From Technological Networks to Social Networks

Kwang-Cheng Chen, *Fellow, IEEE* Mung Chiang, *Fellow, IEEE* and H. Vincent Poor, *Fellow, IEEE*

Abstract—Social networks overlaid on technological networks account for a significant fraction of Internet use. Through graph theoretic and functionality models, this paper examines social network analysis and potential implications for the design of technological networks, and vice versa. Such interplay between social networks and technological networks suggests new directions for future research in networking.

Index Terms—Social networks, network science, random graph, Internet, big data, statistical information processing, networked data, data analysis, social behavior, network economy, information epidemics, statistical communication theory.

I. INTRODUCTION

NEEDLESS to say, the Internet is one of the most important inventions in the field of information and communication technology (ICT), and has enabled a highly, globally, and somewhat unexpectedly connected world which is greatly changing human interaction and therefore society. Since the introduction of Web browsers in the 1990s, the number of Internet users has grown at an explosive pace, as have the number and variety of Web-based applications. The emergence of Web 2.0 has spurred dramatic growth in the numbers of networking sites and *active* users. With the aid of smartphones and mobile computing, users are spending more and more time on the Internet. Thus, we are moving into a highly *connected age. Network science* [1]–[9] is concerned with the in-depth topological study of connected systems, ranging from the human brain and nervous systems to food webs, the World Wide Web (WWW), power grids, economic interactions and business, social networks, etc. The future of network science must be based on domain-specific models and on the pursuit of falsification. For example, while a random graph is elegant, it is often neither a relevant approach to design nor the only generative model to explain network phenomena. And as much as the metrics of static graphs are important, engineering protocols governing the functionalities of feedback, coordination, and robustness are just as crucial as the topological properties of the graph. When electrical engineers and computer scientists are designing, implementing, and operating the Internet, similar topological studies are of considerable interest; for example, autonomous information acquisition for physical layer topology played a key role in the early stages of Internet development [10], [11]. Comprehensive reviews of work on network topologies and ways to infer and model Internet systems can be found in [12] and [13]. A key message from the community has been that the interplay between technological and social networks is much richer than simply common features in their topological properties, but rather extends well into the study of robustness, protocols, measurements, dynamics, and functionalities [14].

In the early stages of the Internet age, portals such as Yahoo played a key role in attracting users' attention and provided a focal point for advertising and e-commerce marketing. Web 2.0 enabled an interactive user experience to dramatically advance Internet services such as Google and Facebook, which involve certain kinds of relationships among either individuals or data. If we consider each individual or each item of data to be a node, and each relationship between two of them as a link, then a new kind of network is formed, namely a *social network*. Consequently, as depicted in Figure 1, networks can be viewed as having both social overlay and technological underlay networks. That is, we have two modes of connections in the complex connected world: physical connections and virtual connections. It is noteworthy that anytime, anyplace, anyone networking has attracted tremendous user activity, with more than 60% of total Internet traffic now coming from social networks and peer-to-peer (P2P) transactions. Thereby, human factors play a critical role in the development of future networking technologies, which gives us a strong motivation to fully understand social networks and thus to optimally design technological networks to support user experience and services.

There are important concepts that cut across social and technological networks, namely:

- The emergence of global coordination through local actions based on local views is a recurring theme, from influence models in social networks to routing and congestion control in the Internet, and from consumer reaction to pricing signals to power control in wireless networks.
- Resource sharing models, in the form of additive sharing $x + y \geq 1$, mulitplicative sharing $x/y \geq 1$, or binary sharing $x, y \in \{0,1\}$, $x + y \leq 1$, are introduced for network pricing as well as the classical problems of congestion control, power control, and contention control. Indeed, congestion control in Transfer Control Protocol (TCP) has been interpreted as a form of dynamic pricing in network access.
- "The wisdom of (independent and unbiased) crowds" is another common theme in social networks. There are two types of "wisdom" here: diversity gain in reducing
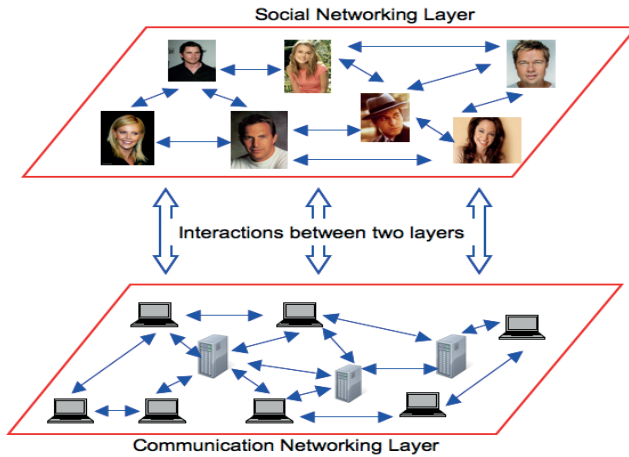
Fig. 1.   Interplay between social networking and technological networking

the chance of some bad event (typically represented mathematically as $1 - (1 - p)^N$ where N is the size of the crowd), and efficiency gain in smoothing out some average metric (typically represented mathematically as a factor $\sqrt{N}$ in the metric). Both types are observed in social networks and the latest generation of 4G cellular and 802.11n wireless networks.

- Spatial hierarchy is used in both small world models and in how the Internet is structured.
- The (positive) network effect is often highlighted in social and economic networks. It also finds a concrete realization in how content is shared over the Internet through peer-to-peer protocols and scaling-up of data centers.

Research on social networks can be traced back to Herbert Simon [15] and Anatol Rapoport [16] in the 1950s. Harrison White later provided pioneering insights. Stanley Milgram conducted his famous small-world experiments in the 1960's [17], seeking to verify the folk wisdom that any two people on the planet are separated by a low number of degrees of separation. And slightly over a decade ago, Watts and Strogatz energized the small-world networks research community by introducing graphical modeling to the field [1]. Shortly thereafter, Barabasi and Albert introduced scale-free networks [2], which spurred further progress in network science. Today, there is a considerable literature about social networks and their analysis, primarily in the social sciences [18], statistical physics, and some in computer science or signal processing [19], [20]. However, a comprehensive overview from the communication networking community is lacking, and thus this paper seeks to serve the purpose of looking at social networks and their analysis in this context, and in particular their impact on future communication systems and networks.

Since the fundamental theory of network science has been well-documented in many papers and books, particularly [8], [21], [22], and [23], this paper is organized instead in a vertical way to link the critical nature of social networks with well-known technological communication networks.

The rest of this paper is organized as follows. We briefly review graph theoretic modeling and analysis, with applications to the Internet and wireless networks, in Section II. Scale-free networks and thus the Internet's Achilles heel are introduced and we describe a functional model to enhance robustness and security of networks in Section III. Starting from small world networks, we demonstrate interplay with technological networks in Section IV. We then examine common methodologies in pagerank and code-division multiple-access (CDMA), and important concepts of recommender systems and influence in social networks, in Section V. Social behavior, network economy, and various aspects of network design are explored in Section VI.

II. Graph Theoretic Modeling

Modern social network analysis [23] is based on both large-scale data analysis and graph theory, the latter of which was invented by L. Euler. In this paper we will focus on this latter approach, providing background on the former in Appendix A. As a relationship in a social network can be appropriately modeled as a random variable, random graph theory [24] which was extensively studied by P. Erdös in 1959 [25] is therefore widely adopted to understand the behavior of social networks and complex networks via construction of a probabilistic networking model [3]–[6], [21]–[23]. As in Section II.A, a Bayesian networks approach is widely applied to construct probabilistic networks or graphs [26], [27]. Mathematically speaking, a social network can be described by a graph $G = (V, E)$, where V is the collection of vertices and E is the collection of edges. A vertex v_n represents a person, an agent, or a data entity $(\phi_1, \phi_2, \ldots, \phi_L)$, where ϕ_1 may be the height of a person, ϕ_2 may be the weight of this person, etc. Sometimes, $v_1, \ldots, v_N$ can be viewed as a *dictionary*. An edge between two nodes or vertices, which can be directed or undirected, logically represents the corresponding relationship. According to the well-known *Cox Axiom* [28] from artificial intelligence, we may use a random variable to represent such a relationship, and thus probabilistic network construction and random graphical modeling will play a central role later in this paper.

A. Graph Model of Social Networks

For a social network, as noted in Section I, we can obtain a corresponding graph $G = (V, E)$. This graph of N nodes can be completely described by the *adjacency* (or *connectivity*) matrix $\boldsymbol{A}$ which is an $N \times N$ square matrix, with entry $a_{ij}, i, j = 1, \ldots, N$, equaling 1 if link (or edge) l_{ij} exists, or 0 otherwise. Some statistical properties of this graph are of general interest:

a) Node degree, degree distribution, and correlations: The degree (or connectivity) k_i of the node i is the number of edges (i.e. links) incident on this node, and equals $\sum_{j \in V} a_{ij}$. For a directed graph, we may separately define outgoing links and incoming links for a node. The most fundamental topological characterization of a graph is the *degree distribution* $P(k)$, the probability that a randomly and uniformly selected node has degree k. The mth moment is $\langle k^m \rangle$. The degree distribution suffices to determine statistical properties of uncorrelated networks. However, many realistic networks are actually correlated;
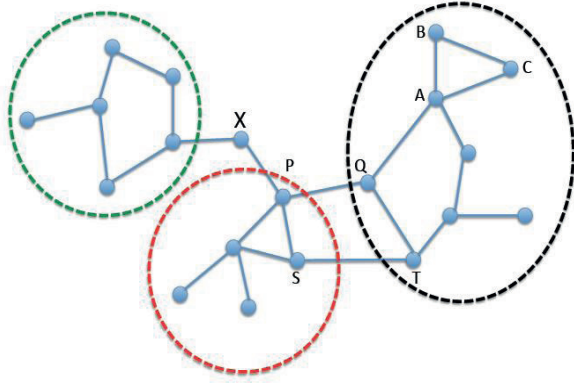
Fig. 2. Graph model with three communities as circled; node X has high degree-centrality and thus strong influence for message passing; $P - Q$ and $S - T$ have less betweenness-centrality; the transitivity relationship that A knows B and A knows C suggests B and C are likely acquainted

that is, the degrees k and k' of neighboring nodes can depend on one another, which can be specified through a conditional probability $P(k'|k)$. To avoid fluctuations in realistic networks, the average nearest neighbor's degree $k_{nn}(k) = \sum_{k'} k' P(k'|k)$ can be used to characterize this behavior.

b) Clustering coefficient: To measure a clique such as two individuals with a common friend, the graph clustering coefficient for node i is

$$c_i = \frac{\sum_{j,k} a_{ij} a_{jk} a_{ki}}{k_i (k_i - 1)}. \tag{1}$$

c) Diameter and betweenness: Suppose d_{ij} is the distance (metric) from node i to node j. The diameter of the graph is then defined as $D_{max}(G) = \max_{i,j} d_{i,j}$. $P - Q$ and $S - T$ in Figure 2 together illustrate the notion of betweenness. If we remove $P - Q$, then $S - T$ has higher betweenness.

d) Centrality: Sometimes, a microscopic view of a graph is useful. Centrality is a critical factor, which can be defined in several ways using degree, closeness, betweenness, and neighbors' characteristics. The node X in Figure 2 illustrates centrality between two clusters.

e) Spectrum: The eigenvalues of the adjacency (or connectivity) matrix $\boldsymbol{A}$ can reveal many useful properties, according to algebraic graph theory.

Many large networks, no matter whether they are technological networks such as the Internet, the Web, power grids, or social networks such as acquaintance networks, collaboration networks, or even biological networks such as metabolic networks or food webs, have some common features:

- Power-law degree distribution, scale-freeness or self-similarity: The degree distribution of these large networks can often be represented in the form $P(k) \sim k^{-\gamma}$ with $\gamma > 1$. For example, in [2], it is shown that $\gamma \approx 2.3$ for the WWW and $\gamma \approx 2.4$ for the Internet. Networks with this property are called *scale-free networks*. Biological networks such as protein interaction networks and metabolic networks, share the same property. Power-law

degree distributions can occur either through critical phenomena or constrained optimization. Scale-free networks usually exhibit self-similarity, i.e., they consist of self-repeating patterns on all length scales.
- Network transitivity or clustering: This will be discussed in Section III below.
- Small-world phenomena: This will be discussed in Section V.

1) Community Detection: In this subsection, we look into network transitivity or clustering in a network. Network transitivity in social networks means that two acquaintances of a single individual have a better chance of knowing each other than do two people randomly selected from the population. This effect is typically quantified by the cluster coefficient which was defined earlier, and for which a more global definition is

$$c = \frac{3 \cdot (\text{number of triangles on the graph})}{\text{number of connected triples of vertices}}. \tag{2}$$

A network structure has substantial clustering when c does not go to zero as the network size becomes large. Another way to study clustering is via the community structure. Communities exist in social networks and typically represent social groups with common interests, background, etc. Detecting and identifying community structure in a (social) network may assist in better understanding of network structure and function. Hierarchical clustering is the most straightforward and this concept has been widely applied in the operation of mobile ad hoc networks (MANETs). Another way to view community structure is to identify the edges most central to communities in a graph. Similarly, we may identify the edges most "between" communities [29]. In addition to network transitivity or clustering, social networks have another feature, *assortative mixing*, which means positive correlation between the degrees of adjacent vertices [30]. For the Internet, vertices of high degree tend to be connected to those vertices of low degree on average. Consequently, the Internet is disassortative and is unlike social networks in this sense. This nature of social networks comes from *degree correlation*, as individuals who belong to a small group are likely connected to others in the same group of low degree. Similarly, those who are in a large social group (community) and have high degree tend to be connected to each other. This is also known as the *rich club phenomenon*. In Section IV, we will discuss whether attacks or failures of high-degree nodes may make the entire network fall apart. A recent study of Internet topology at the autonomous system (AS) level shows a clear rich club phenomenon [31]. Fortunately, in the AS graph, there exist a large number of alternative routing paths among rich club members to maintain network efficiency, redundancy, and robustness. This is a good example to show interplay between social network analysis and practical technological networks.

Data from Web and Internet services, such as Web text, user activities, market data analysis, etc., represent another kind of social network. A typical situation is to form a two-dimensional (or multi-dimensional) contingency or co-occurrence table based on such data. We may consider this contingency table as an empirical joint probability distribution that can be used to identify simultaneous clustering of rows

and columns, which is known as co-clustering [32]. In this large-scale data analysis, we may leverage the concepts of mutual information and information divergence from information theory to develop the co-clustering algorithms. In many situations, clustering or detection of community structure is not clear. *Modularity* can serve as the parameter to partition the graph model of a (social) network. Using eigenvalues of modularity, we may decompose the network into a number of communities [33]. Techniques can be further developed to identify overlapping community structure [34]. The study of community structure is very useful in engineering applications. For example, we can form a mobile call graph based on phone calls in a cellular network. This data and the resulting graph allow us to explore the relationships and strengths of ties among individuals and behaviors such as information spread. This interaction between social networks (phone users) and technological networks (cellular networks) gives useful information [35].

2) Evolution and Control of Graphs: Graph theoretic analysis typically focuses on two fundamental properties, degree and distance, both for social networks and for technological networks. However, almost all networks evolve over time, which involves addition or deletion of nodes and edges. It is therefore of interest to understand time varying properties of graph theoretic network evolution models. Traditional analysis suggests two conjectures about network evolution: constant average degree and slowly growing diameter. After examining various social networks (e.g. email networks, citation networks, the Internet Movie Database network, etc.) and technological networks, two observations have been summarized in [36] regarding evolution of the corresponding graphs:

- Densification power laws: The networks become denser over time with the average degree increasing, which follows the *densification power law*. That is, $e(t) \propto n(t)^{\alpha}$, where $e(t)$ and $n(t)$ denote, respectively, the number of edges and nodes of the graph evolving over time, and $1 < \alpha < 2$.
- Shrinking diameters: In many cases, the effective diameter decreases as the network grows.

The preceding discussion has considered the graph model of bidirectional and binary edges. However, some social networks are constructed on relationships, such as friendship, acquaintance, influence, etc., for which it is not appropriate to use bidirectional and/or binary edges/links. In this situation, particularly to understand *influence* (more in Section VI.C), the edge of the graph or the link of the network is commonly represented as a tie to show the strength of a relationship. *Stochastic actor-based models* are therefore introduced for such network dynamics. This type of network evolves as a stochastic process driven by the actors (or agents, nodes); that is, actors change their outgoing ties. The resulting network forms a Markov process. In-degree or out-degree can be used to analyze behavior by setting proper objective functions [37]. Based on the theory of homophily, the relationship strength is probabilistically modeled as a hidden effect of nodal profile similarity [38]. The graphical model in machine learning can be adopted to represent the relationship strength among agents, typically involving ties in one direction. As indicated in the appendix, since the behavior is stochastic, statistical inference is employed to obtain desired results from such social network analysis. We can further add weighting to such unidirectional-link graphs to design referral systems to search dynamic social networks, in which each node requires only local knowledge [39].

B. Graph-Based Technological Network Design

1) Scaling in Data Centers and P2P: We reflect upon three ways of drawing and sizing topologies, one for each key type of wireline technology network. We also explore the root causes behind these different choices.

a) *The Internet backbone*: Overprovision link *capacities* and then run IP routing, possibly multipath routing, carefully. Since routing is not responsive to link load in real time, and the job of congestion control is given to the transport layer, the end hosts react to varying loads on a fast timescale via TCP.

b) *Data center network*: Scale-up by scaling-out and building a large network with small switches through interconnection networks. Overprovision *connectivity* by increasing the number of paths available, and then run massive amounts of multipath routing, either carefully or randomly. Why not do this for the Internet backbone too? Because overprovisioning connectivity is even more expensive than overprovisioning capacity on the Internet's spatial scale, unless you overlay, as in P2P.

c) *P2P multicast overlay network*: Overprovision *connectivity* rather than capacities, by increasing both the number of paths *and* the number of source nodes, then run massive amounts of multi-tree construction by picking not just routing paths but *source-destination relationships*, carefully or randomly. More than simply creating a richly connected topology and then choosing many paths, this creates many concurrent multicast trees.

