Section 1.1
Complex networks

Anthony Bonato, Ryerson University
Fan Chung, University of California, San Diego

INTRODUCTION

The study of complex networks analyzes graph-theoretical properties arising in real-world networks, ranging from technological, social, and biological. Web pages and their links, protein-protein interaction networks, and on-line social networks such as Facebook and LinkedIn are some of the commonly studied examples of such networks.

Never before have we confronted graphs of not only such tremendous sizes but also extraordinary richness and complexity, both at a theoretical and a practical level. Numerous problems arise. For example, what are basic structures of such large networks? How do they evolve? What are the underlying principles that dictate their behavior? How are subgraphs (that we observe) related to the large (and often incomplete) host graphs? What are the main graph invariants that capture the myriad properties of such large graphs?

To deal with these questions, graph theory comes into play. There has been a great many advances in the field over the past thirty years in combinatorial, probabilistic, and spectral methods. Still, the traditional random graphs mostly consider the same degree distribution for all nodes or edges while real-world graphs are uneven and clustered. The classical algebraic and analytic methods are efficient in dealing with highly symmetric structures, whereas real graphs are quite the opposite. Guided by examples of complex networks, many new and challenging directions in graph theory emerge. Here we include several selected topics that have been developing. This article is based in part on the surveys [BoTi12, Ch10], and further references can be found in [Bo08, ChLu04].

1.1.1 Examples of complex networks

Nowadays we are surrounded by various information networks that are of prohibitively large size. Dealing with graphs arising from these networks, the usual graph parameters such as the exact number of nodes is no longer as important. Instead, only partial or locally available information can be obtained. This leads to a whole range of new graph parameters and problems, which are motivated by examples of complex networks. Some of these complex networks we will mention here. Our main focus in this survey will be on the web graph and on-line social networks.

EXAMPLES

E1: The **web graph** has nodes consisting of web pages, and edges corresponding to links between them. The web graph may be viewed as directed or undirected, depending on the context.

E2: The **collaboration graph** has nodes as co-authors and two authors are connected by an edge if they have written a joint paper together. The collaboration graph (according to the Math Reviews database) has 401,000 nodes and 676,000 edges. The reader is referred to the website of Grossman [Gr23] for many interesting properties of the collaboration graph.

E3: On-line social networks have nodes consisting of users on some social networking site such as Facebook, and edges consisting of friendship links between them. Twitter may be viewed as a directed graph, where users follow each other, but following may not be reciprocal.

E4: Protein-protein interaction networks have nodes consisting of proteins in a living cell, with two proteins joined if they share some biochemical interaction. For a survey of protein-protein interaction networks, see [Pr05].

E5: Other important examples of complex networks are **router graphs** (nodes are routers and edges correspond to physical connections between them), **call graphs** (nodes are phone numbers and directed edges correspond to calls placed between them), **citation graphs** (nodes are academic papers in a given discipline and there are directed edges between cited papers).

1.1.2 Properties of complex networks

Complex networks arise is diverse arenas but have a completely unexpected coherence. The prevailing characteristics of the complex graphs are *large-scale, small-world, and possessing power-law degree distribution*, which we will now describe. Throughout, graphs are taken as finite and undirected, unless otherwise stated.

D1: The order and size of complex networks varies considerably, but all of them are **large-scale**.

The web graph has over a trillion nodes, with billions of pages appearing and disappearing each day. Facebook [Fa12] has over 955 million users, and over 70 billion friendship links. Some of the nodes of Twitter corresponding to well-known celebrities including Lady Gaga and Justin Bieber have out-degree over 26 million [Tw12].

Because of the large-scale property of complex networks, an examination of almost all nodes, called a **sweep** or **crawl**, is costly and rarely done, and only performed off-line. However, on-line and local computation should ideally be computed in constant time or cost (independent of n, the number of nodes), or of order $O(\log n)$ or $O(\log \log n)$.

D2: Complex graphs are often **sparse**, which means that the graphs have a linear number of edges (that is, $|E(G) \leq c|V(G)|$, for some small constant c. In fact, the constant is usually less than 10 in most examples.) Some networks, such as on-line social networks, however, tend to be dense graphs.

In extremal graph theory and random graph theory, dense graphs are quite well understood, partly due to Szemeredi's regularity lemma. To deal with certain complex graphs with billions of nodes such as the web graph, research has focused primarily on the study of sparse graphs.

