure 1). The Erdős-Rényi model, however, raises the question as to whether networks observed in nature are truly random. Could the Internet for example, offer us the fast and seamless service if computers were randomly connected to each other? Or, to carry the analysis even further, could you read this article if the chemicals in your body suddenly decided to react randomly with each other, bypassing the rigid chemical web they normally obey? Intuitively the answer is no, since we suspect that behind each complex system there is an underlying network with nonrandom topology. The challenge of network structure studies, however, is to unearth the signatures of order from the collection of millions of nodes and links that form a complex network.

The World Wide Web and the Internet as Complex Networks

The World Wide Web (WWW) contains over a billion documents (web pages), which represent the nodes of a complex web. These documents are connected by Uniform Resource Locators (URLs) which are used to navigate from one document to another (Figure 2a). To analyze the World Wide Web's properties, we need to map how web pages link to each other. This information is routinely collected by search engines, such as Google and Alta Vista. Given, however, that search engines are reluctant to share their maps for research purposes, an independent map of the WWW was needed [13]. A robot, or a *web crawler,* was designed to start from a given webpage and collect the entire page's outgoing links. The robot then follows each outgoing link to visit more pages, collecting their respective outgoing links, and so on [13]. Through this iterative process a small but representative fraction of the WWW was mapped out.

Since the WWW is a directed network, each document is characterized by the number k_{out} of its outgoing links and the number k_{in} of its incoming links. The outgoing (incoming) degree distribution thus represents the probability $P(k)$ that a randomly selected webpage has exactly k_{out} (k_{in}) links. Relying on random graph theory, $P(k)$ was expected to follow a Poisson distribution. The collected data, however, indicated that $P(k)$ actually follows a power-law distribution as shown in Figure 2c and described by

$$P(k) \sim k^{-\gamma}, \tag{1}$$

where $\gamma_{out} \cong 2.45$ $(\gamma_{in} \cong 2.1)$.

As illustrated in Figure 1, major topological differences exist between a network with a Poisson connectivity distribution and one with a power-law connectivity distribution. Indeed, most nodes in an undirected random network have approximately the same number of links, $k \approx \langle k \rangle$, where $<k>$ represents the average degree. The exponential decay of Poisson distribution $P(k)$ thus guarantees the absence of nodes with significantly more links than $\langle k \rangle$ and imposes a natural scale in the network. In contrast, the power-law distribution implies that nodes with few links are abundant, but that a small number of nodes have a large number of links. A map of our highway system for example, where cities are nodes and highways are links, illustrates an exponential network. Most cities are located at the intersection of two to five highways. On the other hand, a scale-free network is similar to the airline routing maps displayed in flight magazines. While most airports are served by few carriers, few hubs such as Chicago or Frankfurt have links to almost all other U.S. or European airports, respectively. Thus, just like the smaller airports, the majority of WWW documents have a small number of links, and while these links are not sufficient by themselves to ensure that the network is fully connected, the few highly connected hubs guarantee that the WWW is held together.

Unlike Poisson distributions, a power-law distribution does not possess an intrinsic scale and its average degree, $<k>$, does not convey much information about the network structure. The absence of an intrinsic scale in k in networks with power-law degree distribution prompted researchers to denote scale-free networks [14]. A scale-free network is therefore a network whose degree distribution obeys a power law. Empirical measurements, however, indicate that real networks deviate from simple power law behavior. The most common deviation is the flattening of the degree distribution at small k values, while a less common one is the exponential cutoff for high values of k. Thus, a proper fit to the degree distribution of real networks has the form $P(k) \sim (k+k_0)^{-\gamma}\exp(-k/k_x)$, where k_0 is the small degree cutoff and k_x is the lengthscale of the high degree exponential cutoff. Scale-free behavior is therefore evident only between k_0 and k_x.

The scale-free topology of the WWW has led researchers to look for inhomogeneous topologies in other complex systems such as the Internet. Unlike the WWW, the Internet is a physical network, whose nodes are routers and domains, and whose links are the phone lines and optical cables that connect the nodes together (Figure 2b). Due to its physical nature, the Internet was expected to be structurally different from the WWW, where adding a link to an arbitrary remote web page is as easy as linking to one on a computer in the next room. The Internet network however, also appears to follow a power law degree distribution as first noticed in [15] where the authors analyzed the Internet topology at the router and domain levels (see Figure 2b). The degree distribution was shown to follow a power law with an exponent $\gamma=2.5$ for the router network and $\gamma=2.2$ for the domain map, which indicate that the wiring of the Internet is also dominated by several highly connected hubs [15].

19 degrees of separation

Stanley Milgram, a Harvard sociologist, surprised the world in 1967 with a bold claim: any two people are typically five to six handshakes away from each other [16]. That is, in spite of having more than six billion inhabitants of our planet, we live in a *small-world*. This feature of social networks is known as the *six-degrees of separation* property [17]. In addition, sociologists have repeatedly argued that nodes in social networks are grouped in small clusters. These clusters represent circles of friends and acquaintances, within each a node is connected to all other nodes, but with only sparse links to the outside world [18]. While the existence of such local clustering and small world behavior agrees with our intuition about social networks, these features were not expected to be relevant beyond social systems. Nevertheless, the question soon arose as whether the Internet and the WWW follow the small-world paradigm.

In order to answer this question for the WWW, a complete map of the WWW network is needed. As found in [19],[20] however, the most powerful search engines are only able to cover about 16% of the web. When faced with the problem of inferring properties of a large population from a finite sample, physicists routinely use the tools of statistical mechanics. In this case, small computer models of the WWW were constructed, making sure that the link distribution matches the measured functional form

[13]. The shortest distances between any two nodes were then identified and averaged over all node pairs in order to obtain the average node separation *d*. Repeating this process for networks of different sizes using *finite size scaling*, a standard procedure of statistical mechanics, it was inferred that $d = 0.35 + 2.06.log(N)$, where *N* is the number of WWW nodes. When this model was completed in 1999, the WWW had 800 million nodes. The typical shortest path between two randomly selected pages is thus around 19 assuming that such a path exists, which is not always guaranteed due to the web's directed nature. An extensive study by an IBM-Compaq-AltaVista collaboration [21] empirically found that for 200 million nodes this distance is 16, as opposed to 17 as predicted in [13].

These results clearly indicate that the WWW represents a small world, and that the typical number of clicks between two webpages is around 19, despite the presence today of over one billion online pages. Moreover, the WWW displays a high degree of clustering as shown in [22]. That is, the probability that two neighbors of a given node are also linked together is much greater than the value expected for a random network. Finally, results reported in [1] indicate that the Internet also possess a small world structure.