The progression of the above three designs can be summarized as follows: (1) Fix a topology, make pipes fatter and use the pipes intelligently. (2) Enrich the topology by increasing connectivity. (3) Create many topologies to choose from at the same time. In the end, (2) and (3) can get close to their bisection bandwidth limit and peer upload capacity limit, respectively, but (1) cannot get close to full utilization of backbone bandwidth. Furthermore, if there is enough overprovisioning of connectivity, you even choose among the connections randomly (like VL2 and Bit Torrent) and be relatively close to the limit. Overprovisioning connectivity pays off better than overprovisioning capacity, if you can *afford* it.

For the Internet backbone, digging trenches is the dominant cost of the Internet Service Providers (ISPs). And links are long haul, covering thousands of miles, and constrained by the presence of fibers and population. It is very expensive to create connectivity. In a data center, the network inside a large building is a relatively small fraction of the cost, compared to server, electricity and cooling costs. So overprovisioning connectivity makes economic sense. P2P is an overlay network, so connectivity is a logical concept and even cheaper to overprovision in that case. And connectivity can be

dynamically managed through control signals without digging any trenches.

In addition to cost structure, traffic demand's predictability and flexibility are other root causes for these fundamentally different choices in network design. In the Internet, traffic matrices are relatively predictable. Traffic matrix fluctuation on the Internet is over time rather than space, and thus can be mitigated by either capacity overprovisioning or time dependent pricing. In data centers, traffic demands are quite volatile and not well understood, another reason to overprovision connectivity. In P2P, one has the option of *changing* the traffic matrix, by choosing different peers, so leveraging that flexibility gives the biggest "bang for the buck."

2) Wireless Network Design: The most general structure for wireless networks is the MANET, in the design of which routing plays a central role. Routing algorithms are usually developed based on geographical distance or metrics considering fading effects, to establish a global end-to-end routing table for each node, which is difficult to maintain even with flooding of query messages in the networks. An immediate option to leverage graph theoretic concepts from social networks is to consider who wants the information prior to determination of the routing. The interested recipients of information might be geographically close or not. For example, on the military battlefield, the same group of geographically-close infantry might want to share some information. However, a tank commander who is farther away might want to share the same information. That is, the routing of messages proceeds via social groups. To implement this concept, Costa et al. introduced socially aware routing by adding an extra layer [40], known as the interest-based routing layer supporting publish-subscribe, to complete the networking functions. The rationale is that socially actively tied nodes (i.e. those with more social interactions) tend to have common interests in information, which may be useful for routing with time-varying node degree (i.e. connections in the technological network). This also suggests that the community structure of a social network on top of a technological network can help the networking functions of this technological network. To address the time variation in implementation, we can form a predictor for social interaction to update routing metrics. Such a predictor is quite similar to estimation and filtering in digital communication systems [41], for example, Kalman filtering in [40]. There are many more examples in which the features and lessons of social network analysis can be applied to design network functions [42], or in which random graph concept can aid in understanding the scaling laws of wireless networks [43].

III. ROBUSTNESS MODEL

A. Scale-free Networks and the Internet's Achilles Heel

1) Power Law Degree Distribution: Beginning in the late 1990s, it came to be believed that the Internet topology exhibits a *power law distribution* for its node degrees [44]. Here, the "topology" of the Internet may mean any of the following three different graphs:

- The graph of webpages connected by hyperlinks.
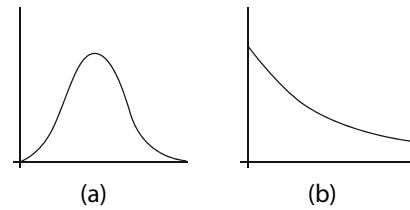- The AS graph connected by peering relationships.

Fig. 3. (a) Gaussian vs. (b) long tailed distribution. The Gaussian distribution has a characteristic scale, *e.g.*, the standard deviation from the mean, whereas the long tailed distribution does not.

- The graph of routers connected by physical links (the focus of this section).

For the AS graph and the router graph, the actual distributions of the node degrees (think of the histogram of the degrees of all the nodes) are not clear due to measurement noise. For example, the AS graph data behind the initial justification of the power law distribution had more than 50% links missing. The Internet exchange points lead to many peering links among ASs, shortcuts to enable settlement-free exchange of Internet traffic, that cannot be readily measured using standard network measurement probes [45]. But suppose we assume that these graphs exhibit a power law distribution of their node degrees. Even then, the actual graphs and their properties can be quite different.

To talk about the Achilles' heel of the Internet, we have to focus on the graph of routers as nodes, with physical links connecting the nodes. No one knows for sure what that graph looks like, so people use proxies to estimate it through measurements like trace-routes. Studies have shown that such estimates lead to a biased sampling, due to the way the Internet protocol reacts to trace-route measurements [46]. In addition, there are other measurement deficiencies arising from IP alias resolution and layer 2 technologies. There is also no scalable measurement platform with enough vantage points at the network edge to detect high-degree nodes there. Therefore, it remains unclear whether the Internet router graph has a power law degree distribution or not. But for now let us assume the conclusion of node degree distribution following the power law holds for the graph of routers.

Many distributions, like Gaussian and exponential distributions, have a characteristic scale, as shown in Figure 3(a). And the probability that a random variable following such a distribution has a value above a given number x, *i.e.*, the *tail probability*, becomes small very quickly as x becomes large. I.e., it is not likely to be too far away from the mean. This leads to what is called a homogeneous network, defined here as a network in which the degrees of the nodes are more or less similar.

In sharp contrast, as shown in Figure 3(b), a long tailed distribution does not have a characteristic scale, leading to the so-called scale free network that is inhomogeneous in its node degrees. The tail probability $\text{Prob}[X \geq x]$ of a long
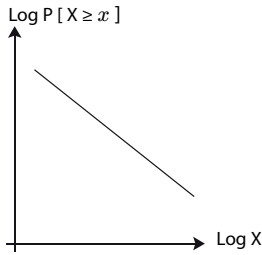
Fig. 4.   The Pareto distribution is a straight line on a log-log plot. This is the visual signature of any long tailed, or power law, distribution.

tailed distribution exhibits the following characteristics:

$$\text{Prob}[X \geq x] \approx kx^{-\alpha},$$

where k is a multiplicative constant, α is the exponent in the exponential decay, and $\approx$ means equal in the limit as x becomes large, *i.e.*, *eventually* following a power law distribution.

The most famous special case of a long tailed distribution is the Pareto distribution, with the following tail distribution for $x \geq k$ [47]:

$$\text{Prob}[X \geq x] = \left(\frac{x}{k}\right)^{-\alpha}. \tag{3}$$

Differentiating the above expression, we see that the Probability Density Function (PDF) of the Pareto distribution also follows a power law, with the power exponent $-(\alpha + 1)$:

$$p(x) = \alpha k^{\alpha} x^{-(\alpha+1)}. \tag{4}$$

In sharp contrast, for any $x \in (-\infty, \infty)$, the Gaussian PDF follows

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

where μ is the mean and σ the standard deviation.

We can plot either the tail distribution (3), or the probability density function (4), on a log-log scale, and get a straight line for the Pareto distribution. This can be readily seen as

$$\log \text{Prob}[X \geq x] = -\alpha \log x + \alpha \log k.$$

The slope is $-\alpha$, as shown in Figure 4. A straight line on a log-log plot is the visual signature of a power law distribution. It has been reported that the power exponent is -2.1 for the in-degree distribution of the webpage graph, -2.4 for the out-degree distribution of the webpage graph, and -2.38 for the Internet router graph.

In 2000, it was suggested in [48] that, unlike networks with an exponential distribution of node degrees, scale free networks are robust to random errors, since the chances are that damaged nodes are not highly connected and thus would not cause too much damage to the network. But a deliberate attack that specifically removes the most highly connected nodes will quickly fragment the network into disconnected pieces. These nodes with large degrees sitting in the center of the network become easy attack targets to break the whole network. Since the Internet, even at the router level, follows a power law degree distribution too, it must be vulnerable to such attacks on its Achilles' heel.

The argument sounds plausible and the implication alarming, except that it does *not* fit reality [49].

2) Internet Reality: Two networks can both have power law degree distributions and yet have very different features otherwise, with very different implications to functional properties, such as robustness to attacks.

For example, what if the high variability of node degrees happens at the network *edge* rather than at the center? That condition is unlikely if networks were randomly generated according to the model of preferential attachment, where nodes attach to more popular nodes. In the set of all graphs with power law degree distributions, degree variability tends to arise out of the difference between highly connected core nodes and sparsely connected edge nodes.

But what if the unlikely topology is the actual design? Cisco cannot make a router that has *both* a large degree and a large bandwidth per connection. Each router has a limitation on its total bandwidth, i.e., on the maximum number of packets it can process at each time, so there is an inevitable tradeoff between the number of ports (node degree) and speed of each port (bandwidth per connection) on the router.

In addition, the Internet takes layers of aggregation to smooth out individual users' demand fluctuations through statistical multiplexing. A user's traffic goes through an access network like WiFi, cellular, DSL, or fiber networks, then through a metropolitan network, then enters the core backbone. A node's bandwidth is related to its placement. The bandwidth per connection in the switches and routers goes up along the way from edge to core.

This implies that nodes in the core of the Internet must have large bandwidth per connection, and thus small degree. For example, the AT&T topology of 2003 shows that the maximum degree of a core router is only 68 while the maximum degree of an edge router is almost five times as large: 313.

In summary, access network node degrees have high variability. Core network node degrees do not. Attacks on high degree nodes can disrupt only access routers and do not disrupt the entire network. Attacks on medium to small degree nodes have a high chance of hitting the access routers because there are many more of them than core routers. That is, the router-level Internet topology does not have an Achilles' heel.

Moreover, there are also protocols that take care of detecting and mitigating failures even when routers are down. It is not just about the connectivity pattern in the graph.

Summarizing, the flaws in "The Internet has an Achilles' heel" thinking are three-fold:

- Incomplete measurements skew the data.
- Degree distribution does not predict robustness.
- Functional protection sits on top of the topology.

3) Functional Model: The Internet might be viewed as "self organizing," but that is achieved by design and protocols based on constrained optimization. There is a way to more precisely define the tradeoff between *performance* and *likelihood* of vastly different topologies all sharing a power law distribution.

One of the several ways to capture the aggregate throughput of the Internet, as its performance metric, is through the
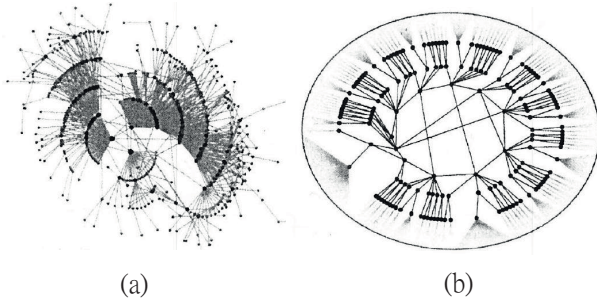
(a)　　　　　　　　(b)

Fig. 5. Achilles' heel or not, from [49]. (a) is a typical topology from a preferential attachment mechanism. (b) is a typical topology from the real Internet. Both satisfy power law degree distribution, but (a) has an Achilles' heel whereas (b) does not.
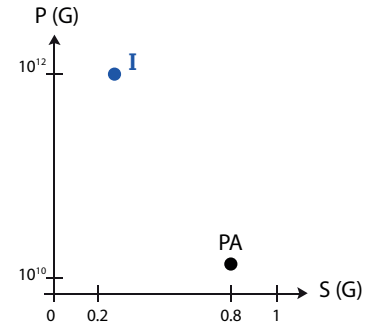

Fig. 6. Performance vs. likelihood of various topologies all with power law distributions. Topology I, which represents the Internet, is much less likely to be chosen if a graph is drawn at random from the set of graphs satisfying a given power law degree distribution. But it also has much higher performance, as measured by total throughput. Topology PA, which represents Preferential Attachment, is much more likely but has much lower performance.

following optimization problem:

$$
\begin{aligned}
&\text{maximize} \sum_i x_i \\
&\text{subject to} \sum_i R_{ki} x_i \le b_k, \quad \forall k \\
&\qquad\quad x_i = \rho y_{S_i} y_{D_i}, \quad \forall i \\
&\text{variables} \ \ \rho, \{x_i\}.
\end{aligned}
\tag{5}
$$

Here, an end-to-end session i's rate x_i is proportional to the overall traffic demand y at its source node S_i and its destination node D_i ($\mathbf{x}$ is generated by $\mathbf{y}$), ρ being the proportionality coefficient and the actual optimization variable. The entry R_{ki} in the routing matrix is 1 if session i passes through router k, and 0 otherwise. The constraint values $\{b_k\}$ capture the router bandwidth-degree feasibility constraint. The resulting optimal value of the above problem is denoted as $P(G)$: the *performance* of the given graph G.

On the other hand, we can define the *likelihood* of a graph G with node degrees $\{d_i\}$ as

$$
S(G) = \frac{s(G)}{s_{max}},
\tag{6}
$$

where

$$
s(G) = \sum_{(i,j)} d_i d_j
$$

captures the pairwise connectivity by summing the degree products $d_i d_j$, over all (i, j) node pairs that have a link between them. And s_{max} is simply the maximum $s(G)$ among all the (simple, connected) graphs that have the same set of nodes and the same node degrees $\{d_i\}$. These $\{d_i\}$, for example, can follow a power law distribution.

We use $P(G)$ to represent performance and $S(G)$ to represent the *likelihood* of a scale free network. As shown in Figure 6, if we were drawing a network from the set of scale free networks at random, high performance topologies like the one exhibited by the real Internet are much less likely to be drawn. But the Internet was not developed by such a random draw. It came through constrained-optimization-based design. Performance of the topology generated by (5) is two orders of magnitude higher than that of the random graph generated by preferential attachment, even though it is less than one-third likely to be chosen if it were chosen at random.

The poor performance of graphs with highly connected core nodes is actually easy to see: a router with a large degree cannot support high bandwidth links, so if it sits in the core of the network it becomes a performance bottleneck.

In summary, the Internet router graph is performance driven and technologically constrained. Concluding that there is an Achilles' heel is a story that illustrates the risk of over-generalizing in a "network science" without domain-specific functional models.

B. Security and Privacy

Current online social networks facilitated by advanced wired/wireless communication media offer a new paradigm of interaction, communication, and sharing media/files/interests. Moreover, current social network sites allow subscribers to create online profiles that can be viewed freely by others. Such a scenario raises two important issues to be immediately addressed:

- Trust evaluation: In human social activities, people make friends according to trust, which is typically evaluated by social distance. Without interacting physically in current online social networks, understanding user behavior and evaluating trust are critical [50].
- Privacy of sensitive information: Since a hacker could easily exploit personal information (such as financial, medical, or location information) of individuals to endanger users or to receive benefits, privacy remains a critical issue to the success of social networks, particularly since individuals offer manage their sensitive profiles in a risky way due to carelessness on lack of awareness of its serious consequences [51].

1) Trust: The typical trust model consists of trust measurement/sampling and evaluation. As we noted above, the challenge with trust is always around how to model and measure uncertainty in social networks. The uncertainty can be generally described by belief, the psychological state in which an individual holds a proposition or premise to be true. Another problem is that it is hard to determine what factors contribute to one's trust state since there is often a lack of obvious clues to track ones trust state and trust behaviors.

The first complete work on modeling belief is that of Cox in 1946 [28], who proposed a reasoning system and proved that it is isomorphic to probability measure. Later, Dempster and Shafer [52], [53] developed the theory of belief functions, which is a method of quantifying uncertainty that generalizes probability theory. In recent decades, the quantitative modeling of uncertainty has been further developed and applied in many areas (sometimes without using the term of belief), including e-commerce [54], peer-to-peer networks [55], and ad hoc and sensor networks [56]–[58]. An important application of trust modeling is in recommendation systems. We may trust someone recommended by trusted friends, which helps us make decisions about products, etc. Josang *et al.* [59] provide a survey of current trust and reputation systems. In the following, we summarize the state of trust modeling in online social networks via two methods: sampling and in/direct trust evaluation.

Sampling:
The empirical study of online social networks has received considerable attention in recent years [50]. The large datacenters and data storage on the cloud creates many opportunities for studying the structure of the online social graph and user activities with abundant data. To exploit this data, efficient and effective sampling methods are required. Practical recommendations for crawling online social networks are given in [60]. Furthermore, [61] provides a framework for sampling of online social networks with multiple relationships, an important feature of social networks.

Trust Evaluation:
To precisely capture the complex interactions among people, online social networks typically adopt indirect trust evaluation by introducing recommendation or confidence values. [62] proposes a trust inference algorithm that uses probabilistic sampling to separately estimate trust information and our confidence in that information. Then the two values are used to compute an estimate of trust based on only those information sources with the highest confidence estimates. Another approach is to introduce a recommendation system to determine reliable third-party information to combine for trust evaluation. [63] builds a social-map-based recommender system, which summarizes users' content-related social behavior over time to offer effective recommendations to new users. [64] further considers the reality that we normally ask trusted friends for recommendations and analyzes trust-based recommendation systems or referral systems. Recommender systems for e-commerce are discussed in Section V.B below.

2) Privacy Issues: Privacy issues in social networks have attracted considerable attention, both in the popular media and in the scientific community[1]. A few key works in the latter category are mentioned in the following paragraphs. This is hardly an exhaustive survey of this field, but these works and others cited therein are representative of some of the main issues, including measurement of privacy leakage and approaches to prevent it.