D3: The **small world phenomenon** can be regarded as a combination of *small distances* and *clustering*. Namely, there is a short path joining any two nodes and if two nodes share a common neighbor, they are more likely to be adjacent.

There are two different notions of "small distance". The **diameter** of a connected graph G is the maximum distance between a pair of nodes, and is written as $\text{diam}(G)$. The **average distance**, denoted by $\text{adist}(G)$, is defined by,

$$\text{adist}(G) = \sum_{u,v \in F} \frac{d(u,v)}{|F|},$$

where F is the set of pairs of distinct nodes u, v of G with the property that the distance $d(u, v)$ between u and v is finite. The directed analogue of this parameter, where distance refers to shortest directed paths (that is, with no back edges), is denoted by $\text{adist}_d(G)$.

F1: As evidence of the small world property for the web graph, in [AlJeBa99] it was reported that some subgraph of the web graph in 1999 has average distance 19, while some other values were reported earlier in [Br-etal00]. The collaboration graph of Math Reviews has diameter 23 and average distance 7.64 (see [Gr23]). This can be phrased as, "eight degrees of separation" among mathematicians. In some complex networks such as on-line social networks and citations networks, distances have been observed to decrease over time; see [LeKlFa05] A recent study of 700 million users of Facebook in [Ba-etal12], the average distance of that network was given as 4.74.

There are also a number of graph parameters associated with clustering. A direct way is to define the **clustering coefficient of** G to be

$$\frac{2}{n} \sum_{x \in V(G)} \frac{|E(N(v))|}{\deg(v)(\deg(v)-1)},$$

where $E(N(v))$ denotes the number of edges in the subgraph induced by the neighborhood of v in G.

D4: A useful isoperimetric graph invariant related with clustering is the **Cheeger constant** which can be defined as the maximum **Cheeger ratios** $h(S)$ over all subset S of nodes in G. The **Cheeger ratio**, written, $h(S)$ is defined by

$$h(S) = \frac{|E(S,\bar{S})|}{\mathrm{vol}(S)}, \qquad \text{and} \qquad h_G = \max_{\substack{S \subseteq V(G) \\ \mathrm{vol}(S) \leq \mathrm{vol}(G)/2}} h(S)$$

where we define the **volume** $\mathrm{vol}(S) = \sum_{v \in S} \deg(v)$, $\mathrm{vol}(G) = \mathrm{vol}(V(G))$ and $E(A,B)$ denotes the set of edges with one endpoint in A and one endpoint in B. We note that in the literature, the **isoperimetric number** $i(S) = \frac{|E(S,\bar{S}|}{|S|}$ ignores the uneven degree distributions of complex graphs and is therefore, less effective for capturing clustering in complex networks.

D5: In a graph G on n nodes, let $N_k = N_k(n)$ denote the number of nodes of degree k. The degree distribution of G follows a **power law** in some range of k if N_k is proportional to k^{-b}, for a fixed **exponent** $b > 2$. If G is directed, then we may consider (possibly different) power laws for the in- and out-degree distributions by defining $N^{in}_{k,G}$ and $N^{out}_{k,G}$, respectively, in the obvious way.

Power law degree distributions are sometimes called **heavy-tailed distributions**, since the real-valued function $f(k) = k^{-\beta}$ exhibits a polynomial (rather than exponential) decay to 0 as k tends to ∞. If G possesses a power law degree distribution, then we simply say G is a **power law graph**.

F2: By taking logarithms, power laws can be expressed as

$$\log(N_{k,G}) \sim \log(t) - \beta \log(k).$$

Hence, in the log-log plot, we obtain a straight line with slope $-\beta$.

F3: Based on their crawl of the domain of Notre Dame University, Albert, Barabási, and Jeong [AlJeBa99] claimed that the web graph exhibited a power law in-degree distribution, with $\beta = 2.1$. An independent crawl corroborating the findings of [AlJeBa99] was reported in [Ku-etal99], which considered 40 million web pages from 1997 data. The exponent of $\beta = 2.1$ was further corroborated by a larger crawl of the entire web (including 200 million web pages) reported in Broder et al. [Br-etal00]. There were several other reports concerning slightly different power law exponent $\beta = 2.45$ and $\beta = 2.72$ for Web graphs [AlJeBa99, Br-etal00]. Kumar, Novak, and Tomkins [KuNoTo06] studied the evolution of Flickr and Yahoo!360, and found that these networks exhibit power-law degree distributions. Power law degree distributions for both the in- and out-degree distributions were documented in Flickr, YouTube, LiveJournal, and Orkut [Mi-etal07], as well as in Twitter [Ja-etal07, Kw-etal10].