Evolving networks

Why do such different systems as the physical network of the Internet and the virtual web of the WWW develop similar scale-free structures? The emergence of the power-law degree distribution may be traced back to two common mechanisms. These mechanisms are absent from the classical random graph models but present in various complex networks [14]. First, traditional graph theoretic models assume that the number of nodes in a network is fixed. In contrast, the WWW continuously expands by adding new webpages, and the Internet grows with the installation of new routers, computers, and communication links. Second, while random graph models assume that the links are randomly distributed, most real networks exhibit *preferential attachment*. Indeed, we are more likely to link our webpage to the highly connected documents on the WWW, since we are likely aware of the existence of such documents. Network engineers also tend to connect their institution's computers to the Internet through high bandwidth nodes, which inevitably attract a large number of other consumers and links.

Based on the increasing number of nodes as well as on preferential attachment, a simple model in which a new node is added to the network at each time step was constructed in [14]. The new node is then linked to some of the nodes already present in the system (Figure 3). The probability $\Pi(k)$ that a new node connects to a node with k links follows a preferential attachment rule such as

$$\Pi(k)=\frac{k}{\Sigma_i k_i} \tag{3}$$

where the sum is over all nodes in the network. Numerical simulations indicate that the resulting network is indeed scale-free, and the probability that a node has k links follows equation (1) with exponent $\gamma = 3$ [14]. The power law nature of the distribution is predicted by a rate equation-based approach, [23]-[25], as well as from an exact solution of the scale-free model introduced above [26]. This simple model illustrates how growth and preferential attachment jointly lead to the appearance of the hub hierarchy which exemplifies the scale-free structure. A node with more links increases its connectivity faster than nodes with fewer links, since incoming nodes tend to connect to it with higher probability as described in (3). This model leads to a *rich-gets-richer* phenomenon that is evident in some competitive systems.

Bose-Einstein condensation

In most complex systems, nodes vary in their ability to compete for links. Some websites for example, through a mix of good content and marketing, tend to quickly acquire a large number of links, and to become more popular than sites that have been around much longer. An example of this phenomenon is the search engine Google. Although a latecomer to the field of web searching, Google became one of the most connected nodes of the WWW in less than two years. In order to incorporate this competitive advantage into the scale-free model, each node is provided with a fitness measure that describes the node's ability to compete for links at the expense of other nodes [27],[28]. A randomly chosen fitness η_i is assigned to each node i to modify the growth rate from equation (3) to

$$\Pi(k_i) = \frac{\eta_i k_i}{\Sigma_j \eta_j k_j}. \qquad (4)$$

The competition generated by the fitness measure leads to *multiscaling*. The connectivity of a given node follows $k_i(t) \simeq t^{\beta(\eta)}$, where $\beta(\eta)$ increases with increasing η. This allows fitter nodes with larger η to join the network at a later time, and yet to become more connected than older but less fit nodes.

The competitive fitness models have close ties to Bose-Einstein condensation, currently one of the most investigated problems in condensed matter physics. Indeed, it was found in [28] that the fitness model can be mapped into a Bose gas by replacing each node with an energy level $\varepsilon_i = e^{-\beta\eta i}$. According to this mapping, links connected to node i are replaced by particles on level ε_i, and the behavior of the Bose gas is uniquely determined by the distribution $g(\varepsilon)$ from which the fitness measures are selected. The functional form of $g(\varepsilon)$ is expected to be system-dependent. Although, the attractiveness of a router to a network engineer is probably generated from a different distribution than that of the fitness of a company competing for customers, a *fits-gets-richer* phenomenon emerges for a wide class of fitness distributions. In such cases, no clear winner is evident even though the fittest node acquires more links than its less fit counterparts. On the other hand, certain fitness distributions may result in Bose-Einstein condensation which corresponds to a *winner-takes-all* phenomenon. The fittest node in a network emerges then as a clear winner, developing a *condensate* by acquiring a (large) finite fraction of the links, independent of the size of the system.

The Achilles' heel of the Internet

As the world economy becomes increasingly dependent on the Internet, a concern arises on whether the Internet's functionality can be maintained under inevitable failures and frequent hacker attacks. The Internet has so far proven remarkably resilient against failures. Even though around 3% of the routers may be down at a particular moment, we rarely observe major Internet disruptions. How did the Internet come to be so robust? While significant error tolerance is built into the protocols that govern packet switching communications, we have recently learned that the scale-free topology of the Internet also plays a crucial role in making it more robust.

Percolation concepts provide one approach to understanding the scale-free induced error tolerance of the Internet. Percolation theory specifies that the random removal of nodes from a network results in an inverse percolation transition. If a critical fraction f_c of nodes is removed, the network fragments into tiny, non-communicating islands of nodes. However, simulations of scale-free networks did not support percolation's theory prediction [29]. Even though up to 80% of the nodes of a scale-free network were removed, the remaining nodes remained part of a compact cluster. The disagreement was resolved in [30],[31] where it was shown that as long as the connectivity exponent γ in equation (1) is smaller than 3 (which is the case for most real networks, including the Internet) the critical threshold for fragmentation is $f_c=1$. This result clearly demonstrates that scale-free networks cannot be broken into pieces by the random removal of nodes. This extreme robustness to random failures is rooted in the inhomogeneous network topology. Since there are far more *weakly-connected* nodes than hubs, random removal will most likely affect the less connected nodes. The removal of a node with a small degree however, does significant disrupt the network topology, just as the closure of a local airport has little impact on international air traffic.

Their inhomogeneous topology, however, makes scale-free networks especially vulnerable to targeted attacks [29]. Indeed, the removal of a small fraction of the most connected nodes (hubs) may break the network up into pieces. These findings clearly illustrate the underlying topological vulnerability of scale-free networks. In fact, while the Internet is not expected to break under the random failure of routers and links, well-informed hackers can easily handicap the network by targeting hubs for attacks.

Note that while error tolerance and fragility to attacks are consequences of the scale-free property, the reverse is not necessarily true. Networks that are resilient to random attacks but that fragment under targeted attacks are not necessarily scale-free. For example, the hub-and-spoke network, in which all nodes connect to a central node, is the most resilient network to random failures. Such a network fails only when the central hub is removed, an event whose probability of occurrence is $1/N$ for a network with N nodes. It is therefore important to define a scale-free network based on the network's degree distribution rather than its robustness [32].

Scale-Free Epidemics

The structure of scale-free networks may help us understand the spread of computer viruses, diseases, and fads. Diffusion theories, by both epidemiologists and marketing experts, predict the presence of a *critical threshold* for the successful propagation of *something* throughout a population or a network. A virus that is less virulent or a fad that is less contagious than the critical threshold inevitably dies out, while those above the threshold will multiply exponentially, eventually penetrating the entire network.

Recently though, research in [33] found that the critical threshold on a scale-free network is actually zero. Therefore, all viruses, even those that are only weakly contagious, eventually spread and persist in the system. This result explains why *Love Bug*, the most damaging computer virus thus far, having shut down the British Parliament in 2000, was still the seventh most common virus even a year after its supposed eradication. Hubs are again the key to this surprising behavior. In fact, since hubs are highly connected, at least one of them is likely to become infected by a single corrupted node. Moreover, once a hub is infected, it broadcasts the virus to numerous other nodes, eventually compromising other hubs that help spread the virus throughout the entire system.