Reference [66] appears to be the first work to present quantitative metrics for privacy and utility. The main focus of this work is on how to share social networks owned by organizations without revealing the identities or sensitive relationships of the users involved. The authors propose a framework for computing a privacy score of a user, which indicates the potential privacy risk caused by her or his participation in the network. The definition of privacy score used here is based on the following: the more sensitive the information revealed by a user, the higher her or his privacy risk. Also, the more visible the disclosed information becomes in the network, the higher the privacy risk. The authors further develop mathematical models to estimate both sensitivity and visibility of the information. [67] also studies patterns of information revelation in online social networks and their privacy implications. These authors analyze the online behavior of more than 4,000 Carnegie Mellon University students who have joined a popular social networking site that caters to colleges. The authors evaluate privacy leakage in terms of the the amount of personally identifiable information (name, date of birth, phone number, email address, etc.) students disclose and study their usage of the site's privacy settings. The authors highlight potential attacks on various aspects of privacy, and show that only a minimal percentage of users change the highly permeable privacy preferences. There are no mathematical models in this treatment. In a related work, [68] further develops a tool to detect and report unintended information loss in online social networks.

There has also been work on privacy in mobile social networks. For example, [69] considers this situation and presents a practical (implementation/networking) technique that allows location-based services to query local mobile devices for users' social network information, without disclosing user identity or compromising privacy and security. The authors of [70] also consider mobile social networks, developing a solution to allow location-based services to query local mobile devices for users' social network information, without disclosing user identity or compromising privacy and security.

As noted above, [66] considers both the utility and the privacy of social networks, recognizing the fact that these two qualities are connected to one another. Indeed, utility and privacy are opposing goals, as one can maximize either at the expense of minimizing the other. That is, one can make any data source completely private by prohibiting access to it, but this will render it useless. On the other hand, a data source will be most useful in general if it is completely and publicly open, thereby completely sacrificing privacy. Thus, there is clearly a tradeoff between these two goals. A fundamental information theoretic characterization of this tradeoff has been developed in [71] and [72], providing a general principled approach within which privacy problems of data sources of all types, including social networks, can be addressed. This framework allows consideration of a variety of privacy-related issues, including leakage due to the availability of multiple data sources (e.g., multiple social networks or other data-bases), multiple searches or queries on the same source, etc. Further work has applied this framework to situations in which a single source provides information to multiple users, each of which has its own privacy concerns [73], and to situations in which information available locally to one user of a data source must be kept private from other users of the same data source.

---

[1]A survey and taxonomy of privacy risks in social networks, together with existing and proposed policy solutions to dealing with these risks, is provided in [65].

This formalism has not yet been applied specifically to social networks, and this represents a very promising area for future research, particularly for those networks of heterogeneity (say, cyber-physical or individual-technological types of networks).

3) Security, Attacks: This subsection examines security attacks that exploit credulity and other aspects of human nature in social networks [51].

a) *Phishing*: By impersonating a trustworthy entity, an adversary could launch phishing attacks, through which the attacker attempts to acquire sensitive information from a victim fraudulently. With information obtained from social networks, phishing has been shown to be four times more effective than traditional blind attempts [74].

b) *Spam*: Because users trust service providers, invitations and requests from service providers are more easily accepted than other messages. Spammers can either exploit sensitive information of users to deliver messages to which users may be particularly susceptible or generate misleading links or information that redirect users to the spammer's website.

c) *Malware*: Malware such as computer viruses and worms have attracted considerable attention. Such attacks rely on acquaintances among individuals and particularly in the form of social networks. The propagation or spread dynamics of malware have been widely investigated, e.g., in [75] and [76]. Epidemics, as discussed in Section V.A-2, are widely considered in modeling and subsequent defense strategies against malware.

d) *Inference*: Another new threat is to (statistically) infer private or secure information from social networking data, such as age privacy leakage by estimation of data from Facebook [77]. Interference attack treats can be strengthened due to (i) multiparty access to shared information, (ii) one party to access associated multiple social networks or databases, or (iii) heterogeneity of networking structure that is common in social networks consisting of individuals and multiple technological networks for these individuals. Current sanitization approaches include tags for data attributes, access control mechanisms, and k-degree l-diversity anonymity models to protect sensitive data [78]–[80].

Finally, as indicated earlier in this section, the Internet is robust due to its functional domain capability to enhance robustness [49]; we can therefore design technological networks from this point of view. Intentional attacks are known to be effective in disintegrating a network by paralyzing some fraction of nodes with the highest degree [48], [81]. A network is expected to recover from the temporal malfunction as long as most of the nodes are still connected [82], which coincides with the percolation phenomenon in statistical physics. A network is connected in the percolation sense if there exists a giant connected component containing a majority of nodes and therefore the percolation-based connectivity can be used to evaluate the vulnerability of the technological network under intentional attack. An intentional attack on the technological communication network is regarded as effective if it is able to divide the network into several small components so that the network fails to be connected in a percolation sense. Further attacks to leverage the heterogeneity of the network structure,

say the network of individuals and physical communication network for social networks, or cyber-physical systems, are therefore possible [77]. A fusion-based defense mechanism (see Section II) can be used to counter such attacks, as the fusion center can combine local intrusion detection results from each node to ensure network robustness in the percolation sense while the network is under attack [83], and game theory can be further used to devise mechanisms to defend against such attacks [77].

IV. COMMUNICATION SYSTEM ARCHITECTURE

A. Small-World Networks

The small-world phenomenon, which refers to the fact that most of us are linked by short chains of acquaintances, was initially investigated in sociology, and then later in other fields [1], [84], [85]. The intuition behind this phenomenon suggests that (a) such short chains are ubiquitous and (b) individuals with purely local information can adapt to access these chains. A characteristic feature of small-world networks is that their diameters tend to be exponentially smaller than their size, being bounded by a polynomial in $\log N$, where N is the number of nodes. In other words, there always exists a short path between any two nodes in such a network. The small-world phenomenon suggests that efficient navigation and search is possible in a large-scale social network, and so is efficient routing in large-scale technological networks.

1) Information Dynamics: Lying atop technological communication networks, large social networks actually serve as conduits for information flow. Such information flow can be termed *information dynamics*. Of immediate interest in this context is to measure the *efficiency* of such networks for exchanging information. We can study this issue as follows. Among many ways of constructing small-world networks, we follow the most widely adopted construction [2], in which a graph G has N vertices and K edges. G is un-weighted, sparse (i.e. $K \ll N(N-1)/2$), and connected. In other words, there exists at least one path connecting any two vertices with a finite number of steps. The shortest path length (or distance metric) ρ_{ij} between node i and node j is the smallest sum-distance among all possible paths between node i and node j. ρ_{ij} is thus calculated from the distance matrix and the adjacency matrix. The efficiency of communication is thus inversely proportional to the shortest path length; i.e., the average efficiency of G can be defined as

$$\mathsf{E}(G) = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{\rho_{ij}}. \tag{7}$$

The above quantity can be viewed as the *global efficiency* of G, $\mathcal{E}_{\text{global}}$. Similarly, we can characterize local properties by looking into the subgraph G_i of the neighbors of a node i. The local efficiency, somewhat similar to the clustering coefficient, is defined as the average efficiency of the local subgraph, i.e.,

$$\mathcal{E}_{\text{local}} = \frac{1}{N} \sum_i \mathsf{E}(G_i). \tag{8}$$

Small world networks have both high global efficiency and high local efficiency [86]. Based on the study in [86], the

technological network Internet and the virtual network WWW have similar global efficiencies, but the WWW has clearly higher local efficiency, which illustrates the important role of community in social networks (as noted in Section III.B).

Information usually spreads as the result of discrete communication events, such as email messages, phone calls and text messages, teleconferences, blogs, or sharing of media. Thus we may reasonably consider information flow over social networks as *event-driven* communications. Kossinets, Kleinberg and Watts, proposed a framework of information pathways for a social communication network [87], based on inferring structural measures from the potential for information to flow between nodes. By recording the time index of each communication event, the *communication skeleton* is defined as a graph G on V with an edge (v_i, v_j) if v_i sends at least one message to v_j during the observation period (say, $[0, T]$). Using algorithms to compute the vector clocks for nodes in $[0, T]$, we may examine latencies from social network data and information dynamics.

A significant advantage of small-world networks lies in search. Watts, Dodds and Newman proposed a model based on social structure to explain this property in the context of social networks in [88], which has six contentions:

(i) Individuals in social networks are endowed with network ties and identities.

(ii) Individuals hierarchically and successively partition into a large number of specific groups.

(iii) There is more than one way to partition into groups.

(iv) Group membership is a primary basis for social interaction, in addition to the formation of identities.

(v) For perception of similarity with other nodes in social networks, a measure of "social distance" can be constructed; however it may not be a metric typically due to the triangle inequality.

(vi) Given only local information about the network, individuals forward a message to a single neighbor.

Two kinds of partial information are available to individuals: social distance and network paths. Through a greedy algorithm as suggested by Milgram, it can be shown that searchability is a generic property in real-world social networks, and the mean search step grows slowly with the population size as the "small"-world property suggests. It is promising to generalize to social networked databases and to technological networks, though the absence of a unique definition of "social distance" in the process of classification can complicate this generalization as indicated in the earlier discussion of community detection in Section III. Email has been used as an important tool to study information flow in online social networks for some time [89], [90], whereas [91] and [92] explored the impact of social group mechanisms and passing strategies on information flow.

The recognition of identity can be generalized in social networks. Recognition of user behavior beyond identity may more directly meet practical needs in e-commerce with privacy. Consequently, we can generalize interaction between social networks and technological networks in Figure 1 to the layer structure in the Figure 7 by including recognition, which can be facilitated by communication theoretic decision and machine intelligence.

The meaning of the cognition layer in Figure 10 can be viewed in the following way. In *social cognitive theory*, behavior is determined by expectations and incentives of the following three types, as the foundation for designing online social networks [93]:

i) Expectations about environmental cues (that is, beliefs about connected events).

ii) Expectations about the consequences due to one's own actions (that is, opinions about individual behaviors likely to influence outcomes), also known as *outcome expectation*.

iii) Expectations about one's own competence for performing behaviors to influence outcomes, also known as *efficacy expectation* (i.e. self-efficacy).

Incentives refer to the values of particular objects or outcomes. Behaviors are regulated by recognizable consequences. Online services should be designed by recognition of social networks that overlay the technological networks.

The early study of information dynamics was about the spread of rumors. Suppose that one of n people knows a message. At the first stage, this person passes this message to someone chosen at random. At each stage, each person who already knows the message communicates it to a person chosen at random, independently of all other present and past choices. S_n denotes the random number of stages before all n people know the message. Without considering network structure, it is shown in [94] that in probability,

$$S_n \to \log_2 n + \log n. \tag{9}$$

On letting $N(t)$ denote the number of people who know this message after t stages, we may approximate the behavior of $N(t)$ by the recursion [95]

$$N(t+1) = n - [n - N(t)]e^{-N(t)/n}. \tag{10}$$

A similar but more useful model for small-world networks, such as social networks or certain technological communication networks, is studied in [35]. For this model, nodes are distributed randomly in a square such that nodes can communicate directly with any other nodes within a radius r; moreover, each node can also communicate with another node randomly selected from all nodes The average message delivery time (i.e. the number of hops for a message to get from one point to another) from a randomly selected node is of interest. It is shown that this quantity, which is the average number of degrees of separation, becomes essentially constant for large separation, which is the small-world phenomenon. Furthermore, consistent with Milgram's famous experiment, shorter social distances are traversed linearly, while longer social distances are traversed in saturation. This analysis can be applied to sociology [96] as another example of the interplay between social networks and technological networks.

2) Information Epidemic: Although general information dynamics over networks are not straightforward, the study of epidemics is useful for understanding such dynamics. Epidemics have been studied as dynamical systems using differential equations [97]. To study epidemics, agents (or nodes) can be categorized into three types: susceptible (S), infected (I), and recovered (R). For information epidemics, an SI model using the following differential equations might
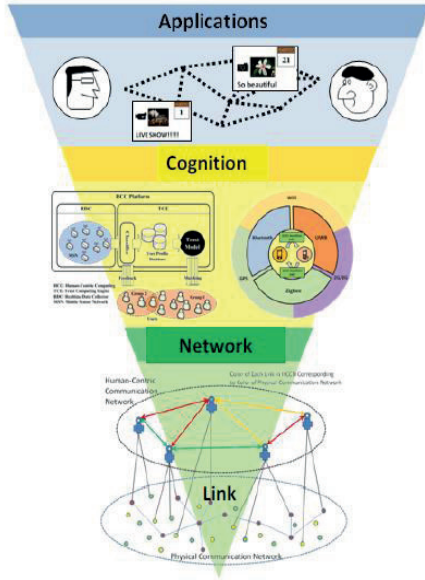
Fig. 7.   Layered structure of general social networks.

fit well. An SI model allows population in two states, say, with the message or without the message. The dynamics of the number of nodes in these states can be modelled as follows:

$$\frac{dS(t)}{dt} = -\beta S(t)I(t)$$
$$\frac{dI(t)}{dt} = \beta S(t)I(t) \qquad (11)$$

On the other hand, SIR models using the following differential equations usually fit well for malware, Internet worms, and virus propagation:

$$\frac{dS(t)}{dt} = -\beta S(t)I(t)$$
$$\frac{dI(t)}{dt} = \beta S(t)I(t) - \gamma I(t)$$
$$\frac{dR(t)}{dt} = \gamma I(t). \qquad (12)$$

The above population models can be refined by considering network structure. Epidemics on small-world networks were first explored in [95] and [98], which provide models that are useful for information dissemination. Dynamics of rumor spreading in networks have been studied via information rate in [99]. Information dissemination over ad hoc networks of opportunistic links was studied in [100]. Information dissemination is useful in mobile social networks, such as described in [101] and [102]. Applying gossip algorithms [103], [104], voter algorithms, consensus building algorithms [105], and information casecade [106], to message passing, cellular network design, etc. are other good examples to connect social network concepts with technological networks.

We noted above that epidemics can be adapted to model malware propagation. An interesting model related to small-world networks was proposed in [107] which models malware for smartphones coming from two directions: short range

wireless like Bluetooth and WiFi, or long-haul wireless like cellular. The epidemics thus involve both spatial and social interaction networks, and the study of such behavior relies on simulations. Consequently, a variant set of epidemic differential equations was developed in [108] to precisely quantify epidemic information dynamics over small-world-like social networks in a computationally efficient way.

B. Interplay with Technological Networks

The study of social networks and associated analysis can be very useful in the evolution of the design of technological networks. We elaborate on this concept as follows.

1) Large Wireless Ad Hoc Networks: According to the analysis of small-world phenomena [1], [84], [109], [110], even adding a small number of random (social) connections to a highly clustered network results in short paths appearing between most pairs of network nodes. For realistic applications, the challenge to apply this small-world phenomenon lies in how to find such a short-path connection for an individual (or a network node) in search or similar tasks. We need a global view to identify such path(s) based on local information (i.e. knowledge about neighbors at most). To address this issue, we may introduce the concept of *gravitation laws* in social networks; namely, the probability of being friends with a friend at a distance d decays according to d^{-r} for some power r. The exponent r controls the nature of the social network; for larger r, the network is less random or more geographically clustered. By using Milgram-style greedy search with local information, the effectiveness of search increases when r increases, but finally gets worse as r keeps increasing since short-path connections are difficult to find. Interestingly, $r = 2$, just as in a gravitation field, is found to be a good choice in many experiments.

Search or navigation in a social network plays a key role under many scenarios, and is also critical to networking functions in wireless (and wired) technological networks, such as routing. From the above discussion, we may conclude the following heuristic rules in realistic network design, for both social networks and technological networks:

- Properly building short-path connections results in small-world networks.
- Each short-path connection prefers being connected to a locally large component.

A good task to which the small-world phenomenon can be applied in technological networks is routing. This approach could lead to efficient network architectures for large ad hoc networks, which is an important technology not only in military communications, but also in resilient commercial networks, sensor networks, and M2M communications in cyber-physical systems with potentially trillions of wireless devices [111]–[114].

Figure 8 depicts the abstract topology of cloud-based M2M communications. When data aggregators (DAs) have certain paths through wireless infrastructure, say cellular or short range wireless communications, machines or sensors in the swarm want to find the way to send message(s) to a DA that suggests a short path. Finding such short paths for machines suggests formation of small-world networks to support quality
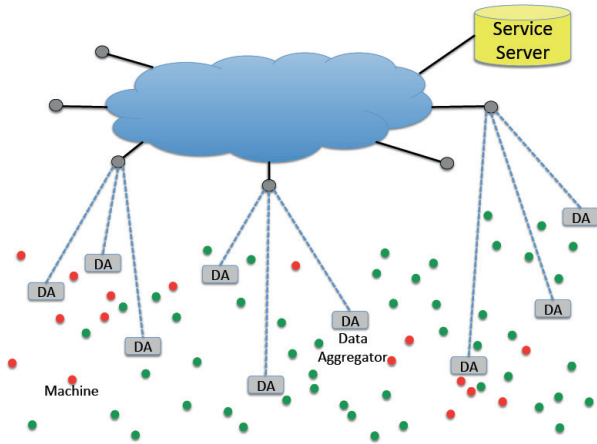
Fig. 8. Machine-to-machine communications showing a cloud with a two-tier network architecture.

of service [115], spectral efficiency (message throughput per unit of network bandwidth), and energy efficiency. Technical questions related to this two-tier architecture include:

- What is the number of hops, say from a machine to the service server?
- What is the message delivery delay?
- How does a machine find the data aggregator to reach the cloud, that is, what is this routing algorithm?
- What is the best placement of data aggregators among machines/sensors?
- What is the appropriate access technology to DAs?
- What are the information dynamics and thus what is the appropriate network design, if there are multiple kinds of sensors/machines or competitive/collaborative information/messages?
- What is the appropriate transmission power (i.e. r in the gravitation law) with or without cooperation (also Section VI)?

Many challenges in the above list arise more generally in large (mobile) ad hoc networks, and researchers have already been explicitly or implicitly heading in this direction. For example, Daly and Haahr showed that applying graph theoretic analysis for social networks helps develop more effective designs for MANETs [116]. As MANETs are usually not fully connected, efficient delay-tolerant message delivery was examined. Since centrality presents a way to quantify relative importance of a node in a social network [117], network centrality measures for information flow in a MANET are used to calculate the (social) metrics of routing algorithms. A more detailed survey of social-based routing in delay-tolerant networks can be found in [118].

All of the above challenges of applying social network understanding to technological network design become even more challenging if we add spectrum sharing [119] or quality-of-service (QoS) issues into these scenarios. Femtocell networks, relay/mesh networks or other two-tier heterogeneous networks [120], [121] represent other potential application scenarios.

2) In-Network Computation: In large multi-hop information-gathering networks, appropriate information aggregation is possible in order to provide advantages for communications such as a reduction in the number of messages in the network, more reliable network operation, etc. as noted in [122]–[124]. This can be viewed as a tradeoff between computation and communications. One aspect of in-network computation relating to correlation among observations can be traced back to a well-known branch of information theory, source coding of correlated sources [124]–[126]. In large ad hoc networks, particularly those using opportunistic cooperative networking, information sensing may consume communication bandwidth [127] and energy. Conversely, a node has to compute to determine relaying strategies [128]. A different idea is to apply heterogeneous information fusion and inference (HIFI) to spectrum sensing in cognitive radio networks [129]. In this approach, a new spectrum sensing technique is proposed to identify spectrum opportunities for cognitive radios by achieving sensing diversity through just a single sensor. The sensor in such spectrum sensing simultaneously detects the primary transmitter's data transmission and primary receiver's feedback message transmission; thus, this spectrum sensing equivalently implements receiving diversity by creating multiple kinds of observations solely from a single terminal/sensor. It is shown that the resulting spectrum sensing, even with a single sensing node, can achieve and transcend the performance of traditional cooperative spectrum sensing that requires multiple sensors.

3) Synchronization and Control of Networks: It is necessary to disseminate information to network nodes in order to control various technological networks, such as mobile ad hoc networks, distributed systems, and P2P multimedia networks [130]–[133]. This task could be accomplished through the use of information epidemics to reach consensus [134]. To reach consensus and thus collaboration in systems and networks, we can apply methods of information dissemination such as gossip algorithms or voter algorithms to establish synchronization in networks. This idea was illustrated in [135] for network synchronization for 3GPP femto-networks using an improved voter algorithm.

V. OPTIMIZATION OF NETWORKS

In this section, we explain the algorithmic connection between power control in CDMA cellular networks and pagerank in Google searches.

A. Common Methodology in Pagerank and Power Control

1) Distributed Power Control in Wireless Networks: Consider N pairs of transmitters and receivers. Each pair is a logical channel, indexed by i. The transmit power of the transmitter of link i is p_i. The transmitted power impacts both the received power at the intended receiver and the received interference at the receivers of all the other pairs.

Now, consider the channel from the transmitter of link (*i.e.*, transmitter-receiver pair) j to the receiver of link i, and denote the *channel gain* by G_{ij}. Since G_{ii} is the direct channel gain, the bigger it is the better, since it is the channel for the intended
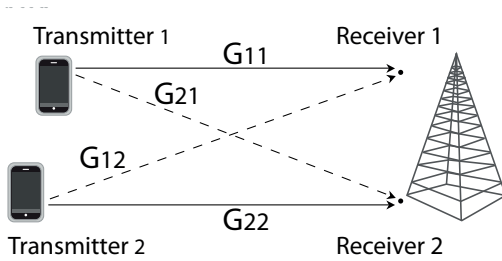
Fig. 9. Uplink interference between two mobile stations at the base station. We can think of the base station as two receivers collocated. G_{11} and G_{22} are direct channel gains, so the bigger the better. G_{12} and G_{21} are interference channel gains, so the smaller the better. Transmit powers are denoted by p_i and received signal-interference-ratio by SIR_i.

transmission for transmitter-receiver link i. All the other G_{ij}'s, for j not equal to i, are gains for interference channels, so the smaller they are the better. This notation is visualized in Figure 9 for a simple case of two mobile stations (MSs) talking to a base station (BS), which can be thought of as two different "logical" receivers physically located together.

We can write the signal-to-interference ratio (SIR), a unitless ratio, at the receiver on logical link i, as

$$\text{SIR}_i = \frac{G_{ii}p_i}{\sum_{j \neq i} G_{ij}p_j + n_i}, \qquad (13)$$

where p_i is the transmit power at transmitter i and n_i is the noise power at receiver i. For proper decoding of packets, the receiver needs to maintain a target level of SIR. We will denote this target as γ_i for link i, so that we want $\text{SIR}_i \geq \gamma_i$ for all i. It is clear that increasing p_1 raises the SIR for receiver 1 but lowers the SIR for all other receivers.

We assume that time is divided into discrete slots, each indexed by t. At each time t, the receiver on link i can measure the received SIR readily, and feed that number $\text{SIR}_i[t]$ back to the transmitter. The dynamic power control (DPC) algorithm can be described through a simple, one line equation: each transmitter multiplies the current power level $p_i[t]$ by the ratio between the target SIR, $\gamma_i[t]$, and the current measured SIR_i, to obtain the new power level to use in the next timeslot [136]:

$$p_i[t+1] = \frac{\gamma_i}{\text{SIR}_i[t]}p_i[t], \quad \text{for each } i. \qquad (14)$$

We see that each user i needs to measure only its own SIR at each iteration, and remember its own target SIR. There is no need for passing any control messages around, such as telling other users what power level you are using.

Intuitively, this algorithm makes sense. First, when the iterations stop, i.e., converge to an equilibrium, that means $\text{SIR}_i = \gamma_i$ for all i.

Second, there is hope that the algorithm will actually converge, based on the direction in which the power levels are moving. The transmit power moves up when the received SIR is below the target, and down when it is above the target. Of course, proving that convergence will happen is not as easy. As one transmitter changes its power, the other transmitters are doing the same, and it is unclear what the next timeslot's SIR values will be. In fact, this algorithm does *not* converge if too many γ_i are too large, i.e., when too many users require large SIRs as their targets.

Third, if satisfying the target SIRs is the only criterion, there are many transmit power configurations that can do that. If $p_1 = p_2 = 1$ mW achieves these two users' target SIRs, $p_1 = p_2 = 10$ mW will do so too. We would like to choose the configuration that uses the least amount of power; i.e., we want a power-minimal solution. And the algorithm above seems to be adjusting power lower when high power is unnecessary.

In general, the questions of "will it converge" and "will it converge to the right solution" are the top two questions that we would like to address in the design of all iterative algorithms. Of course, what "the right solution" means will depend on the definition of optimality [137]. In this case, power-minimal transmit powers that achieve the target SIRs for all users are the "right solution." Power minimization is the objective and achieving target SIR for all users is the constraint. The following is the optimization problem of varying transmit power to satisfy fixed target SIR constraints and minimize the total power:

$$\begin{array}{ll} \text{minimize} & \sum_i p_i \\ \text{subject to} & \text{SIR}_i(\mathbf{p}) \geq \gamma_i, \ \ \forall i \\ \text{variables} & \mathbf{p}. \end{array} \qquad (15)$$

We can represent the target SIR constraints in problem (15) as

$$\mathbf{p} \geq \mathbf{D}(\gamma)\mathbf{F}\mathbf{p} + \mathbf{v},$$

and further group all the terms involving the variables $\mathbf{p}$. The linear programming problem then becomes

$$\begin{array}{ll} \text{minimize} & \mathbf{1}^T \mathbf{p} \\ \text{subject to} & (\mathbf{I} - \mathbf{D}(\gamma)\mathbf{F})\mathbf{p} \geq \mathbf{v} \\ \text{variables} & \mathbf{p}, \end{array} \qquad (16)$$

where $\mathbf{1}$ is a vector of 1s, so the objective function is simply $\sum_i p_i$, $\mathbf{I}$ is the identity matrix, $\mathbf{D}(\gamma)$ is a diagonal matrix with the diagonal entries being the target SIR values $\{\gamma_i\}$, $\mathbf{F}$ is a matrix capturing the given channel conditions: $\mathbf{F}_{ij} = \frac{G_{ij}}{G_{ii}}$ if $i \neq j$, and 0 along the diagonal: $\mathbf{F}_{ii} = 0$, and the constant vector $\mathbf{v}$ captures normalized noise:

$$\mathbf{v} = \left(\frac{\gamma_1 n_1}{G_{11}}, \frac{\gamma_2 n_2}{G_{22}}, \dots, \frac{\gamma_N n_N}{G_{NN}} \right)^T.$$

Linear programming problems are conceptually and computationally easy to solve in general. This special case here has even more structure. The $\mathbf{DF}$ matrix is a non-negative matrix; i.e., all entries of the matrix are non-negative numbers. If the largest eigenvalue $\rho(\mathbf{DF})$ is less than 1, then the following three statements are true:

(a) We can guarantee that the set of target SIRs can indeed be achieved simultaneously, which makes sense since $\rho(\mathbf{DF}) < 1$ means that the $\{\gamma_i\}$ in $\mathbf{D}$ are not "too big," relative to the given channel conditions captured in $\mathbf{F}$.

(b) We can invert the matrix defining the linear constraints in our optimization problem (16) to compute $(\mathbf{I} - \mathbf{DF})^{-1}\mathbf{v}$. But of course there is no easy to way to directly run this matrix inversion distributively across the MSs.

(c) The inverse of the matrix can be expressed as a sum of terms, each term a multiplication of matrix $\mathbf{DF}$ by itself. More precisely: $(\mathbf{I} - \mathbf{DF})^{-1} = \sum_{k=0}^{\infty}(\mathbf{DF})^k$. This is an infinite sum, so we are saying that the partial sum $\sum_{k=0}^{K}(\mathbf{DF})^k$ will

converge as K becomes very large. Furthermore, the tail term in this sum, $(\mathbf{DF})^k$, approaches 0 as k becomes large:

$$\lim_{k \to \infty} (\mathbf{DF})^k = 0. \tag{17}$$

The key insight is that we want to invert the matrix $(\mathbf{I} - \mathbf{DF})$, because that will lead us to a power-minimal solution to achieving all the target SIRs:

$$\mathbf{p}^* = (\mathbf{I} - \mathbf{DF})^{-1}\mathbf{v} \tag{18}$$

is a solution of problem (15), *i.e.*, for any solution $\hat{\mathbf{p}}$ satisfying the constraints in problem (15), $\mathbf{p}^*$ is better

$$\hat{\mathbf{p}} \geq \mathbf{p}^*. \tag{19}$$

Now, the matrix inversion step is not readily implementable in a distributed fashion, but fortunately, it can be achieved by applying the following update.

(1) First, $\mathbf{p}^* = (\mathbf{I} - \mathbf{DF})^{-1}\mathbf{v}$ can be represented as a power series, as stated in Statement (c) above, i.e.

$$\mathbf{p}^* = \sum_{k=0}^{\infty} (\mathbf{DF})^k \mathbf{v}. \tag{20}$$

(2) Then, we can readily check that the following iteration over time gives exactly the above power series (20), as time t increases:

$$\mathbf{p}[t+1] = \mathbf{DFp}[t] + \mathbf{v}. \tag{21}$$

(3) Finally, on rewriting the vector form of update equation (21) in scalar form for each transmitter i, and it can be seen that it is exactly the DPC distributed algorithm (14).

What we saw in DPC is an iterative *power method* (the word "power" here has nothing to do with transmit power, but instead refers to raising a matrix to some power), a common method to develop iterative algorithms arising out of a linear systems model. First we formulated the problem as a linear program, then defined the solution of the problem as the solution to a linear equation, implemented the matrix inversion through a sequence of matrix powers, turned each of those steps into an easy computation at each time slot, and thereby achieved the matrix inversion through iteration in time. Indeed, the same algorithm is used in Google's pagerank algorithm for ranking webpages.

2) Google's Pagerank: We will be constructing several related matrices: $\mathbf{H}, \hat{\mathbf{H}}$, and $\mathbf{G}$, step by step (this matrix $\mathbf{G}$ is not the channel gain matrix now; it denotes the Google matrix, to be defined below [138], [139]). Eventually we will be computing an eigenvector of $\mathbf{G}$ as the importance score vector. Each matrix is $N \times N$, where N is the number of webpages. These are extremely large matrices, considering that there were about $N = 40$ billion webpages in 2011.

The first matrix we define is $\mathbf{H}$: its (i, j)th entry is $1/O_i$, where O_i is the number of outgoing links from node i, if there is a hyperlink from webpage i to webpage j, and 0 otherwise. This matrix describes the network topology, i.e., which webpage points to which webpage. It also evenly spreads the importance of each webpage among its outgoing neighbors, i.e., the webpages to which it points.

Let π be an $N \times 1$ column vector denoting the importance scores of the N webpages. We start by guessing that the consistent score vector is $\mathbf{1}$ (a vector of all 1s), which assumes that each webpage is equally important. This provides an initial vector $\pi[0]$.

From this initial point we iteratively multiply π^T on the right by the matrix $\mathbf{H}$. This has the effect of spreading the importance score from the last iteration evenly among the outgoing links, and re-calculating the importance score of each webpage in this iteration by summing up the importance score from incoming links. If we index the iterations by k, the update at each iteration is simply

$$\pi^T[k] = \pi^T[k-1]\mathbf{H}. \tag{22}$$

Do the iterations in (22) converge? The answer is "not quite yet." We need two adjustments to $\mathbf{H}$.

First, some webpages do not point to any other webpages. These are "dangling nodes" in the hyperlink graph. There might not be a set of consistent ranking scores. One solution is to replace each row of 0, like the last row in $\mathbf{H}$ above, with a row of $1/N$. Intuitively, this is saying that even if a webpage does not point to any other webpage, we will force it to spread its importance score evenly among all the webpages.

Mathematically, this amounts to adding a matrix $\frac{1}{N}(\mathbf{w1}^T)$ to $\mathbf{H}$, where $\mathbf{1}$ is simply a vector of 1s, and $\mathbf{w}$ is a vector with the ith entry being 1 if webpage i points to no other webpages (a dangling node) and 0 otherwise (not a dangling node). The resulting matrix

$$\hat{\mathbf{H}} = \mathbf{H} + \frac{1}{N}(\mathbf{w1}^T),$$

has all non-negative entries and each row adds to 1. So we can think of each row as a probability vector, with the (i, j)th entry of $\hat{\mathbf{H}}$ indicating the probability that a browser currently on webpage i will click on a link and go to webpage j.

Second, there might be many consistent score vectors all compatible with a given $\hat{\mathbf{H}}$. One solution to this problem is to add a little *randomization* to the iterative procedure and the recursive definition of importance. Intuitively, we say there is a chance of $(1 - \theta)$ that you will be jumping to some other random webpage, without clicking on any of the links on the current webpage.

Mathematically, we add yet another matrix $\frac{1}{N}\mathbf{11}^T$, a matrix of 1s scaled by $1/N$ (and clearly a rank-1 matrix), to $\hat{\mathbf{H}}$. But this time it is a *weighted* sum, with a weight $\theta \in [0, 1]$, where, again, $(1 - \theta)$ quantifies how likely it is to randomly jump to some other webpage. The resulting matrix is called the *Google matrix*:

$$\mathbf{G} = \theta\hat{\mathbf{H}} + (1 - \theta)\frac{1}{N}\mathbf{11}^T. \tag{23}$$

Now we can show that, independently of the initialization vector $\pi[0]$, the iterative procedure

$$\pi^T[k] = \pi^T[k-1]\mathbf{G} \tag{24}$$

will converge as $k \to \infty$, and converge to the unique vector π^* representing the consistent set of importance scores. Obviously, π^* is a left eigenvector of $\mathbf{G}$ corresponding to an eigenvalue of 1:

$$\pi^{*T} = \pi^{*T}\mathbf{G}. \tag{25}$$

One can then normalize π^*: take $\pi_i^*/\sum_j \pi_j^*$ as the new value of π_i^*, and rank the entries in descending order, before

outputting them on the search result webpage in that order. The matrix $\mathbf{G}$ is designed such that there is a solution to (25) and that (24) converges from any initialization.

3) A Parallel: As the above discussion indicates, pagerank is similar to distributed power control. They both apply the power method to solve a system of linear equations. The matrix $\mathbf{G}$ in pagerank is much larger, but centralized computation is practical for search. The solutions to those equations capture the right engineering configuration in the network, whether that is the relative importance of webpages in a hyperlink graph or the best transmit power vector in a wireless interference environment. This conceptual connection can be sharpened to an exact, formal parallel below.

First, we can rewrite the characterization of π^* as the solution to the following linear equation (rather than as the dominant left eigenvector of the matrix $\mathbf{G}$ (25), the viewpoint we have been taking thus far):

$$(\mathbf{I} - \theta\mathbf{H})^T\pi = \mathbf{v}. \qquad (26)$$

Compare (26) with the characterization of optimal power vector in distributed power control:

$$(\mathbf{I} - \mathbf{DF})\mathbf{p} = \mathbf{v}.$$

Of course, the vectors $\mathbf{v}$ are defined differently in these two cases: based on webpage viewing behavior in pagerank and receiver noise in power control. But we see an interesting parallel: the consistent score vector π, and the optimal transmit power vector $\mathbf{p}$ are both solutions to a linear equation with the following structure: an identity matrix minus a scaled version of the network connectivity matrix.

In pagerank, the scaling is done by one scalar θ, and the network connectivity is represented by the hyperlink matrix $\mathbf{H}$. This makes sense since the key factor here is the hyperlink connectivity pattern among the webpages.

Alternatively, in power control, the network connectivity is represented by the normalized channel gain matrix $\mathbf{F}$. This makes sense since the key factor here is the strength of interference channels.