Because biological viruses spread on scale-free social networks, scientists need to consider the interplay between network topology and epidemics. Specifically, in a scale-free contact network, the traditional public-health approach of random immunization will probably fail because it will likely miss some of the hubs. The topology of scale-free networks suggests an alternative approach. By targeting the hubs, or the most connected individuals, the immunizations become effective after reaching only a small fraction of the population [34]-[36]. Identifying the hubs in a social network, however, is much more difficult than in other types of systems such as the Internet. Nevertheless, a clever solution was proposed in [35]. By immunizing a small fraction of the random acquaintances of randomly selected individuals, the hubs will be immunized with high probability since hubs are by definition acquainted with many people. This approach though, raises ethical dilemmas. In fact, even if the hubs were identified, it is not obvious that they should have priority for immunizations and cures.

In some business contexts, people want to start epidemics rather than stop them. For example, viral marketing campaigns target social hubs in order to spread the adoption of a product as quickly as possible. In the 1950s, a study funded by the pharmaceutical giant Pfizer discovered the role that hubs play by how quickly a community of doctors adopts a new drug. Indeed, marketers have intuitively known for some time, that certain customers are much more effective in spreading promotional buzz about new products and fads. The architecture of scale-free networks may provide the scientific framework and mathematical tools to probe that phenomenon more rigorously.

Human dynamics and the temporal behavior of single nodes

We focused in the previous sections on one aspect of complex networks, namely their topology. While impressive research advances were achieved in this direction, our limited understanding of *network dynamics* continues to haunt us. Indeed, most complex systems of practical interest, from the cell, to the Internet, to social networks, are fascinating because of their overall temporal behavior. While such systems have nontrivial network topologies, the role of their topology is to serve as a skeleton on which the myriad of dynamical processes, from information to material transfer, take place. Topological network theory, while indispensable towards describing these dynamical processes, does not yet fully account for the complex behavior displayed by these systems. We thus need to characterize the dynamical processes taking place on complex networks, and to understand the interplay between topology and dynamics. Most importantly, we need to uncover, should they exist, the generic organizing principles of network dynamics. Advances in understanding network dynamics will probably be empirical and data-driven, as was the case with network topology. Indeed, the scale-free topology was not discovered until extensive datasets became available to help explore large real networks [14]. In order to understand the network dynamics, we probably have to develop tools to carefully monitor the dynamics of real systems, in the hope that some universal properties emerge. In the following, we describe recent advances in the context of human dynamics, since the dynamics of many social, technological, and economic networks are driven by individual human actions. The quantitative understanding of human behavior is therefore a central question of modern science, and one that network research can hardly ignore.

Current models of human dynamics in areas such as risk assessment and communications assume that human actions are randomly distributed in time, and well-approximated by Poisson processes [37]-[39]. In the following, evidence is presented to show that the timing of many human activities, ranging from communication to entertainment and work patterns, follow non-Poisson statistics [40]-[53]. These statistics are characterized by bursts of rapidly occurring events separated by long periods of inactivity. This *bursty* nature of human behavior is a consequence of a decision-based queuing process. When individuals execute tasks based on some perceived priority, the tasks' timing follows a heavy-tailed distribution with most tasks being rapidly executed, while a few tasks experience long waiting times [40],[53]. In contrast, priority-blind execution is well approximated by uniform inter-event statistics.

Humans participate in a large number of distinct daily activities. These activities range from electronic communication, to browsing the web, to initiating financial transactions and engaging in entertainment and sports. Factors ranging from work and sleep patterns, to resource availability and interaction with other individuals determine the timing of each daily human activity. Seeking regularities in human dynamics apart from the obvious daily and seasonal periodicities, does appear difficult if not impossible. A quantitative understanding of network dynamics driven by human activities may, therefore, appear hopeless. We show next that this is not entirely true and that human dynamics is driven by interesting reproducible mechanisms that serve as a starting point towards understanding human network dynamics.

As mentioned earlier, current models of human activity are based on Poisson processes, and assume that in a *dt* time interval, an individual engages in a specific action with probability *qdt*, where *q* is the overall frequency of the monitored activity. According to such models, the time interval between two consecutive actions by the same individual, called the waiting or inter-event time, follows an exponential distribution (Figure 4) [37],[53]. Poisson processes are widely used to quantify the consequences of human actions, such as modeling traffic flow patterns or frequency of accidents [37], call center staffing [38], inventory control [39], or to estimate the number of congestion-caused blocked calls in mobile communications [41], [54]. An increasing number of recent measurements, however, indicate that the timing of many human actions systematically deviate from the Poisson-based prediction. In fact, the waiting or inter-event times are better approximated by a heavy-tailed or Pareto distribution (Figure 4). The differences between Poisson and heavy-tailed behavior are striking. A Poisson distribution decreases exponentially, forcing the consecutive events to follow each other at regular time intervals and forbidding very long waiting times. In contrast, the slowly decaying heavy-tailed processes allow for long periods of inactivity that separate bursts of intense activity (Figure 4).

In order to provide evidence for non-Poisson activity patterns in individual human behavior, consider two human activity patterns: the e-mail-based communication based on a data set capturing the sender, recipient, time and size of each e-mail [42]-[43], and the letter-based communication pattern of Einstein [55]. As Figure 5 shows, the response time distribution of both e-mails and letters is best approximated with

$$P(\tau) \sim \tau^{-\alpha} \qquad (5)$$

where $\alpha = 1$ for e-mail communications, and $\alpha = 3/2$ for letters. These results indicate that an individual's communication pattern has a bursty non-Poisson character. For a short time period, a user sends out several e-mails/letters in quick succession, followed by long periods of no communication.

This behavior is not limited to e-mail and letter-based communications. Measurements capturing the distribution of the time differences between consecutive instant messages sent by individuals during on-line chats [44] show a similar pattern. Professional tasks, such as the timing of job submissions on a supercomputer [45], directory listings and file transfers (FTP) requests initiated by individual users [46], the timing of print jobs submitted by users [47], or the return visits of individual users to a website [48], are also reported to display non-Poisson features. Similar patterns emerge in economic transactions that describe the number of hourly trades in a given security [49], or the time interval distribution between individual trades in currency futures [50]. Finally, heavily-tailed distributions characterize entertainment events, such as the time intervals between consecutive online games played by the same user [51].

The wide range of human activity patterns that follow non-Poisson statistics suggests that the observed bursty character of such patterns reflects some fundamental and potentially generic feature of human dynamics. Next, we describe that the bursty nature of human dynamics is a consequence of a *queuing* process driven by human decision making. Whenever an individual is presented with multiple tasks, and is asked to choose among them based on some perceived priority parameter, the waiting time of the various tasks is Pareto distributed [40],[53]. In contrast, first-come-first-serve as well as random task executions, both common in service-oriented or computer-driven environments, lead to a uniform Poisson-like dynamics.

Most human-initiated events require an individual to prioritize different activities. At the end of each activity, an individual decides what to do next: send an e-mail, shop, or place a phone call, thus allocating time and resources for the chosen activity. Consider then an agent operating with a priority list of L tasks. Once a task is executed, it is removed from the list, and as soon as new tasks emerge, they are added to the list. In order to compare the urgency of the different tasks on the list, the agent or individual assigns a priority parameter x to each task. The waiting time of a given task before it is executed depends on the method used by the agent to choose the order of task execution. In this respect, three selection protocols [54] are relevant for human dynamics:

(i) The simplest protocol is the first-in-first-out protocol, which executes tasks in the order they were added to the list. This protocol is common in service-oriented processes, such as the execution of orders in a restaurant, or in directory assistance and consumer support applications. A task remains on the list before it is executed, as long as it takes to perform all tasks that are ahead of it. If the time required to perform a task is chosen from a bounded distribution (the second moment of the distribution is finite), the waiting-time distribution develops an exponential tail, indicating that most tasks will experience approximately the same waiting time.

(ii) In the second protocol, tasks are executed in a random order, irrespective of their priority, or time spent on the list. This protocol is common, for example, in educational settings where students are called on randomly, and in some Internet packet-routing protocols. In this case, the waiting-time distribution of individual tasks is also exponential.

(iii) In most human-initiated activities, task selection is not random. Individuals usually execute the highest priority items on their list. The resulting execution dynamics is quite different from (i) and (ii). High-priority tasks are executed soon after they are added to the list, while low-priority items must wait until all higher priority tasks are cleared. This protocol forces lower-priority items to stay on the list longer than higher-priority ones. In the following, we show that this selection mechanism, practiced by humans on a daily basis, is a likely source of the heavy tails observed in human-initiated processes. Let us consider two models for this protocol:

Model A [53]: Assume that an individual has a priority list with L tasks, and that each task is assigned a priority parameter $x_i, i = 1...., L$ chosen from a $\rho(x)$ distribution. At each time step, the agent or individual selects the highest priority task from the list and executes it, thus removing it from the list. At that moment, a new task is added to the list, its priority again chosen from $\rho(x)$. This simple model ignores the possibility that the agent occasionally selects a low priority item for execution before all higher priority items are completed, a situation common for tasks with deadlines. The deadline-driven execution may be incorporated in this scenario by assuming that the agent executes the highest priority item with probability *p*, then executes with probability 1- *p* a randomly selected task, independent of its priority. The limit $p \to 1$ therefore describes the deterministic (iii) protocol, when the highest priority task is always chosen for execution, while $p \to 0$ corresponds to the random choice protocol discussed in (ii).

Model B [55]-[56]: Assume that tasks arrive to the priority list at the rate λ following a Poisson process with exponential arrival-time distribution. The arrival of each new task increases the length of the list from $L \; to \; L+1$. The tasks are executed at the rate μ, reflecting the overall time an individual devotes to its priority list. Once a task is executed, the length of the priority list decreases from $L \; to \; L-1$. Each task is assigned a discrete priority parameter x_i upon arrival, such that the highest-priority unanswered task is always chosen for execution. The lowest priority task must wait the longest before execution, therefore dominating the waiting-time probability density for large waiting times.

The only difference between model A and model B is in the length of the queue L. In model A, the queue length is fixed and remains unchanged during the model's evolution, while in model B, the queue length L can fluctuate as tasks arrive or are executed. This small difference however, has a large impact on the distribution of the waiting times of tasks on the priority list. Indeed, numerical and analytical results indicate that both models give rise to power-law waiting-time distribution. Model A predicts that in the $p \to 0$ limit, the waiting times follow equation (4) with $\alpha = 1$ [8],[53]. This agrees with the observed scaling for e-mail communications, web browsing, and several other human activity patterns. In contrast, we find for model B that the waiting-time distribution follows (4), with exponent $\alpha = 3/2$, as long as $\mu \leq \lambda$. This is in agreement with the results obtained for the mail correspondence patterns of Einstein [40],[55]-[56]. These results indicate that human dynamics are described by at least two universality classes, characterized by empirically distinguishable exponents. In searching to explain the observed heavy-tailed human activity patterns, we have limited our study to single queues. In reality, none of our actions is performed independently, as most of our daily activities are embedded in a web of actions by other individuals. Indeed, the timing of an e-mail sent to user A may well depend on the time we receive an e-mail from user B. A future goal of human dynamics research is to understand how various human activities and their timing are affected by the fact that individuals are embedded in a network environment. Such understanding will ultimately bring together the study of network topology and network dynamics.

From Networks Theory to a Theory of Complexity

As it stands today, network theory is not a proxy for a theory of complexity. Network theory currently addresses the emergence and structural evolution of the skeleton of a complex system. The ultimate understanding of the overall *behavior* of a complex network must however account for its architecture as well as the nature of dynamical processes taking place on such a network. We believe that we are at the threshold of unraveling the characteristics of these dynamical processes. Indeed, the data collection capabilities that originally catalyzed the explosion of research on network topology are now far more penetrating. These capabilities are allowing us to capture not only a system's topology, but also the simultaneous dynamics of its components, such as the communication and travel patterns of millions of individuals, [57]-[60] or the expression level of all genes in a cell. Such capabilities may one day lead to a systematic program, starting from a measurement-based discovery process, and potentially ending in a theory of complexity with predictive power. Therefore, structural network theory is by no means the end of a journey, but rather an unavoidable step towards the ultimate goal of understanding complex systems. Should a theory of complexity be ever completed, I believe that it will incorporate the newly discovered fundamental laws governing the architecture of complex systems [61].

At the nodes' and links' level, each network is driven by apparently random and unpredictable events. Despite this microscopic randomness, a few fundamental laws and organizing principles are helping to explain the topological features of such diverse systems as the cell, the Internet, or society. This new type of universality is driving the multidisciplinary explosion in network science. Will there be a similar degree of universality in the dynamics of large complex systems? Are there generic organizing principles that are just as intriguing and powerful as those uncovered in the past few years in the quest to understand networks topologies? As shown in this article, results indicating some degree of universality in the behavior of individual nodes in complex systems are emerging. So far, these results refer to human dynamics only, but the questions they pose will probably initiate research in other areas as well. Similarly exciting results have emerged lately on the nature of fluctuations in complex networks [62]-[65]. These findings provide us hope for interesting potential developments on dynamics, just as network theory provided us on structure and topology.

Networks represent the architecture of the complexity. To fully understand complex systems however, we need to move beyond this architecture, and to uncover the laws governing the underlying dynamical processes. Most importantly, we need to understand how these two layers of complexity, architecture and dynamics, evolve together. There are formidable challenges for physicists, biologists, computer scientists, engineers, and mathematicians, inaugurating a new era that Stephen Hawking recently called the *century of complexity*.