Thus, in power control, the scaling is done by many scalars in the diagonal matrix $\mathbf{D}$: the target SIR for each user. To make the parallelism exact, we can also think of a generalization of the Google matrix $\mathbf{G}$ where each webpage has its own scaling factor θ.

The general theme for solving these two linear equations can be stated as follows. Suppose we want to solve a system of linear equations $\mathbf{Ax} = \mathbf{b}$ but do not want to directly invert the square matrix $\mathbf{A}$. We might be able to split $\mathbf{A} = \mathbf{M} - \mathbf{N}$, where $\mathbf{M}$ is invertible and its inverse $\mathbf{M}^{-1}$ can be much more easily computed than $\mathbf{A}^{-1}$.

The following *linear stationary iteration* over time indexed by k [140]:

$$\mathbf{x}[k] = \mathbf{M}^{-1}\mathbf{N}\mathbf{x}[k-1] + \mathbf{M}^{-1}\mathbf{b}$$

will converge to the desired solution:

$$\lim_{k\to\infty}\mathbf{x}[k] = \mathbf{A}^{-1}\mathbf{b},$$

from any initialization $\mathbf{x}[0]$, provided that the largest eigen-

value of $\mathbf{M}^{-1}\mathbf{N}$ is smaller than 1. Both DPC and pagerank are special cases of this general algorithm.

But we still need to show that (26) is indeed equivalent to (25): a π that solves (26) also solves (25), and vice versa. First, starting with a π that solves (26), we can easily show the following string of equalities:

$$\begin{aligned}
\mathbf{1}^T\mathbf{v} &= \mathbf{1}^T(\mathbf{I} - \theta\mathbf{H})^T\pi \\
&= \mathbf{1}^T\pi - \theta(\mathbf{H1})^T\pi \\
&= \mathbf{1}^T\pi - \theta(\mathbf{1} - \mathbf{w})^T\pi \\
&= \pi^T(\theta\mathbf{w} + (1-\theta)\mathbf{1}),
\end{aligned}$$

where the first equality uses (26) and the third equality uses the fact that summing each row of $\mathbf{H}$ gives a vector of 1s (except for those rows corresponding to dangling webpages). The other two equalities are based on simple algebraic manipulations.

But $\mathbf{1}^T\mathbf{v} = 1$ by design, so we know

$$\pi^T(\theta\mathbf{w} + (1-\theta)\mathbf{1}) = 1.$$

Now we can readily check that $\pi^T\mathbf{G}$, using its definition in (26) and the above equation, equals $\theta\pi^T\mathbf{H} + \mathbf{v}$.

Finally, using one more time the assumption that π satisfies (26), *i.e.*, $\mathbf{v} = (\mathbf{I} - \theta\mathbf{H})^T\pi$, we complete the argument:

$$\pi^T\mathbf{G} = \theta\pi^T\mathbf{H} + \pi^T(\mathbf{I} - \theta\mathbf{H}) = \theta\pi^T\mathbf{H} - \theta\pi^T\mathbf{H} + \pi^T = \pi^T.$$

Therefore, any π solving the linear equation (26) is also a dominant left eigenvector of $\mathbf{G}$ that solves (25). And vice versa that a π solving (25) also solves (26), can be similarly shown.

B. Recommender Systems

A recommender system is a critical component in e-commerce to recommend media or documents for users' rental, purchase, and reading, and as a marketing tool for targeted advertising. A recommender system makes a prediction of any user's "personal" favorites based on either a computationally difficult large database or very limited information (i.e. a cold start). Since the Netflix Prize competition announced in October 2006, ample design experience has been tested using real data and shared in the research community. It is related to the data analysis and inference discussed in Appendix A, and additional techniques will be discussed in this subsection.

Recommender systems rely on different types of data, a vector $\boldsymbol{p}_u \in \mathbb{R}^l$ representing user u and another vector $\boldsymbol{q}_i \in \mathbb{R}^l$ representing items of interest such as movies, music, etc. Such data may be from *explicit feedback* such as users' rating, or *implicit feedback* inferred from user behavior such as purchase history, search patterns, etc. The elements of $\boldsymbol{q}_i$ measure the extent to which the item possesses these factors, positive or negative. The elements of $\boldsymbol{p}_u$ measure the level of user u's interests in these items, positive or negative. User u's rating of item i is estimated as

$$\hat{r}_{ui} = \boldsymbol{q}_i^T\boldsymbol{p}_u. \qquad (27)$$

A singular value decomposition (SVD) approach that is widely used in CDMA and multiple-input multiple-output (MIMO)

systems appears useful but suffers from computational challenges due to an incomplete and sparse user-item matrix. To avoid over-fitting, the recommender system minimizes the squared error on the set of known ratings (i.e., the training set τ), to learn vectors $\boldsymbol{p}_u$ and $\boldsymbol{q}_i$ to predict future ratings, i.e.,

$$\min_{p^*, q^*} \sum_{(u,i) \in \tau} (r_{ui} - \boldsymbol{q}_i^T \boldsymbol{p}_u)^2 + \lambda(\|p_u\|^2 + \|\boldsymbol{q}_i\|^2). \quad (28)$$

This learning algorithm is implemented exactly as is equalization in digital communications, through stochastic gradient descent or alternating least square algorithms. From the Netflix Prize competition, it was learned that matrix factorization models are superior to classical nearest-neighbor techniques [141]. Matrix factorization can further incorporate techniques such as bias adjustment, additional input source for cold start, temporal dynamics, and confidence levels to enhance performance. Another key to final success in the Netflix Prize competition was to adopt collaborative filtering that combines neighborhood and latency models to address quite different structures of data [142]–[145]. Actually, the key success of collaborative filtering is to apply a community structure or clustering in a more precise way. In particular, the predicted value of r_{ui} is taken as the weighted sum over neighboring data, while downplaying neighbors lacking information, instead of just taking an average without discrimination.

It is of interest to consider communication theoretic construction of this matrix formation of recommender systems by treating the user u as an input signal, after a "correlation matrix" $\boldsymbol{R} = [r_{ui}]$ to match (i.e. correlate) users and items, to obtain a user's items of interest:

$$\boldsymbol{p}_u \boldsymbol{R} = \boldsymbol{q}_i. \quad (29)$$

Ignoring potential computational load and sparsity, this equation might suggest a new way of prediction that is equivalent to multiuser detection (MUD) [146] or MIMO communications; recall that MUD and MIMO are equivalent mathematically and in terms of processing interference. If we denote $\boldsymbol{X}$ input signals to channels, $\boldsymbol{A}$ the signal amplitude gains, $\boldsymbol{R}$ the cross-correlation matrix, $\boldsymbol{Y}$ the received signal, $\boldsymbol{N}$ the noise, then MUD for CDMA involves the model

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{R}\boldsymbol{X} + \boldsymbol{N}. \quad (30)$$

Similarly, if $\boldsymbol{H}$ represents the channel gains, then a model for a MIMO system is

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{N}. \quad (31)$$

In a possible implementation in recommender systems, we may use a training set to get this "correlation matrix," and then obtain recommendations for a user from the above equation, though effective computation remains a challenge. Methods in exploring communication theory [147]–[151] are therefore potentially useful in the equivalent knowledge discovery problems.

C. Influence in Social Networks

The diffusion and spreading of ideas, innovation, information, and rumors to influence other people is an important socio-economic process taking place over the Internet
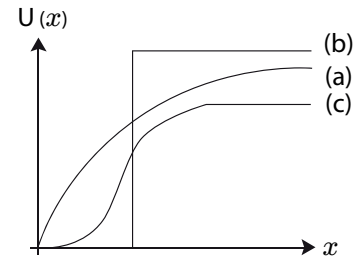

Fig. 10. Three types of utility function shapes: (a) concave, (b) discontinuous, and (c) sigmoidal. Eventually practical utility functions all become concave as marginal returns become diminishing. Maximizing concave and smooth utility functions is easier than maximizing sigmoidal or discontinuous utility functions.

and serves as the basis for many e-commerce activities. An influence model can be based on an overall population or a specific topology, and may be based on a deterministic interaction model or a random interaction model. Seven widely used models are the following [8]: information cascade, synchronization, tipping, diffusion, contagion, random walk, and infection.

One of the early trials in network structure for mathematical modeling is described in [152] where it is assumed that g interacting users form a network described by an interaction matrix. Social influence comes from two factors: communications with other users (e.g., the model in Section II.A) and comparison due to similarity to a user's peers. By straightforward graph theoretic consideration, degree centrality or distance centrality can easily explain the importance of an individual or a node in a network. A more mathematically mature question is to identify the k-node set of maximum influence under independent cascade and linear combination from neighboring users [153], and consequently cascades of information and their patterns [154]. If we consider information dynamics as passing from one individual to another, this is equivalent to broadcasting over a tree [155]. Therefore, if we define the neighborhood of an agent as socially expanding, then this agent is influential. The existence of influential individuals is characterized by the existence of positive recurrent states of a Markov chain [156]. Since it is useful but difficult to observe an individual information propagation path and underlying networks, diffusion network inference is defined as an optimization to find a subset of the entire graph, $\hat{G}$, of at most k degrees,

$$\hat{G} = \arg\max_{|G| \leq k} P(C|G) \quad (32)$$

where C represents a set of cascades. Algorithms have been developed to solve this problem, see e.g., [157].

VI. SOCIAL BEHAVIOR

A. Negative Externalities and Pricing Feedback

1) Utility Functions: A *utility function* is a common modeling tool in economics used to capture "how happy" a consumer would be if a certain amount of resources is allocated to her or him. A typical utility function is shown as the curve (a), a log function, in Figure 10. We consider multiple consumers

or users of resources and denote the utility function of user i as $U_i(x_i)$, where x_i is some performance metric such as throughput.

Maximizing the sum of utilities across all users, $\sum_i U_i(x_i)$, is referred to as *social welfare maximization* [158], [159]. It is of interest to consider from whence these utility function models come. One source is human subject experiments. For example, researchers run tests with a focus group, trying different rates, delays, and jitters of voice calls, and ask them to quantify how happy they are in each case. This leads to utility models for various multimedia applications.

Another is *demand elasticity* modeling. Given a utility function $U(x)$, we also have an associated *demand function*, capturing the volume of demand as a function of price offered. Think about the following *net utility maximization*: a user chooses the x that maximizes the difference between utility and total price paid,

$$U(x) - px,$$

where p is the unit-price. If U is a concave function, it is easy to solve this optimization over one variable by taking the derivative with respect to x and setting it 0: $U'(x) = p$. Since U' is invertible, we can write x, the resulting demand, as a function of p. We call this function U'^{-1} the demand function D, and it is always a decreasing function with higher prices and induced lower demand, i.e.,

$$x = U'^{-1}(p) = D(p).$$

So the utility model determines the demand model. It also determines demand elasticity, defined as normalized price sensitivity of demand:

$$\eta = -\frac{\partial D(p)/\partial p}{D(p)/p}.$$

For example, if utility is logarithmic, then the demand function is $x = 1/p$ and the elasticity is 1.

The third basis on which we can choose utility models is *fairness*. There is a class of utility functions called α-*fair utility functions*, parameterized by a positive number $\alpha \geq 0$, such that maximizing them will yield optimizers that satisfy the definition of α-*fairness*. These are also called *isoelastic* utility functions.

Utility functions are increasing functions ($U' \geq 0$). They are often, but not always, assumed to be smooth (i.e., continuous and differentiable) and concave ($U'' \leq 0$). In some cases, the utility is 0 when x falls below a threshold and is a constant otherwise, leading to a discontinuous utility function, like curve (b) in Figure 10. In other cases, it starts out as a convex function for smaller x: not only is the user happier as x becomes larger, but the rate at which her happiness rises also increases. But after an inflexion point, it becomes concave: more x is still better, but the incremental happiness for each unit of additional x drops as x becomes larger. Such functions are called sigmoidal functions, as shown in curve (c) in Figure 10. Due to the principle of diminishing marginal return, utility functions tend to eventually become concave, possibly flat, for sufficiently large x.

2) Tragedy of the Commons: The positive network effect is often summarized as follows: the benefit of adding nodes to a network increases as the square of the size of the network. The underlying assumptions are that the benefits increase proportionally as the number of links in the network increase, and the number of links is a quadrature function of the number of nodes.

There is also a well-known negative network effect: the "tragedy of the commons." This concept was noted by Lloyd in 1833 and popularized by Hardin's article in 1968 [160]. To describe the effect, consider a group of N farmers sharing a common parcel of land to feed their cows. If there are too many cows, the land will be overgrazed and eventually all cows will die. Should a farmer get a new cow? The benefit of having a new cow goes entirely to her or him, whereas the cost of overgrazing is shared by all N farmers, say, $1/N$ of the cost. So each farmer has the incentive to acquire more cows, even though this collectively leads to overgrazing, the worst case scenario for everyone. This is called the *externality* effect, since it is not represented by pricing signals.

One solution to this problem is to charge each farmer a cost proportional to N, to compensate for the inadequate incentive. This amounts to changing the net utility calculation of each farmer from

$$\text{maximize } U(x) - x$$

to

$$\text{maximize } U(x) - Nx.$$

This is called *internalizing* the externality.

Examples abound in the Internet data plans for mobile phones use different pricing schemes to internalize the externality of interference and congestion. TCP uses congestion pricing to internalize the externality of congestion in the Internet shared by many users; P2P protocols like BitTorrent use tit-for-tat to internalize the externality of free riders in file sharing networks, etc.

3) Structure of Usage Price: Recently both major cellular wireless providers in the U.S. terminated flat rate pricing and shifted to usage pricing [161] for their 3G and 4G data plans. We take a brief look at a model of usage pricing in this subsection.

Let $f \in F$ be a consumer data flow, and let the data rate for flow f in the interval $[(t-1), t]$ be given by x_f^t. The data volume consumption over time T is then given by $x_f = \sum_{t=1}^{T} x_f^t$, and a capacity constraint C applies at every time instant t on a single bottleneck link (often the access link):

$$\sum_f x_f^t \leq C, \ \ \forall t.$$

The shape of the utility function depends on the response of the content or application to varying data-rates, and the utility level represents the consumer's need for the application or content. This motivates us to assume that the consumer's utility level varies over time, but the shape of the utility function does not. Let $\sigma_f^t U_f(x_f^t)$ be the utility to a consumer associated with flow f at time instant t, with factor σ_f^t denoting the time dependency of the consumer's utility level.

Faced with time-varying consumer utilities, the ISP can charge a time-dependent, flow-dependent price $r_f^t(x_f^t)$, as a function of the allocated data rate x_f^t. Consumers maximize

the net utility for each flow f,

$$\text{maximize } \sigma_f^t U_f(x_f^t) - r_f^t(x_f^t) \qquad (33)$$
$$\text{variable } x_f^t.$$

The most common form of the price r is a flat rate baseline g, followed by a linear increase as a function of data-rate with slope h, i.e.,

$$r_f^t(x_f^t) = g_f^t + h_f^t x_f^t. \qquad (34)$$

The *flat price* g_f^t is fixed for the duration of the time interval, irrespective of the allocated data rate. The usage based component is based on a price h_f^t per unit data consumption.

The demand function for this form of the price is

$$D_f^t(g_f^t, h_f^t) = \begin{cases} U_f'^{-1}(h_f^t/\sigma_f^t), & \text{if } g_f^t \le \sigma_f^t U_f(y_f^t) - h_f^t y_f^t \\ 0, & \text{otherwise.} \end{cases} \qquad (35)$$

The condition $\sigma_f^t U_f(x_f^t) - g_f^t - h_f^t x_f^t \ge 0$ ensures that consumers have non-negative utility, by making g sufficiently small. To simplify notation, we often use $D_f^t(h_f^t) = U_f'^{-1}(h_f^t/\sigma_f^t)$, with the implicit assumption that the flat price is low enough to ensure non-negative consumer net-utility.

The revenue maximization problem for a monopoly ISP can be defined by the following:

$$\text{maximize } \sum_t \sum_f (g_f^t + h_f^t x_f^t)$$
$$\text{subject to } \sum_f x_f^t \le C, \;\; \forall t$$
$$x_f^t = U_f'^{-1}(h_f^t/\sigma_f^t), \;\; \forall t, f \qquad (36)$$
$$\sigma_f^t U_f(x_f^t) - g_f^t - h_f^t x_f^t \ge 0, \;\; \forall t$$
$$\text{variables } \{g_f^t, h_f^t, x_f^t\}.$$

The variables $\{g_f^t, h_f^t\}$ are controlled by the ISP, and $\{x_f^t\}$ are the reactions from the users to the ISP prices. Since we assumed a monopoly ISP market, the ISP has *complete pricing power*, which is not the case in practice.

Obviously, ISP revenue increases with a higher flat fee component g_f^t, which can be set so that the consumer net utility is zero. The revenue from each flow is then $g_f^t + h_f^t x_f^t = \sigma_f^t U_f(x_f^t)$, which can be realized by any combination of flat and usage fee that can support a data-rate of x_f^t.

If the usage fee h_f^t is such that the consumer demand $D_f^t(h_f^t)$ is greater than the ISP provisioned data rate x_f^t, then some flows' data packets have to be dropped. However, the ISP can avoid packet drops by setting a sufficiently high usage price to reduce the consumer demand so that the aggregate demand is within the available capacity. It follows that $x_f^t = D_f^t(h_f^t/\sigma_f^t)$.