Acknowledgements This research was supported by NSF ITR 0426737, NSF ACT/SGER 0441089, and the NSF NetWorkBench awards and by an award from the James S. McDonnell Foundation. I wish to also thank Alexei Vazquez and Suzanne Aleva for their assistance with the preparation of this review.

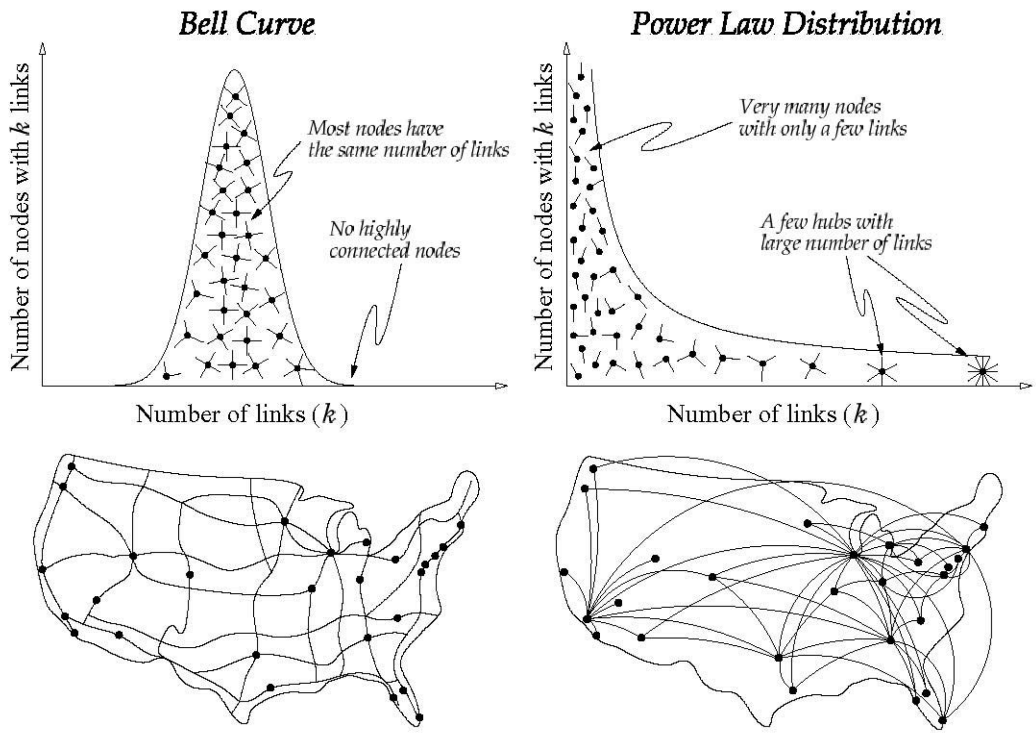

Figure 1: Random and Scale-Free Networks. The degree distribution of a random network follows a Poisson curve close in shape to the Bell Curve, telling us that most nodes have the same number of links, and nodes with a very large number of links don't exist (top left). Thus a random network is similar to a national highway network, in which the nodes are the cities, and the links are the major highways connecting them. Indeed, most cities are served by roughly the same number of highways (bottom left). In contrast, the power law degree distribution of a scale-free network predicts that most nodes have only a few links, held together by a few highly connected hubs (top right). Visually this is similar to the air traffic system, in which a large number of small airports are connected to each other via a few major hubs (bottom right). After [1].

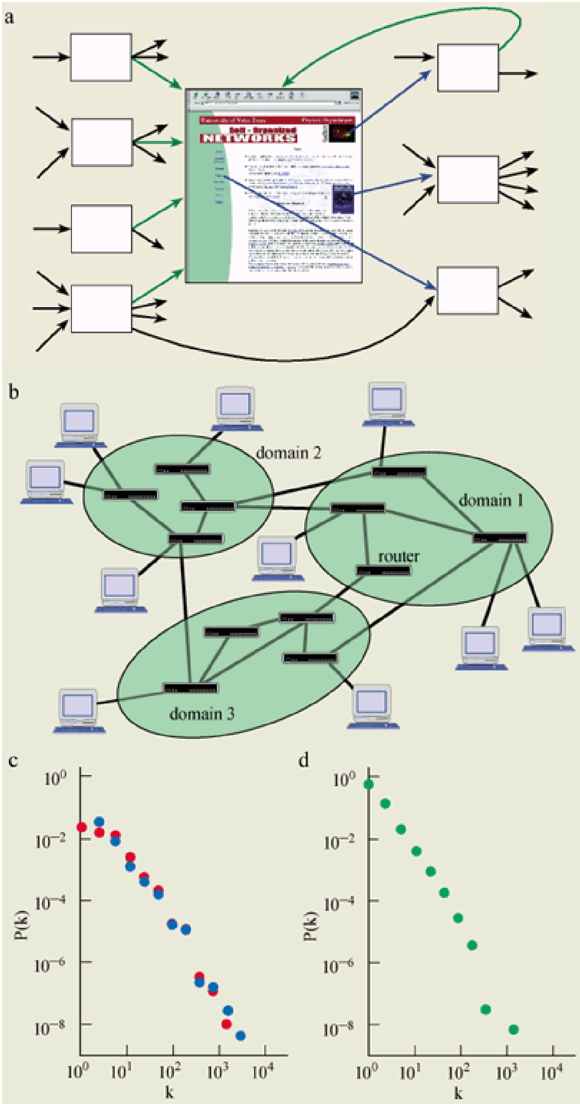

Figure 2. (a) The nodes of the World Wide Web are Web documents, each of which is identified by a unique uniform resource locator, or URL. Most documents contain URLs that link to other pages. These URLs represent outgoing links, three of which are shown (blue arrows). Currently there are about 80 documents worldwide that point to the Web site www.nd.edu/~networks, represented by the incoming green arrows. While we have complete control over the number of the outgoing links, *k*out, from our Web page, the number of incoming links, *k*in, is decided by other people, and thus characterizes the popularity of the page. (b) The Internet itself, on the other hand, is a network of routers that navigate packets of data from one computer to another. The routers are connected to each other by various physical or wireless links and are grouped into several domains. (c) The probability that a Web page has *k*in (blue) or *k*out (red) links follows a power law. The results are based on a sample of over 325 000 Web pages collected by Hawoong Jeong. (d) The degree distribution of the Internet at the router level, where *k* denotes the number of links a router has to other routers. This

research by Ramesh Govindan from University of Southern California is based on over 260,000 routers and demonstrates that the Internet exhibits power-law behavior. After [66].

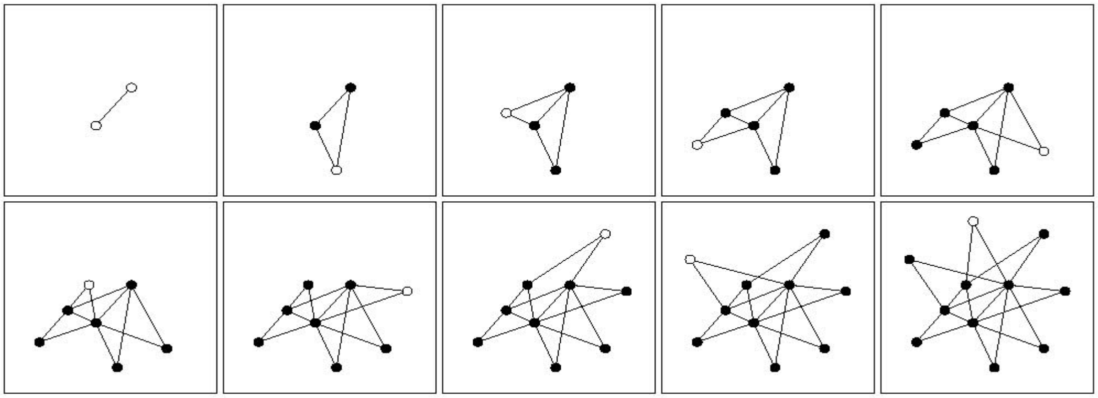

Figure 3. The Birth of a Scale-Free Network. The scale-free topology is a natural consequence of the ever-expanding nature of real networks. Starting from two connected nodes (top left), in each panel a new node (shown as an empty circle) is added to the network. When deciding where to link, new nodes prefer to attach to the more connected nodes. Thanks to growth and preferential attachment, a few highly connected hubs emerge. After [1].

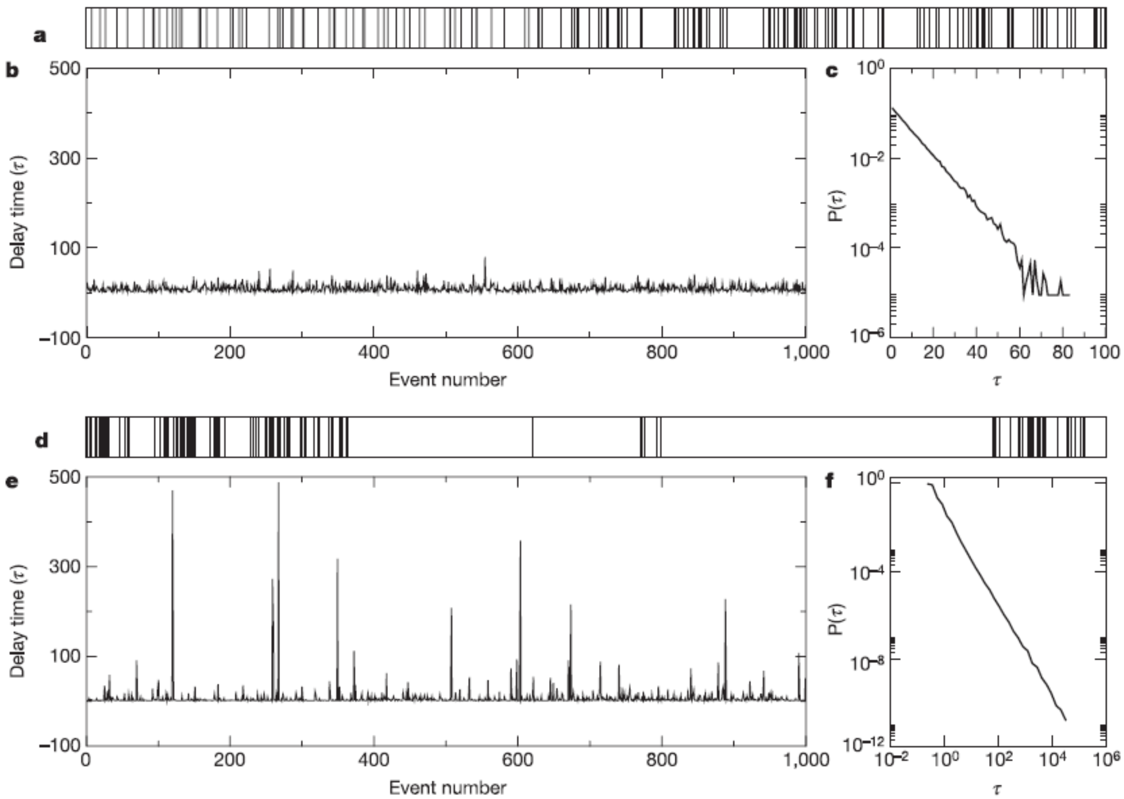

Figure 4: The difference between the activity patterns predicted by a Poisson process (top) and the heavy tailed distributions observed in human dynamics (bottom). **a** Succession of events predicted by a Poisson process, which assumes that in any moment an event takes place with probability q. The horizontal axis denotes time, each vertical line corresponding to an individual event. Note that the interevent times are comparable to each other, long delays being virtually absent. **b** The absence of long delays is visible on the plot showing the delay times τ for 1,000 consecutive events, the size of each vertical line corresponding to the gaps seen in **a.** **c** The probability of finding exactly n events within a fixed time interval is $P(n; q)=e^{-qt}(qt)^n/n!,$ which predicts that for a Poisson process the inter-event time distribution follows $P(\tau)=qe^{-q\tau}$, shown on a log-linear plot in **c** for the events displayed in **a, b-d,** The succession of events for a heavy tailed distribution. **e** The waiting time τ of 1,000 consecutive events, where the mean

event time was chosen to coincide with the mean event time of the Poisson process shown in **a-c**. Note the large spikes in the plot, corresponding to long delay times. **b** and **e** have the same vertical scale, allowing to compare the regularity of a Poisson process with the bursty nature of the heavy tailed process. **f,** Delay time distribution $P(\tau) \sim\tau^{-2}$ for the heavy tailed process shown in **d, e,** appearing as a straight line with slope -2 on a log-log plot. After [53].