Therefore, the ISP revenue maximization problem (36) simplifies to

$$\text{maximize } \sum_t \sum_f \sigma_f^t u_f(D_f^t(h_f^t))$$
$$\text{subject to } \sum_f D_f^t(h_f^t) \le C, \;\; \forall t \qquad (37)$$
$$\text{variables } \{h_f^t\}.$$

The capacity inequality should be achieved with equality at optimal prices. It suffices to have the optimal usage fee h^t be the same across all flows f, since it will be used to avoid capacity waste in the sum of demands across all the flows. The optimal flat fee g_f^t is flow dependent, allowing the ISP to fully extract the consumer net-utility.
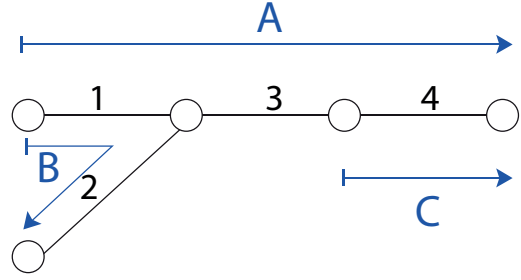

Fig. 11. A simple network with four links and three sessions. Sessions A and B share link 1, and sessions A and C share link 4. Given the fixed capacities on the four links, it is not trivial to design a distributed algorithm that allocates the capacities in an efficient and fair way among the four competing sessions.

Therefore, an optimal pricing scheme that achieves the maximum in (37) is given by the following: for each t, the per-unit usage price h^t is set such that the resulting demands fully utilize the link capacity C, i.e.,

$$\sum_f x_f^t = \sum_f D_f^t(h^t/\sigma_f^t) = C,$$

and the flat rate baseline prices $\{g_f^t\}$ are set such that the maximum revenue is generated out of the consumer demand:

$$g_f^t = \sigma_f^t U_f(x_f^t) - h^t x_f^t.$$

Let R_f^* be the revenue from a flat component of the optimal price, and R_s^* the revenue from the usage component. Using the above solution structure, we can derive the ratio between the flat and usage components. In an exemplary special case, if utility functions are alpha-fair with $\alpha_f = \alpha$ for all f, the ratio of flat revenue to usage dependent revenue becomes [162]

$$\frac{R_f^*}{R_s^*} = \frac{\alpha}{1 - \alpha}. \qquad (38)$$

This reveals that usage dependent revenue dominates with linear utilities ($\alpha \to 0$), while revenue from flat rate components dominates with logarithmic utilities ($\alpha \to 1$). The flat price is a significant component in the extraction of consumer net-utility, if the consumer price sensitivity is low.

4) Network Capacity Allocation: Now we move from one bottleneck link to a general network topology, and from a socio-economic context of charging users (and reactions from humans) to a technological context of Internet protocol (and reactions from operating systems). Any protocol trying to control congestion in a network, whether distributedly like TCP [163] or through a centralized command system, must consider this fundamental issue: each link l's fixed capacity c_l is shared by multiple sessions, and each of these end-to-end sessions traverses multiple links. We assume each source i has one session and uses single path routing. Each link l is shared by a set of sessions $S(l)$, and each session i uses a set of links $L(i)$ along its path decided by IP routing.

Consider the simple example in Figure 11. In this graph, session A originating from node 1 traverses links 1, 3 and 4. And link 1 is shared by sessions A and B. Even when link 3 is not fully utilized, we cannot just increase session A's rate since the bottleneck link for that session might be link 1.

How can we allocate each link's capacity so that the sessions collectively use as much capacity as they can without causing congestion, and their competition is *fairly* coordinated? In Figure 11, assume each link's capacity to be 1 Mbps. One solution that satisfies all four links' capacity constraint, thus a *feasible* solution, is $[\frac{1}{2}, \frac{1}{2}, \frac{1}{2}]$ for the three sessions A, B, and C. In this equal distribution of end-to-end rates, the efficiency of link capacity utilization is 3 Mbps across all the links. For the same capacity utilization, another feasible solution is $[1, 0, 0]$, which starves sessions B and C, and thus is not viewed as a fair allocation. It turns out a standard notion of fairness, called proportional fairness, would give the allocation $[\frac{1}{3}, \frac{2}{3}, \frac{2}{3}]$ to the three competing sessions.

Now we need to write down the problem statement more precisely. This optimization is the basic Network Utility Maximization (NUM) problem [164], [165]. It is a networked version of the social welfare problem. It has been found that TCP Reno and TCP Vegas can be reverse-engineered as solutions to specific NUM problems.

How do we measure efficiency? We use the utility functions, and then sum up each individual TCP session's utility to the end user.

How do we capture the link capacity constraint? On each link l, there is a limited capacity c_l in bits-per-second (bps). The load must be smaller than c_l. There are several ways to express the load on a link in terms of the transmission rates at the sources and the routing decision. We can write the load on link l as the sum of source rates x_i across those sources using this link: $\sum_{i \in S(l)} x_i$. Or, we can use R_{li} as a binary-valued indicator, so that $R_{li} = 1$ if source i's session traverses link l, and $R_{li} = 0$ otherwise. Then the load on link l is simply $\sum_i R_{li} x_i$. In this notation, we can readily see that the constraints

$$\sum_i R_{li} x_i \le c_l, \quad \forall l, \tag{39}$$

are equivalent to the following matrix notation:

$$\mathbf{R}\mathbf{x} \le \mathbf{c}, \tag{40}$$

where $\le$ between two vectors means component-wise $\le$ between the corresponding entries of the vectors.

For example, in the network topology in Figure 11, the link capacity constraint in matrix form becomes

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_A \\ x_B \\ x_C \end{pmatrix} \le \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{pmatrix}.$$

Now we have completely specified the link capacity allocation problem that prescribes what congestion control should be solving, as follows:

$$\begin{array}{ll} \text{maximize} & \sum_i U_i(x_i) \\ \text{subject to} & \mathbf{R}\mathbf{x} \le \mathbf{c} \\ \text{variables} & x_i \ge 0, \quad \forall i. \end{array} \tag{41}$$

We refer to this problem as the basic NUM problem. (41) is easy to solve because it is a convex optimization problem and it is also decomposable. *Decomposition* here refers to breaking up one optimization problem into many smaller ones, somehow coordinated so that solving them will be equivalent
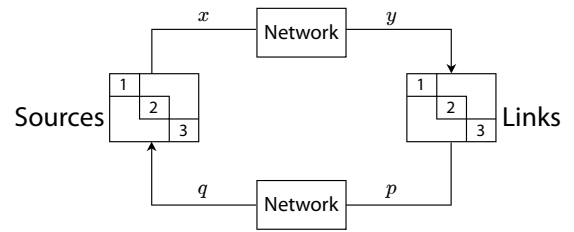

Fig. 12. The feedback control loop in the distributed solution of NUM problem. Each source autonomously adapts its window size (or, transmission rate x_i) based on the path congestion price feedback q_i, while each link autonomously adapts its congestion price p_l based on its own load y_l.

to solving the original. Each of these smaller problems is much easier to solve, often locally at each node in a network. And if their coordination can be done without explicit message passing, we have a truly distributed way to solve the problem.

The following solution to (41) aligns selfish interests into a social welfare maximization. At each of the discrete time slots t, the source of each session simply decides its transmission rate from its demand function, with respect to the current price along its path. This path price q_i is the sum of link prices p_l along all the links this session traverses: $q_i = \sum_{l \in L(i)} p_l$. Thus,

$$x_i[t] = D_i(q_i[t]) = U'^{-1}(q_i[t]). \tag{42}$$

The path price serves as the *congestion signal* and the feedback from the network. We hope it can be obtained without explicit message passing in actual implementation.

At the same time, the router on each link l updates the "price" on that link, as follows:

$$p_l[t] = \{p_l[t-1] + \beta(y_l[t] - c_l)\}^+ \tag{43}$$

where y_l is the load on link l: $y_l[t] = \sum_{i \in S(l)} x_i[t]$. This is an interpretation in the language of pricing, not actual money changing hands between network entities. The parameter $\beta \ge 0$ is the *stepsize* that controls the tradeoff between convergence guarantee and convergence speed.

The feedback loop in the pair of equations (42) and (43) is illustrated in Figure 12 .

The above algorithm not only solves the basic NUM problem, but solves it in a very nice way: fully distributed and intuitively motivated.

- As clearly shown in (43), each link needs only to measure its own total load. It does not need to know any other link's condition, not even the load coming from each of the sources using it.
- As clearly shown in (42), each source needs only to know the total price along the path it is using. It does not need to know any other path's or source's condition, not even the price per link it is using. If the path price q_i can be measured locally at each source i without explicit message passing, this would be a completely distributed solution, as is the case for loss as the price in TCP Reno, and delay as the price in TCP Vegas.

The above also makes sense with an economic interpretation. If at time t, there is more load than capacity on link l, then the price p_l will increase according to (43), and the price for all paths containing link l will rise too, in the next timeslot

$t + 1$. A higher price will reduce demand according to (42), and x_i will drop at all sources that use link l, helping to restore the balance between demand and supply on link l. Indeed, this pricing signal balances out the elastic demand and the fixed supply of capacities. What is interesting is that it carries out this task distributedly through a network consisting of many links and sessions.

B. Network Economy

There are many other key questions and issue that have become increasingly important in the industry:

- "What," "whom," "how" and "how much" should ISPs charge consumers and enterprises for wireless or wireline Internet access?
- The formation of the Internet is driven in part by economic considerations: different tier-1 ISPs form peering and transit relationships based on business and political decisions as much as on technical ones.
- The invention, adoption, and disappearance of Internet technologies are driven by the economics of vendor competition and consumer adoption.
- The investment of network infrastructure, from purchasing wireless licensed spectrum to deploying triple play broadband access, is driven by the economics of capital expenditure, operational expenditure, and return on investment.

The economics of the Internet are interesting because the technology-economics interactions are *bidirectional*: economic forces shaping the evolution of technology, while disruptive technologies can rewrite the balance of economic equations. It is also challenging to study because of the lack of publicly available data on ISPs' cost structure and the difficulty of collecting well-calibrated consumer data.

There is a rapidly growing research field and industry practice on network access pricing. In static pricing, there are several variants:

- Basic usage pricing in the form of tiered and then metered/throttled plans.
- The hourly rate model, *e.g.*, Mobinil in Egypt charges for data connection by the hour of usage.
- Reservation and expected capacity pricing relies on resource allocation driven by the needs of different sessions.
- Priority pricing, in which one can pay more to get a higher speed, such as the priority pass service by SingTel in Singapore. A turbo mode of anti-throttling is also being considered in the U.S. for heavy users whose speed is throttled once usage exceeds some threshold.
- Location-dependent pricing, which is used in certain transportation industries, including downtown London and Singapore.
- Time-dependent pricing, which time shifts delay-tolerant traffic as being piloted in Princeton University's TUBE project.
- Two-sided pricing, in which content providers subsidize content consumers. This type of pricing is used by *e.g.*, Telus in Canada and TDC in Denmark.

In dynamic pricing, network access prices are continuously adjusted to reflect the state of the network. Congestion control can be interpreted as a type of dynamic pricing. If user applications' performance needs, such as delay elasticity, are also taken into account, we can view congestion-dependent pricing as a generalization of time-dependent pricing.

In addition to Internet pricing, spectrum auctions and pricing in wireless communications have been research subjects of high interest in recent years, such as described in [166] and [167] , and economics of sharing femtocells in cellular systems as discussed in [168].

In communication engineering, we consider social behavior in technological networks to deal with radio resource allocation/management. For example, wireless nodes or cognitive radios accessing common spectrum is analogous to natural resource sharing in the biological world [169]. The tragedy of the commons is well known to exist in network bandwidth sharing. However, functional domain control using game theory can restore efficient system operations, as described, for example, in [170] and [171] with applications to practical systems in [172] and [173]. Non-cooperative games are also widely applied to various aspects of medium access, such as the prediction of selfish behavior leading to equilibrium performance [174] and asynchronous dynamic spectrum access to achieve even better performance [175]. Evolutionary game analysis shows an analogy between ecological systems and cognitive radios, indicating that the system must have carefully designed utility functions to avoid selfish user behaviors or equivalent denial-of-service attacks [176].

C. Impact of Online Social Networks on Society

The applications of online social networks have dramatically changed the world. The success of the recent democratic revolution in Egypt was based in part on Twitter and serves as an example to show the impact of a social network running on top of state-of-the-art technological networks. This introduced new insight into collective intelligence or crowd intelligence [177], [178]. Social ants, bees, and schools of fishes, based on very limited local information, can complete complex tasks. Wisdom of the crowd, embodied is such applications as Wikipedia, illustrates the interplay between social networks and technological networks.

VII. Conclusion

In this survey, we have presented some methodologies and cases involving the application of social network analysis to the design of technological networks, and vice versa. Social network topology based on trust, friendship, kinship, cooperation, and so on, has a strong influence on physical technological networks, as a significant portion of traffic in technological networks is generated by social networks and related activities. There are, of course, other aspects related to social networks that we have not discussed in this paper, such as multimedia social networks, P2P sharing, economics, etc. Nevertheless, we hope that this paper will provide a collection of viewpoints that will be useful in future networking research.

Appendix A
Online Social Networks and Statistical Data Analysis

In current social networks, data mining has been widely applied in knowledge discovery from databases (KDD) which is the process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data [179]. Graph-based data mining is useful to deal with network-structured data [180]. For more effective e-commerce (e.g. direct marketing, advertisement, etc.), we wish to infer more useful results from such vast amounts of data, to support applications and services over the Internet and location-based mobile Internet. For example, it is proposed in [181] to use a social network and Markov random field model to obtain a customer's network value for direct marketing. In any case, online social networks involve the behavior and interaction among people, groups, and Internet services, and thus contain vast amounts of digital data. Among many other methodologies, say those in [182] and [183], statistical data analysis and inference techniques enable the gleaning of useful information from large amounts of data in a computationally efficient way.

In most data processing problems, we are interested in extracting information from data that are distorted or corrupted in some unknown manner before taking some actions or exerting a control. We are interested in the *state of nature* Y, where information is embedded in a social network, and hence generates data that are represented by a matrix $\boldsymbol{X} = [x_{ij}]$. The indices $i \in \{1, \dots, n\}$ and $j \in \{1, \dots, p\}$ are called the *observation index* and *attribute index*, respectively. For simplicity, we also use X_k to represent $(x_{1k}, \dots, x_{nk})^T$ as an attribute indexed by k. (See Figure 13)

We formulate a problem as following: Begin with a state of nature Y belonging to a state space $\mathcal{S}$, a network space described by a graph $\mathcal{G}$ and an induced observation space $\Omega_{\mathcal{G}}$. Each vertex in the observation space corresponds to an observable attribute X_j, while edges connect those attributes that we judge to be directly related. In general, Y can be described by $(X_1, \dots, X_p)$ via model $f : Y = f(X_1, \dots, X_p)$ but $f(\cdot)$ may be unknown. A sampling process $K \in \mathcal{K}$ is a relationship between the state space and observation space that introduces some noise or distortion in the data we collect. The goal of data processing is to infer information about Y based on observing the data set $\boldsymbol{X}$. If we define a loss function L, the data processing algorithm seeks to implement a function $\hat{Y} = \hat{f}(X_1, \dots, X_p)$ to satisfy

$$\min_{K \in \mathcal{K}, \mathcal{G}, \hat{Y}} \mathsf{E}\left[L\left((K(Y, \boldsymbol{X}), \hat{Y})\right)\right]. \qquad (44)$$

This optimization problem involves the following challenges:

a) The choice of sampling process K. This problem also may involve a problem of an undetermined model relating Y to $(X_1, \dots, X_p)$.
b) The choice of attribute set, or determination of proper network structure describing the model of Y and $(X_1, \dots, X_p)$.
c) Determination of an estimate $\hat{Y}$ of Y to minimize the overall expected loss.
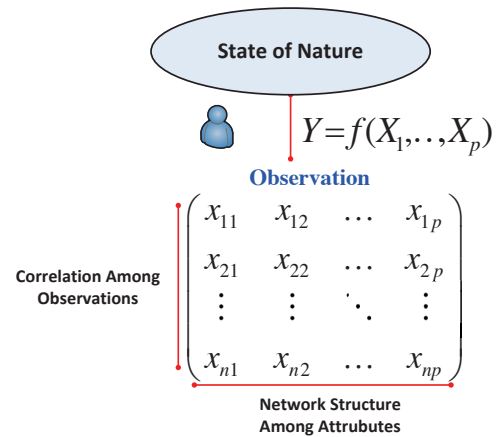d) Maintaining the privacy of data.

Fig. 13. Data processing on social networks involves four properties:1) Undetermined model; 2) Network structure among attributes; 3) Correlation among observations; and 4) Dimensionality.

By appropriate representation of data and exploring the network structure among attributes, we may apply various techniques from decision theory to analyze the data and to infer desirable information from it [184]–[188], techniques which are commonly used in communication systems and networks. To tackle extremely large amounts of data, further dimension reduction techniques such as manifold learning can be adopted in processing [189].

References

[1] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, 1998.
[2] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
[3] S. H. Strogatz, "Exploring complex networks," *Nature*, vol. 410, pp. 268–276, Mar. 2001.
[4] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of Modern Physics*, vol. 74, pp. 47–97, Jan. 2002.
[5] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.
[6] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," *Physics Reports*, vol. 424, no. 4V5, pp. 175–308, 2006.
[7] A.-L. Barabási, "Scale-free networks: A decade and beyond," *Science*, vol. 325, no. 5939, pp. 412–413, 2009.
[8] M. Chiang, *Networked Life: 20 Questions and Answers*. Cambridge University Press, 2012.
[9] T. Berners-Lee, W. Hall, J. A. Hendler, K. O'Hara, N. Shadbolt, and D. J. Weitzner, "A framework for web science," *Found. Trends Web Sci.*, vol. 1, pp. 1–130, Jan. 2006.
[10] Y. Breitbart, M. Garofalakis, C. Martin, R. Rastogi, S. Seshadri, and A. Silberschatz, "Topology discovery in heterogeneous IP networks," in *Proc. IEEE INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, vol. 1, 2000, pp. 265–274 vol.1.
[11] Y. Breitbart, M. Garofalakis, B. Jai, C. Martin, R. Rastogi, and A. Silberschatz, "Topology discovery in heterogeneous IP networks: The netinventory system," *IEEE/ACM Trans. Netw.*, vol. 12, no. 3, pp. 401–414, Jun. 2004.
[12] B. Donnet and T. Friedman, "Internet topology discovery: A survey," *IEEE Commun. Surveys Tuts.*, vol. 9, no. 4, pp. 56–69, quarter 2007.
[13] H. Haddadi, M. Rio, G. Iannaccone, A. Moore, and R. Mortier, "Network topologies: inference, modeling, and generation," *IEEE Commun. Surveys Tuts.*, vol. 10, no. 2, pp. 48–69, 2008.
[14] J. Kleinberg, "The convergence of social and technological networks," *Commun. ACM*, vol. 51, pp. 66–72, Nov. 2008.
[15] H. A. Simon, "A formal theory of interaction in social groups," *American Sociological Review*, vol. 17, no. 2, pp. pp. 202–211, 1952.