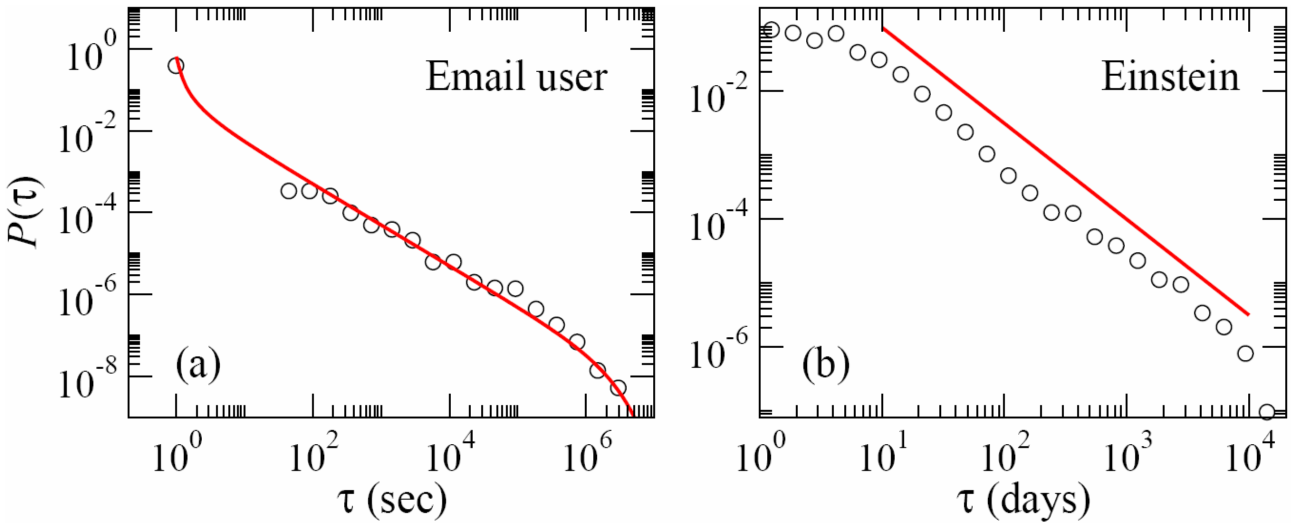

Figure 5: *Left*: The response time distribution of an email user, where the response time is defined as the time interval between the time the user first sees an email and the time she sends a reply to it. The first symbol in the upper left corner corresponds to messages that were replied to right after the user has noticed it. The continuous line corresponds to the waiting time distribution of the tasks, as predicted by Model A discussed in the paper, obtained for p=0.999999+0.000005. Right: Distribution of the response times for the letters replied to by Einstein. The distribution is well approximated with a power law tail with exponent α=3/2, as predicted by Model B. Note that while in most cases the identified reply is indeed a response to a received letter, there are exceptions as well: some of the much delayed replies represent the renewal of a long lost relationship. After [40].

Albert-László Barabási is the Director of the Center for Complex Network Research, and he is at the Department of Physics and Computer Science, University of Notre Dame, Notre Dame, IN 46556, USA. He studied physics and engineering, at the University of Bucharest, and obtained an M.Sc. in Physics in 1991 from Eotvos Lorand University, Budapest under T. Vicsek. He obtained his Ph.D. in Physics in 1994 from Boston University under H.E. Stanley. His was elected in 2003 a Fellow of the American Physical Society, and in 2004 a member of the Hungarian Academy of Sciences. In 2005, he received the FEBS Anniversary Prize for Systems Biology and in 2006 the Neumann Medal for Computer Science. His book Linked: The new science of networks, is available in eleven languages.

References

[1] A.-L. Barabási, *Linked: The New Science of Networks*. Plume Books, MA, 2003.

[2] G. Caldarelli, *Scale-Free Networks*. Oxford University Press, Oxford, 2007.

[3] R. Albert and A.-L. Barabási, “Statistical mechanics of complex networks,” *Review of Modern Physics*, vol. 74, pp. 47-97, 2002.

[4] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach*, Cambridge University Press, Cambridge, 2004.

[5] A.-L. Barabási and Z. N. Oltvai, “Network biology: understanding the cell’s functional organization,” *Nature Rev. Gen.,* vol. 5, pp. 101-113, 2004.

[6] M. E. J. Newman, A.-L. Barabási and D. J. Watts, *The Structure and Dynamics of Complex Networks*. Princeton University Press, Princeton, 2006.

[7] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW.* Oxford University Press, New York, 2003.

[8] S. H. Strogatz, “Exploring complex networks,” *Nature*, vol. 410, pp. 268-276, 2001.

[9] B. Ben-Naim, H. Frauenfelder, and Z. Toroczski, Ed. *Complex Networks*, Lecture Notes in Physics, vol. 650, Springer; Germany, 2004.

[10] P. Erdős and A. Rényi, “On random graphs I,” *Publ. Math. Debrecen*, vol. 6, pp. 290-297, 1959.

[11] P. Erdős and A. Rényi, “On the evolution of random graph,” Publ. Math. Inst. Hung. Acad. Sci. vol. 5, pp. 17-61, 1960.

[12] B. Bollobás, *Random Graphs*, Academic, London, 1985.

[13] R. Albert, H. Jeong and A.-L. Barabási, “Diameter of the World Wide Web,” *Nature*, vol. 401, pp. 130-131, 1999.

[14] A.-L. Barabási and R. Albert, “Emergence of scaling in random networks,” *Science,* vol. 286, pp. 509-512, 1999.

[15] M. Faloutsos, P. Faloutsos and C.Faloutsos, “On power-law relationships of the internet topology,” in *Proc. ACM SIGCOMM 99*, 1999, pp. 251-262.

[16] S. Milgram, “The small world problem,” *Psychology Today*, vol. 1, pp. 60-67, 1967.

[17] J. Guare, *Six Degrees of Separation*, Vintage Books, New York, 1990.

[18] M. S. Granovetter, “The strength of weak ties,” *Am. J. Sociol.,* vol. 78, no. 6, pp. 1360-1380, 1973.

[19] S. Lawrence and C. L. Giles, “Searching the World Wide Web,” *Science*, vol. 280, pp. 98-100, 1998.

[20] S. Lawrence and C. L. Giles, “Accessibility of information on the web,” *Nature*, vol. 400, pp. 107-109, 1999.

[21] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. rajalopagan, R. Stata, A. Tomkins and J. Weiner, “Graph structure in the web,” *Int. J. of Comp. and Telecom. Netw.*, vol. 33, pp. 309-320, 2000.

[22] B. A. Huberman, and L. Adamic, “Growth dynamics of the World Wide Web,” *Nature*, vol. 401, pp. 131, 1999.

[23] A.-L. Barabási, R. Albert and H. Jeong, “Mean-field theory for scale-free random networks,” *Physica A*, vol. 272, pp. 173-187, 1999.

[24] P. L. Krapivsky, S. Redner, and F. Leyvraz, “Connectivity of growing random networks,” *Phys. Rev. Lett.*, vol. 85, pp. 4629-4632, 2000.

[25] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, “Structure of growing networks with preferential linking,” *Phys. Rev. Lett.*, vol. 85, pp. 4633-4636, 2000.

[26] B. Bollobás, O. Riordan, J. Spencer and G. Tusnady, “The degree sequence of a scale-free random graph process,” *Random Structures & Algorithms*, vol. 18, no. 3, pp. 279-290, 2001.

[27] G. Bianconi and A.-L. Barabási, “Competition and multiscaling in evolving networks,” *Europhys. Lett.*, pp. 54, pp. 436-442, 2001.

[28] G. Bianconi and A.-L. Barabási, “Bose-Einstein Condensation in Complex Networks,” *Phys. Rev. Lett.,* vol. 86, no. 24, pp. 5632-5635, 2001.

[29] R. Albert, H. Jeong and A.-L. Barabási, “The internet’s Achilles’ heel: Error and attack tolerance in complex networks,” *Nature*, vol. 406, pp. 378-382, 2000.