[16] A. Rapoport, "Contribution to the theory of random and biased nets," *The Bulletin of Mathematical Biophysics*, vol. 19, pp. 257–277, 1957.

[17] S. Milgram, "The small world problem," *Psychology Today*, vol. 1, pp. 61–67, May 1967.

[18] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, "Network analysis in the social sciences," *Science*, vol. 323, no. 5916, pp. 892–895, 2009.

[19] J. Bilmes, "Dynamic graphical models," *IEEE Signal Process. Mag.*, vol. 27, no. 6, pp. 29–42, Nov. 2010.

[20] D. Barber and A. Cemgil, "Graphical models for time-series," *IEEE Signal Process. Mag.*, vol. 27, no. 6, pp. 18–28, Nov. 2010.

[21] M. Newman, *Networks: An Introduction*, Oxford University Press, 2010.

[22] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.

[23] M. O. Jackson, *Social and Economic Networks*. Princeton University Press, Aug. 2008.

[24] B. Bollobas, *Random Graphs*, W. Fulton, A. Katok, F. Kirwan, P. Sarnak, B. Simon, and B. Totaro, Eds. Cambridge University Press, 2001.

[25] P. Erdös and A. Rényi, "On random graphs, I," *Publicationes Mathematicae (Debrecen)*, vol. 6, pp. 290–297, 1959.

[26] M. Singh and M. Valtorta, "Construction of Bayesian network structures from data: A brief survey and an efficient algorithm," *International Journal of Approximate Reasoning*, vol. 12, no. 2, pp. 111 – 131, 1995.

[27] F. V. Jensen and T. D. Nielsen, *Bayesian Networks and Decision Graphs*, 2nd ed. Springer, 2007.

[28] R. T. Cox, "Probability, frequency and reasonable expectation," *American Journal of Physics*, vol. 14, no. 1, pp. 1–13, Jan. 1946.

[29] M. Girvan and M. E. J. Newman, "Community structure in social and biological networks," *Proc. National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, Jun. 2002.

[30] M. E. J. Newman and J. Park, "Why social networks are different from other types of networks," *Phys. Rev. E*, vol. 68, p. 036122, Sep. 2003.

[31] S. Zhou and R. Mondragon, "The rich-club phenomenon in the internet topology," *IEEE Commun. Lett.*, vol. 8, no. 3, pp. 180–182, Mar. 2004.

[32] I. S. Dhillon, S. Mallela, and D. S. Modha, "Information-theoretic co-clustering," in *Proc. ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2003, pp. 89–98.

[33] M. E. J. Newman, "Modularity and community structure in networks," *Proc. National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, Jun. 2006.

[34] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, pp. 814–818, Jun. 2005.

[35] J. P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A. L. Barabási, "Structure and tie strengths in mobile communication networks," *Proc. National Academy of Sciences*, vol. 104, no. 18, pp. 7332–7336, May 2007.

[36] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graph evolution: Densification and shrinking diameters," *ACM Trans. Knowl. Discov. Data*, vol. 1, no. 1, article 2, Mar. 2007.

[37] T. A. Snijders, G. G. van de Bunt, and C. E. Steglich, "Introduction to stochastic actor-based models for network dynamics," *Social Networks*, vol. 32, no. 1, pp. 44–60, 2010.

[38] R. Xiang, J. Neville, and M. Rogati, "Modeling relationship strength in online social networks," in *Proc. 19th International Conference on World Wide Web*, 2010, pp. 981–990.

[39] B. Yu and M. P. Singh, "Searching social networks," in *Proc. Second International Joint Conference on Autonomous agents and Multiagent Systems*, 2003, pp. 65–72.

[40] P. Costa, C. Mascolo, M. Musolesi, and G. Picco, "Socially-aware routing for publish-subscribe in delay-tolerant mobile ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 5, pp. 748 –760, Jun. 2008.

[41] H. V. Poor, *An Introduction to Signal Detection and Estimation (2nd ed.)*, Springer-Verlag, 1994.

[42] E. Stai, V. Karyotis, and S. Papavassiliou, "Socially-inspired topology improvements in wireless multi-hop networks," in *2010 IEEE International Conference on Communications Workshops (ICC)*, May 2010, pp. 1–6.

[43] S. Cui, A. M. Haimovich, O. Somekh, H. V. Poor, and S. Shamai, "Throughput scaling of wireless networks with random connections," *IEEE Trans. Inf. Theory*, vol. 56, no. 8, pp. 3793–3806, Aug. 2010.

[44] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," in *Proc. Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, 1999, pp. 251–262.

[45] M. Roughan, W. Willinger, O. Maennel, D. Perouli, and R. Bush, "10 lessons from 10 years of measuring and modeling the internet's autonomous systems," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 9, pp. 1810–1821, Oct. 2011.

[46] D. Achlioptas, A. Clauset, D. Kempe, and C. Moore, "On the bias of traceroute sampling: Or, power-law degree distributions in regular graphs," in *Proc. Thirty-seventh Annual ACM Symposium on Theory of Computing*, 2005, pp. 694–703.

[47] M. Mitzenmacher, "A brief history of generative models for power law and lognormal distributions," *Internet Mathematics*, vol. 1, pp. 226–249, 2003.

[48] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, pp. 378–382, Jul. 2000.

[49] D. Alderson, L. Li, W. Willinger, and J. Doyle, "Understanding internet topology: principles, models, and validation," *IEEE/ACM Trans. Netw.*, vol. 13, no. 6, pp. 1205–1218, Dec. 2005.

[50] L. Gyarmati and T. A. Trinh, "Measuring user behavior in online social networks," *IEEE Netw.*, vol. 24, no. 5, pp. 26–31, Sep. 2010.

[51] H. Gao, J. Hu, T. Huang, J. Wang, and Y. Chen, "Security issues in online social networks," *IEEE Internet Comput.*, vol. 15, no. 4, pp. 56–63, Jul. 2011.

[52] A. P. Dempster, "New methods for reasoning towards posterior distributions based on sample data," *The Annals of Mathematical Statistics*, vol. 37, no. 2, pp. 355–374, Apr. 1966.

[53] G. Shafer, "Belief functions and parametric models," *J. Royal Statistical Society. Series B (Methodological)*, vol. 44, no. 3, pp. 322–352, 1982.

[54] P. Resnick and R. Zeckhauser, "Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system. the economics of the Internet and e-commerce," in *Advances in Applied Microeconomics*, M. R. Baye, Ed. Amsterdam, Elsevier Science, 2002, vol. 11, pp. 127–157.

[55] Y. Wang and J. Vassileva, "Trust and reputation model in peer-to-peer networks," in *Proc. P2P '03*, Sep. 2003, pp. 150–157.

[56] S. Buchegger and J.-Y. L. Boudec, "Performance analysis of the CONFIDANT protocol," in *Proc. ACM MobiHoc '02*, Jun. 2002, pp. 226–236.

[57] Y. L. Sun, W. Yu, Z. Han, and K. J. R. Liu, "Information theoretic framework of trust modeling and evaluation for ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 2, pp. 305–317, Feb. 2006.

[58] G. Theodorakopoulos and J. S. Baras, "On trust models and trust evaluation metrics for ad hoc networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 2, pp. 318–328, Feb. 2006.

[59] A. Josang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision," *Decision Support Systems*, vol. 43, no. 2, p. 618V644, Mar. 2007.

[60] M. Gjoka, M. Kurant, C. T. Butts, and A. Markopoulou, "Practical recommendations on crawling online social networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 9, pp. 1872–1892, Oct. 2011.

[61] M. Gjoka, C. T. Butts, M. Kurant, and A. Markopoulou, "Multigraph sampling of online social networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 9, pp. 1893–1905, Oct. 2011.

[62] U. Kuter and J. Golbeck, "Using probabilistic confidence models for trust inference in web-based social networks," *ACM Trans. Internet Technology*, vol. 3, no. 1, Oct. 2010.

[63] S. Zhao, M. X. Zhou, X. Zhang, Q. Yuan, W. Zheng, and R. Fu, "Who is doing what and when: Social map-based recommendation for content-centric social web sites," *ACM Trans. Intelligent Systems and Technology*, vol. 3, no. 1, Oct. 2011.

[64] H. Ma, I. King, and M. R. Lyu, "Learning to recommend with explicit and implicit social relations," *ACM Trans. Intelligent Systems and Technology*, vol. 2, no. 3, Apr. 2011.

[65] D. Riphagen, "Privacy Risks for Users of Social Network Sites," Master's thesis, Systems Engineering, Policy Analysis and Management, Delft University of Technology, Delft, The Netherlands, 2008.

[66] K. Liu and E. Terzi, "A framework for computing the privacy scores of users in online social networks," *ACM Trans. Knowledge Discovery from Data*, vol. 5, pp. 6:1–6:30, Dec. 2010.

[67] R. Gross and A. Acquisti, "Information revelation and privacy in online social networks (the Facebook case)," in *Proc. 2005 ACM Workshop on Privacy in the Electronic Society*, 2005, pp. 71–80.

[68] J. Becker and H. Chen, "Measuring privacy risk in online social networks," in *Proc. Web 2.0 Security & Privacy*, Oakland, CA, May 2009.

[69] F. Adu-Oppong, C. K. Gardiner, A. Kapadia, and P. P. Tsang, "Social circles: Tackling privacy in social networks," in *Proc. Symposium on Usable Privacy and Security*, Pittsburgh, PA, Jul. 2008.

[70] A. Beach, M. Gartrell, and R. Han, "Solutions to security and privacy issues in mobile social networking," in *Proc. International Conference on Computational Science and Engineering, 2009. CSE '09*, vol. 4, Aug. 2009, pp. 1036–1042.

[71] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "An information-theoretic approach to privacy," in *Proc. 48th Annual Allerton Conference on Communication, Control, and Computing*, Oct. 2010, pp. 1220 –1227.

[72] L. Sankar, S. Rajagopalan, and H. V. Poor, "A theory of utility and privacy of data sources," in *Proc. 2010 IEEE International Symposium on Information Theory*, Jun. 2010, pp. 2642–2646.

[73] R. Tandon, L. Sankar, and H. V. Poor, "Discriminatory lossy source coding: Side information privacy," in *Proc. 2011 IEEE Global Communications Conference*, Dec. 2011, pp. 1–5.

[74] T. N. Jagatic, N. A. Johnson, M. Jakobsson, and F. Menczer, "Social phishing," *Communications of the ACM*, vol. 40, no. 10, pp. 94–100, Oct. 2007.

[75] C. Zou, D. Towsley, and W. Gong, "Modeling and simulation study of the propagation and defense of internet e-mail worms," *IEEE Trans. Dependable Secure Comput.*, vol. 4, no. 2, pp. 105–118, Apr.-Jun. 2007.

[76] P. Van Mieghem, J. Omic, and R. Kooij, "Virus spread in networks," *IEEE/ACM Trans. Netw.*, vol. 17, no. 1, pp. 1–14, Feb. 2009.

[77] P.-Y. Chen, S.-M. Cheng, and K.-C. Chen, "Smart attacks in smart grid communication networks," *IEEE Commun. Mag.*, vol. 50, no. 8, pp. 24–29, Aug. 2012.

[78] H. Hu, G. Ahn, and J. Jorgensen, "Multiparty access control for online social networks: Model and mechanisms," *IEEE Trans. Knowl. Data Eng.*, vol. PP, no. 99, p. 1, 2012.

[79] R. Heatherly, M. Kantarcioglu, and B. Thuraisingham, "Preventing private information inference attacks on social networks," *IEEE Trans. Knowl. Data Eng.*, vol. PP, no. 99, p. 1, 2012.

[80] M. Yuan, L. Chen, P. Yu, and T. Yu, "Protecting sensitive labels in social network data anonymization," *IEEE Trans. Knowl. Data Eng.*, vol. PP, no. 99, p. 1, 2012.

[81] S. Xiao, G. Xiao, and T. H. Cheng, "Tolerance of intentional attacks in complex communication networks," *IEEE Commun. Mag.*, vol. 46, no. 1, pp. 146–152, Jan. 2008.

[82] P. Smith, D. Hutchison, J. Sterbenz, M. Schöandller, A. Fessi, M. Karaliopoulos, C. Lac, and B. Plattner, "Network resilience: a systematic approach," *IEEE Commun. Mag.*, vol. 49, no. 7, pp. 88–97, Jul. 2011.

[83] P.-Y. Chen and K.-C. Chen, "Intentional attack and fusion-based defense strategy in complex networks," in *Proc. 2011 IEEE Global Communications Conference*, Dec. 2011, pp. 1–5.

[84] J. M. Kleinberg, "Navigation in a small world," *Nature*, vol. 406, p. 845, 2000.

[85] D. J. Watts, "Networks, dynamics, and the small-world phenomenon," *American Journal of Sociology*, vol. 105, no. 2, pp. pp. 493–527, 1999.

[86] V. Latora and M. Marchiori, "Efficient behavior of small-world networks," *Phys. Rev. Lett.*, vol. 87, p. 198701, Oct. 2001.

[87] G. Kossinets, J. Kleinberg, and D. Watts, "The structure of information pathways in a social communication network," in *Proc. 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2008, pp. 435–443.

[88] D. J. Watts, P. S. Dodds, and M. E. J. Newman, "Identity and search in social networks," *Science*, vol. 296, no. 5571, pp. 1302–1305, 2002.

[89] B. Wellman, "Computer networks as social networks," *Science*, vol. 293, no. 5537, pp. 2031–2034, 2001.

[90] M. F. Schwartz and D. C. M. Wood, "Discovering shared interests using graph analysis," *Commun. ACM*, vol. 36, pp. 78–89, Aug. 1993.

[91] F. Wu, B. A. Huberman, L. A. Adamic, and J. R. Tyler, "Information flow in social groups," *Physica A: Statistical Mechanics and its Applications*, vol. 337, no. 1V2, pp. 327 – 335, 2004.

[92] L. Adamic and E. Adar, "How to search a social network," *Social Networks*, vol. 27, no. 3, pp. 187 – 203, 2005.

[93] C. Leberknight, H. Inaltekin, M. Chiang, and H. V. Poor, "The evolution of online social networks: A tutorial survey," *IEEE Signal Process. Mag.*, vol. 29, no. 2, pp. 41–52, Mar. 2012.

[94] A. Frieze and G. Grimmett, "The shortest-path problem for graphs with random arc-lengths," *Discrete Applied Mathematics*, vol. 10, no. 1, pp. 57–77, 1985.

[95] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Phys. Rev. Lett.*, vol. 86, pp. 3200–3203, Apr. 2001.

[96] H. Inaltekin, M. Chaing, and H. V. Poor, "Delay of social search on small-world random geometric graphs," to appear in *J. Math. Sociol.*

[97] J. Murray, *Mathematical Biology: I. An Introduction*. Springer, 2005.

[98] A. Frieze and G. Grimmett, "The shortest-path problem for graphs with random arc-lengths," *Discrete Applied Mathematics*, vol. 10, no. 1, pp. 57–77, 1985.

[99] D. Shah and T. Zaman, "Rumors in a network: Who's the culprit?" *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 5163–5181, Aug. 2011.

[100] P.-Y. Chen and K.-C. Chen, "Information epidemics in complex networks with opportunistic links and dynamic topology," in *Proc. 2010 IEEE Global Communications Conference*, Dec. 2010, pp. 1–6.

[101] J. Fan, J. Chen, Y. Du, W. Gao, J. Wu, and Y. Sun, "Geo-community-based broadcasting for data dissemination in mobile social networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. PP, no. 99, p. 1, 2012.

[102] K. C.-J. Lin, C.-W. Chen, and C.-F. Chou, "Preference-aware content dissemination in opportunistic mobile social networks," in *Proc. IEEE INFOCOM 2012*, Mar. 2012, pp. 1960–1968.

[103] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Gossip algorithms: design, analysis and applications," in *Proc. IEEE INFOCOM 2005*, vol. 3, Mar. 2005, pp. 1653–1664.

[104] ——, "Randomized gossip algorithms," *IEEE Trans. Inf. Theory*, vol. 52, no. 6, pp. 2508–2530, Jun. 2006.

[105] R. Olfati-Saber, J. Fax, and R. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.

[106] M. F., M. Kurant, A. Markopoulou, C. Westphal, and U. C. Kozat, "Proactive seeding for information cascades in cellular networks," in *Proc. IEEE INFOCOM 2012*, Mar. 2012, pp. 1719–1727.

[107] P. Wang, M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding the spreading patterns of mobile phone viruses," *Science*, vol. 324, no. 5930, pp. 1071–1076, 2009.

[108] S.-M. Cheng, W. C. Ao, P.-Y. Chen, and K.-C. Chen, "On modeling malware propagation in generalized social networks," *IEEE Commun. Lett.*, vol. 15, no. 1, pp. 25–27, Jan. 2011.

[109] J. Kleinberg, "The small-world phenomenon: an algorithm perspective," in *Proc. Thirty-Second Annual ACM Symposium on Theory of Computing*, 2000, pp. 163–170.

[110] ——, "Small-world phenomena and the dynamics of information," *Science*, vol. 14, no. 12, pp. 1–14, 2001.