[30] R. Cohen, K. Reez, D. Ben-Avraham, and S. Havlin, “Resilience of the Internet to random breakdowns,” *Phys. Rev. Lett.*, vol. 85, no. 21, pp. 4626-4628, 2000.

[31] R. Cohen, K. Reez, D. Ben-Avraham, and S. Havlin, “Breakdown of the Internet under intentional attack,” *Phys. Rev. Lett.* vol. 86, no. 16, pp. 3682-3685, 2001.

[32] L. Li, D. Alderson, W. Willinger, and J. Doyle, “A first principles approach to understanding the Internet’s router-level topology,” *Computer Comm. Rev.* vol. 34, no. 4, pp. 3-14, 2004.

[33] R. Pastor-Satorras and A. Vespignani, “Dynamical and Correlation Properties of the Internet,” *Phys. Rev. Lett.,* vol. 87, no. 25, pp. 258701-1-4, 2001.

[34] Z. Dezső and A.-L. Barabási, “Can we stop the AIDS epidemic?,” *Phys. Rev. E,* vol. 65, pp. 055103-1-4 (2002).

[35] S. Havlin and D. ben-Avraham, “Efficient Immunization Strategies for Computer Networks and Populations,” *Phys. Rev. Lett.*, vol. 91, no. 24, 247901-1-4, 2003.

[36] R. Pastor-Satorras and A. Vespignani, “Immunization of complex networks,” *Phys. Rev. E,* vol. 65, pp. 036104-1-8, 2002.

[37] F. A. Haight, *Handbook on the Poisson Distribution*, Wiley, New York, 1967.

[38] P. Reynolds, *Call Center Staffing*, The Call Center School Press, Lebanon, Tennessee, 2003.

[39] J. H. Greene, *Production and Inventory Control Handbook* 3rd ed, McGraw-Hill, NewYork, 1997.

[40] A. Vázquez, J. G. Oliveira, Z. Dezső, K.-I. Goh, I. Kondor and A.-L. Barabási, “Modeling bursts and heavy-tails in human dynamics,” *Phys. Rev. E,* vol. 73, no. 3, pp. 036127-1-19, 2006.

[41] H. R. Anderson, *Fixed Broadband Wireless System Design*, Wiley, New York, 2003.

[42] J. P. Eckmann, E. Moses and D. Sergi, “Entropy of dialogues creates coherent structure in e-mail traffic,” in *Proc. Natl. Acd. Sci,* . USA, col. 101, 2004, pp. 14333-14337.

[43] H. Ebel, L. I. Mielsch and S. Bornholdt, “Scale-free topology of e-mail network,” Phys. Rev. E, vol. 66, pp. 35103-1-4, 2002.

[44] C. Dewes, A. Wichmann and A. Feldman, “An Analysis of Internet Chat Systems,” in *Proc. 2003 ACM SIGCOMM Conf. on Internet Measurement (IMC-03)*, ACM Press, New York, 2003 pp. 51-64.

[45] S. D. Kleban and S.H. Clearwater, “Hierarchical Dynamics, Interarrival Times and Performance,” in P*roc. of ACM/IEEE Supercomputing*, November 15-21, Phoenix, AZ, 2003, pp. 28-28.

[46] V. Paxson and S. Floyd, "Wide-Area traffic: The failure of Poisson modeling," *IEEE/ACM Transactions in Networking*, vol. 3, no. 3, pp. 226-244, 1995.

[47] U. Harder and M. Paczuski, "Correlated dynamics in human printing behavior," *Physica A*, vol. 361, no. 1, pp. 329-336, 2006.

[48] Z. Dezső, E. Almaas, A. Lukacs, B. Racz, I. Szakadat and A.-L. Barabási, "Fifteen minutes of fame: The dynamics of information access on the web, " http://arxiv.org/abs/physics/0505087, 2005.

[49] V. Plerou, P. Gopikirshnan, L.A.N. Amaral, X. Gabaix, and H.E. Stanley, "Economic fluctuations and anomalous diffusion," *Phys. Rev. E,* vol. 62, pp. 3023-3026, 2000.

[50] J. Masoliver, M. Montero and G. H. Weiss, "Continuous-time random-walk model for financial distributions," *Phys. Rev. E*, vol. 67, pp. 021112/1-9, 2003.

[51] T. Henderson and S. Nhatti, "Modeling user behavior in networked games," in *Proc. ACM Multimedia*, Ottawa, Canada, 2001, pp. 212-220.

[52] A. Vázquez, "Exact results for the Barabási model of human dynamics," Phys. Rev. Lett. 95, pp. 248701/1-4, 2005.

[53] A.-L. Barabási, *The origin of bursts and heavy tails in human dynamics*, Nature, vol. 435, pp. 207-211, 2005.

[54] J. W. Cohen, *The Single Server Queue*, North Holland, Amsterdam, 1969.

[55] J. G. Oliveira and A.-L. Barabási, "Human Dynamics: The Correspondence Patterns of Darwin and Einstein," *Nature,* vol. 437, pp. 1251-1254, 2005.

[56] A. Cobham, "Priority assignment in waiting line problems," *J. Oper. Res. Sec. Am*. Vol. 2, pp. 70-76, 1954.

[57] M.C. González and A.-L, Barabási, Complex Networks: From data to models, Nature, Physics (in press).

[58] G. Palla, A-L. Barabási and T. Vicsek, "Quantifying social group evolution," Nature (in press, 2007).

[59] J.-P. Onnela, J. Saramaki, J. Hyvonen, G. Szabo, D. Lazer, K. Kaski, J. Kertesz, A.-L. Barabasi, "Structure and tie strengths in mobile communication networks," PNAS (in press, 2007) http://lanl.arxiv.or=g/abs/physics/0610104.

[60] Vazquez, A., Balázs, R., András, L. & Barabási A.-L. Phys. Rev. Lett. (in press); preprint at http://arxiv.org/abs/physics/0609184, 2005.

[61] A.-L. Barabási, "Taming complexity," *Nature Physics*, vol. 1, pp. 68-70, 2005.

[62] M. A. Menezes and A.-L. Barabási, "Fluctuations in network dynamics," *Phys. Rev. Lett.*, vol. 92, no. 2, pp. 028701/1-4, 2004.

[63] M. A. de Menezes and A.-L. Barabási, "Separating internal and external dynamics of complex systems," *Phys. Rev. Lett*. vol. 93, no. 6, pp. 068701/1-4, 2004.

[64] Z. Eisler, J. Kertesz, S.-H. Yook and A.-L. Barabási, "Multiscaling and non-universality in fluctuations of driven complex systems," *Europhysics Letters*, vol. 69, pp. 664-670, 2005.

[65] J. Duch, A. Arenas, " Scaling of fluctuations in traffic on complex networks," *Phys. Rev. Lett.,* vol. 96, no. 21, Art. No. 218702, 2006.

[66] A.-L. Barabási, "The physics of the Web", *Physics World,* vol. 14, pp. 33-38, 2001.