[111] G. Wu, S. Talwar, K. Johnsson, N. Himayat, and K. Johnson, "M2m: From mobile to embedded internet," *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 36–43, Apr. 2011.

[112] S.-Y. Lien, K.-C. Chen, and Y. Lin, "Toward ubiquitous massive accesses in 3gpp machine-to-machine communications," *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 66–74, Apr. 2011.

[113] E. Lee, "Cyber physical systems: Design challenges," in *Proc. 2008 11th IEEE International Symposium on Object Oriented Real-Time Distributed Computing*, May 2008, pp. 363–369.

[114] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer Networks*, vol. 54, pp. 2787–2805, Oct. 2010.

[115] S.-C. Lin, L. Gu, and K.-C. Chen, "Providing statistical qos guarantees in large cognitive machine-to-machine networks," in *IEEE Global Communications Conference Workshop 2012*, Dec. 2012.

[116] E. Daly and M. Haahr, "Social network analysis for information flow in disconnected delay-tolerant manets," *IEEE Trans. Mobile Comput.*, vol. 8, no. 5, pp. 606–621, May 2009.

[117] S. P. and Borgatti, "Centrality and network flow," *Social Networks*, vol. 27, no. 1, pp. 55–71, 2005.

[118] Y. Zhu, B. Xu, X. Shi, and Y. Wang, "A survey of social-based routing in delay tolerant networks: Positive and negative social effects," *IEEE Commun. Surveys Tuts.*, vol. PP, no. 99, pp. 1–15, 2012.

[119] Y.-C. Liang, K.-C. Chen, G. Li, and P. Mahonen, "Cognitive radio networking and communications: An overview," *IEEE Trans. Veh. Technol.*, vol. 60, no. 7, pp. 3386–3407, Sep. 2011.

[120] F. S. Chu and K. C. Chen, "Radio resource management of self-organizing ofdma wireless mesh networks," *Wireless Communications and Mobile Computing*, vol. 11, no. 3, pp. 306–320, 2011.

[121] S.-M. Cheng, S.-Y. Lien, F.-S. Chu, and K.-C. Chen, "On exploiting cognitive radio to mitigate interference in macro/femto heterogeneous networks," *IEEE Wireless Commun. Mag.*, vol. 18, no. 3, pp. 40–47, Jun. 2011.

[122] A. Giridhar and P. Kumar, "Computing and communicating functions over sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 755–764, Apr. 2005.

[123] E. Fasolo, M. Rossi, J. Widmer, and M. Zorzi, "In-network aggregation techniques for wireless sensor networks: a survey," *IEEE Wireless Commun. Mag.*, vol. 14, no. 2, pp. 70–87, Apr. 2007.

[124] T. Berger, Z. Zhang, and H. Viswanathan, "The ceo problem [multi-terminal source coding]," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

[125] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.

[126] A. Ramamoorthy, "Minimum cost distributed source coding over a network," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 461–475, Jan. 2011.

[127] K.-C. Chen and F.-M. Tseng, "Machine-to-machine communications: Technologies and challenge," to appear in *Ad Hoc Networks*, 2013.

[128] X. Gong, T. Chandrashekhar, J. Zhang, and H. V. Poor, "Opportunistic cooperative networking: To relay or not to relay?" *IEEE J. Sel. Areas Commun.*, vol. 30, no. 2, pp. 307–314, Feb. 2012.

[129] C.-H. Huang and K.-C. Chen, "Dual-observation time-division spectrum sensing for cognitive radios," *IEEE Trans. Veh. Technol.*, vol. 60, no. 8, pp. 3712–3725, Oct. 2011.

[130] P. Eugster, R. Guerraoui, A.-M. Kermarrec, and L. Massoulie, "Epidemic information dissemination in distributed systems," *IEEE Computer*, vol. 37, no. 5, pp. 60–67, May 2004.

[131] P. De, Y. Liu, and S. Das, "An epidemic theoretic framework for vulnerability analysis of broadcast protocols in wireless sensor networks," *IEEE Trans. Mobile Comput.*, vol. 8, no. 3, pp. 413–425, Mar. 2009.

[132] W. Gao, Q. Li, B. Zhao, and G. Cao, "Social-aware multicast in disruption-tolerant networks," *IEEE/ACM Trans. Netw.*, vol. PP, no. 99, p. 1, 2012.

[133] P.-Y. Chen and K.-C. Chen, "Optimal control of epidemic information dissemination in mobile ad hoc networks," in *Proc. 2011 IEEE Global Communications Conference*, Dec. 2011, pp. 1–5.

[134] R. Olfati-Saber, J. Fax, and R. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, Jan. 2007.

[135] H. H. Lee and K.-C. Chen, "Time synchronization in heterogeneous wireless networks," in *Proc. 2012 IEEE Wireless Communications and Networking Conference*, 2012.

[136] G. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Trans. Veh. Technol.*, vol. 42, no. 4, pp. 641–646, Nov. 1993.

[137] M. Chiang, P. Hande, T. Lan, and C. W. Tan, "Power control in wireless cellular networks," *Foundations and Trends in Networking*, vol. 2, pp. 381–533, Apr. 2008.

[138] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Computer Networks and ISDN Systems*, vol. 30, pp. 107–117, Apr. 1998.

[139] A. N. Langville and C. D. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton University Press, 2006.

[140] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences (Classics in Applied Mathematics)*. Society for Industrial and Applied Mathematics, Jan. 1987.

[141] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *IEEE Computer*, vol. 42, no. 8, pp. 30–37, Aug. 2009.

[142] R. Bell, Y. Koren, and C. Volinsky, "Modeling relationships at multiple scales to improve accuracy of large recommender systems," in *Proc. 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2007, pp. 95–104.

[143] R. Bell and Y. Koren, "Scalable collaborative filtering with jointly derived neighborhood interpolation weights," in *Proc. Seventh IEEE International Conference on Data Mining*, Oct. 2007, pp. 43–52.

[144] Y. Koren, "Factorization meets the neighborhood: a multifaceted collaborative filtering model," in *Proc. 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2008, pp. 426–434.

[145] Y. Hu, Y. Koren, and C. Volinsky, "Collaborative filtering for implicit feedback datasets," in *Proc. Eighth IEEE International Conference on Data Mining*, Dec. 2008, pp. 263–272.

[146] S. Verdu, *Multiuser Detection*, 1st ed, Cambridge University Press, 1998.

[147] U. Mitra and H. V. Poor, "Activity detection in a multi-user environment," *Wireless Personal Communications*, vol. 3, pp. 149–174, 1996.

[148] X. Wang and H. V. Poor, "Blind multiuser detection: A subspace approach," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 677–690, Mar. 1998.

[149] W.-C. Wu and K.-C. Chen, "Identification of active users in synchronous cdma multiuser detection," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 9, pp. 1723–1735, Dec. 1998.

[150] T. Oskiper and H. V. Poor, "Online activity detection in a multiuser environment using the matrix CUSUM algorithm," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 477–493, Feb. 2002.

[151] E. Biglieri and M. Lops, "Multiuser detection in a dynamic environment - part I: User identification and data detection," *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3158–3170, Sep. 2007.

[152] R. Th.A.J. and Leenders, "Modeling social influence through network autocorrelation: constructing the weight matrix," *Social Networks*, vol. 24, no. 1, pp. 21–47, 2002.

[153] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proc. Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2003, pp. 137–146.

[154] J. Leskovec, A. Singh, and J. Kleinberg, "Patterns of influence in a recommendation network," in *Advances in Knowledge Discovery and Data Mining*, ser. Lecture Notes in Computer Science, W.-K. Ng, M. Kitsuregawa, J. Li, and K. Chang, Eds. Springer, 2006, vol. 3918, pp. 380–389.

[155] W. Evans, C. Kenyon, Y. Peres, and L. J. Schulman, "Broadcasting on Trees and the Ising Model," *The Annals of Applied Probability*, vol. 10, no. 2, pp. 410–433, May 2000.

[156] L. Dolecek and D. Shah, "Influence in a large society: Interplay between information dynamics and network structure," in *Proc. IEEE International Symposium on Information Theory*, Jul. 2009, pp. 1574–1578.

[157] M. Gomez Rodriguez, J. Leskovec, and A. Krause, "Inferring networks of diffusion and influence," in *Proc. 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2010, pp. 1019–1028.

[158] L. McKnight and J. Bailey, *Internet Economics*, MIT Press, 1997.

[159] J. Walrand, "Economic models of communication networks," in *Performance Modeling and Engineering*, Z. Liu and C. H. Xia, Eds. Springer, 2008, pp. 57–89.

[160] G. Hardin, "The tragedy of the commons," *Science*, vol. 162, no. 3859, pp. 1243–1248, 1968.

[161] R. Edell and P. Varaiya, "Providing internet access: What we learn from index," *IEEE Netw.*, vol. 13, no. 5, pp. 18–25, Sep./Oct. 1999.

[162] P. Hande, M. Chiang, R. Calderbank, and J. Zhang, "Pricing under constraints in access networks: Revenue maximization and congestion management," in *Proc. IEEE INFOCOM 2010*, Mar. 2010, pp. 1–9.

[163] W. R. Stevens, *TCP/IP Illustrated (Vol. 1): The Protocols*, Addison-Wesley Longman Publishing Co., 1993.

[164] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, "Rate control for communication networks: Shadow prices, proportional fairness and stability," *The Journal of the Operational Research Society*, vol. 49, no. 3, pp. 237–252, 1998.

[165] M. Chiang, S. Low, A. Calderbank, and J. Doyle, "Layering as optimization decomposition: A mathematical theory of network architectures," *Proc. IEEE*, vol. 95, no. 1, pp. 255–312, Jan. 2007.

[166] Z. Han, R. Zheng, and H. V. Poor, "Repeated auctions with bayesian nonparametric learning for spectrum access in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 3, pp. 890–900, Mar. 2011.

[167] J. Huang, Z. Han, M. Chiang, and H. V. Poor, "Auction-based resource allocation for cooperative communications," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 7, pp. 1226–1237, Sep. 2008.

[168] S.-Y. Yun, Y. Yi, D.-H. Cho, and J. Mo, "The economic effects of sharing femtocells," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 595–606, Apr. 2012.

[169] D. Liau, K.-C. Chen, and S.-M. Cheng, "A predator-prey model for dynamics of cognitive radios," to appear in *IEEE Commun. Lett.*, 2013.

[170] Q. Ni and C. Zarakovitis, "Nash bargaining game theoretic scheduling for joint channel and power allocation in cognitive radio systems," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 1, pp. 70–81, Jan. 2012.

[171] Y. Xiao, G. Bi, and D. Niyato, "Game theoretic analysis for spectrum sharing with multi-hop relaying," *IEEE Trans. Wireless Commun.*, vol. 10, no. 5, pp. 1527–1537, May 2011.

[172] S.-Y. Lien, Y.-Y. Lin, and K.-C. Chen, "Cognitive and game-theoretical radio resource management for autonomous femtocells with qos guarantees," *IEEE Trans. Wireless Commun.*, vol. 10, no. 7, pp. 2196–2206, Jul. 2011.

[173] J. Huang and V. Krishnamurthy, "Cognitive base stations in LTE/3GPP femtocells: A correlated equilibrium game-theoretic approach," *IEEE Trans. Commun.*, vol. 59, no. 12, pp. 3485–3493, Dec. 2011.

[174] H. Inaltekin, M. Chiang, H. V. Poor, and S. B. Wicker, "Selfish random access over wireless channels with multipacket reception," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 1, pp. 138–152, Jan. 2012.

[175] Y.-Y. Lin and K.-C. Chen, "Asynchronous dynamic spectrum access," *IEEE Trans. Veh. Technol.*, vol. 61, no. 1, pp. 222–236, Jan. 2012.

[176] S.-M. Cheng, P.-Y. Chen, and K.-C. Chen, "Ecology of cognitive radio ad hoc networks," *IEEE Commun. Lett.*, vol. 15, no. 7, pp. 764–766, Jul. 2011.

[177] I. D. and Couzin, "Collective cognition in animal groups," *Trends in Cognitive Sciences*, vol. 13, no. 1, pp. 36–43, 2009.

[178] A. W. Woolley, C. F. Chabris, A. Pentland, N. Hashmi, and T. W. Malone, "Evidence for a collective intelligence factor in the performance of human groups," *Science*, vol. 330, no. 6004, pp. 686–688, 2010.

[179] U. Fayyad, G. Piatetsky-shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," *AI Magazine*, vol. 17, pp. 37–54, 1996.

[180] D. Cook and L. Holder, "Graph-based data mining," *IEEE Intell. Syst.*, vol. 15, no. 2, pp. 32–41, Mar./Apr. 2000.

[181] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proc. Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2001, pp. 57–66.

[182] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proc. 7th ACM SIGCOMM Conference on Internet Measurement*, 2007, pp. 29–42.

[183] H.-P. Kriegel, P. Kröger, and A. Zimek, "Clustering high-dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering," *ACM Trans. Knowl. Discov. Data*, vol. 3, pp. 1:1–1:58, Mar. 2009.

[184] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*, MIT Press, 2009.

[185] T.-Y. Chuang, "Trust with social network learning in e-commerce," in *2010 IEEE International Conference on Communications Workshops*, May 2010, pp. 1 –6.

[186] T.-Y. Chuang and K.-C. Chen, "Robust information fusion on social networks," in *Proc. 2011 IEEE Global Communications Conference*, Dec. 2011, pp. 1–6.

[187] S. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *Proc. IEEE*, vol. 73, no. 3, pp. 433 – 481, Mar. 1985.

[188] B. Frey, *Graphical Models for Machine Learning and Digital Communication*. MIT Press, 1998.

[189] A. Talwalkar, S. Kumar, and H. Rowley, "Large-scale manifold learning," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

Kwang-Cheng Chen (M'89-SM'94-F'07) received the B.S. from the National Taiwan University in 1983, and the M.S. and Ph.D from the University of Maryland, College Park, United States, in 1987 and 1989, all in electrical engineering. From 1987 to 1998, Dr. Chen worked with SSE, COMSAT, IBM Thomas J. Watson Research Center, and National Tsing Hua University, in mobile communications and networks. Since 1998, Dr. Chen has been with National Taiwan University, Taipei, Taiwan, ROC, and is the *Distinguished Professor* and *Deputy Dean* in academic affairs for the College of Electrical Engineering and Computer Science, National Taiwan University. Dr. Chen has been actively involved in the organization of various IEEE conferences as General/TPC chair/co-chair. He has served in editorships with a few IEEE journals and many international journals and has served in various positions within IEEE. Dr. Chen also actively participates in and has contributed essential technology to various IEEE 802, Bluetooth, and 3GPP wireless standards. He has authored and co-authored over 250 technical papers and more than 20 granted US patents. He co-edited (with R. DeMarca) the book *Mobile WiMAX* published by Wiley in 2008, and authored the book *Principles of Communications* published by River in 2009, and co-authored (with R. Prasad) another book *Cognitive Radio Networks* published by Wiley in 2009. Dr. Chen is an IEEE Fellow and has received a number of awards including the *2011 IEEE COMSOC WTC Recognition Award* and has co-authored a few award-winning papers published in the IEEE ComSoc journals and conferences. Dr. Chen's research interests include wireless communications and network science.

Mung Chiang (S'00, M'03, SM'08, F'12) is a Professor of Electrical Engineering at Princeton University, and an affiliated faculty in Applied and Computational Mathematics, and in Computer Science. He received his B.S. (Hons.), M.S., and Ph.D. degrees from Stanford University in 1999, 2000, and 2003, respectively, and was an Assistant Professor 2003-2008 and an Associate Professor 2008-2011 at Princeton University. His research on networking received the IEEE Kiyo Tomiyasu Award (2012), a U.S. Presidential Early Career Award for Scientists and Engineers (2008), several young investigator awards, and a few paper awards including the IEEE INFOCOM Best Paper Award (2012). His inventions have resulted in a few technology transfers to commercial adoption, and he received a Technology Review TR35 Award (2007) and founded the Princeton EDGE Lab in 2009. He initiated the Network Optimization workshop at CISS and the Smart Data Pricing forums. He served as an associate editor for a few journals and an IEEE Communications Society Distinguished Lecturer in 2012-2013. He created a new undergraduate course: Networks: Friends, Money, and Bytes that lead to an open online offering and a flipped classroom at Princeton, and wrote the corresponding textbook: *Networked Life: 20 Questions and Answers*.

H. Vincent Poor (S'72, M'77, SM'82, F'87) received the Ph.D. degree in EECS from Princeton University in 1977. From 1977 until 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990 he has been on the faculty at Princeton, where he is the Michael Henry Strater University Professor of Electrical Engineering and Dean of the School of Engineering and Applied Science. Dr. Poor's research interests are in the areas of stochastic analysis, statistical signal processing and information theory, and their applications in wireless networks and related fields including social networks and smart grid. Among his publications in these areas are the recent books *Smart Grid Communications and Networking* (Cambridge University Press, 2012) and *Principles of Cognitive Radio* (Cambridge University Press, 2013).

Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences, a Fellow of the American Academy of Arts and Sciences, an International Fellow of the Royal Academy of Engineering (U. K), and a Corresponding Fellow of the Royal Society of Edinburgh. He is also a Fellow of the Institute of Mathematical Statistics, the Optical Society of America, and other organizations. In 1990, he served as President of the IEEE Information Theory Society, and in 2004-07 he served as the Editor-in-Chief of the *IEEE Transactions on Information Theory*. He received a Guggenheim Fellowship in 2002, the IEEE Education Medal in 2005, and the Marconi and Armstrong Awards of the IEEE Communications Society in 2007 and 2009, respectively. Recent recognition of his work includes the 2010 IET Ambrose Fleming Medal for Achievement in Communications, the 2011 IEEE Eric E. Sumner Award, and honorary doctorates from Aalborg University, the Hong Kong University of Science and Technology, and the University of Edinburgh.