D6: In a graph G, let $0 = \lambda_0 \leq \lambda_1 \cdots \leq \lambda_{n-1} \leq 2$ denote the eigenvalues of the **normalized Laplacian** $\mathcal{L} = I - D^{-1/2} A D^{-1/2}$ where A is the adjacency matrix and D is the diagonal degree matrix of G. The eigenvalue λ_1 is intimately related to the Cheeger constant h_G by the **Cheeger inequality** (see [Ch97]):

$$2h_G \geq \lambda_1 \geq \frac{h_G^2}{2}.$$

F4: In 1999 Faloutsos et al. [FaFaFa99] made an experimental study of an autonomous systems graph finding a power law distribution for the highest eigenvalues of the adjacency matrix.

F5: Mihail and Papadimitriou [MiPa02] showed that the largest eigenvalues of the adjacency power law graphs are themselves distributed according to a power law. This phenomena is sometimes referred to as the **eigenvalue power law**. In the other direction, the eigenvalues of the normalized Laplacian of a random graph with general degree distribution satisfies **semi-circle law**.

F6: Social networks often organize into separate clusters in which the intra-cluster links are significantly higher than the number of inter-cluster links. In particular, social networks contain communities (characteristic of social organization), where tightly knit groups correspond to the clusters [NePa03]. As a result, it is reported in [Es06] that social networks, unlike other complex networks such as the web graph, possess bad spectral expansion properties realized by small spectral gaps in their adjacency matrices.

1.1.3 Random graphs with general degree distributions

To describe a complex graph (or indeed, any graph), we wish to use as few parameters as possible. This is exactly what graph models are supposed to do. For example, the Erdos-Reńyi random graph model $G(n,p)$ uses only two parameters to describe a family of quite complicated graphs by selecting each pair of nodes to be an edge with probability p independently. We note that a random graph in $G(n,p)$ has the same expected degree at every node, and therefore, $G(n,p)$ does not capture some basic behaviors of complex networks. Still, the methods and approaches of classical random graph theory are instrumental in the study of general random graphs.

D7: The $G(\mathbf{w})$ model for random graphs with expected degree sequence $\mathbf{w}$ is defined by selecting edge between v_i and v_j independently with probability p_{ij} where p_{ij} is proportional to the product $w_i w_j$. For example, $G(n,p)$ is a special case of $G(\mathbf{w})$ by taking $\mathbf{w}$ to be $(pn, pn, \ldots, pn)$. The $G(\mathbf{w})$ with a power law distribution $\mathbf{w}$ is called a **power law random graph**.

F7: [ChLu02] Suppose that a random power law graph G in $G(\mathbf{w})$ has n nodes and expected degree sequence $\mathbf{w}$ following a power law with exponent $2 < \beta < 3$. Let G have average degree $d > 1$ and maximum degree m satisfy

$$\log m \gg \frac{\log n}{\log \log n}.$$

Then, for all values of $\beta > 2$, asymptotically almost surely the graph G is connected and the diameter satisfies

$$\mathrm{diam}(G) = \Theta(\log n),$$

and the average distance satisfies

$$\mathrm{adist}(G) \leq (2 + o(1)) \frac{\log \log n}{\log(1/(\beta - 2))}.$$

Many probabilistic and spectral properties of power law random graphs have been studied (see [ChLu04] for details).

D8: An important random graph model for general degree sequences is the **configuration model**, which is a spin-off of random regular graphs. One way to define **random regular graphs** $\mathcal{G}_k$ of degree k on n nodes is to consider all possible matchings in a complete graph K_{kn}. Note that a matching is a maximum set of vertex-disjoint edges. Each matching is chosen with equal probability. We then obtain a random k-regular graph by partitioning the nodes into subsets of size k. Each k-subset then is associated with a node in a random regular graph $\mathcal{G}_k$. Although such a random regular graph might contain loops (that is, an edge having both endpoints the same node), the probability of such an event is of a lower order and can be controlled. Now, instead of partitioning the set of nodes of the large graph into equal parts, we choose a random matching of a complete graph on $\sum_i d_i$ nodes which are partitioned into subsets of sizes $d_1, d_2, \ldots, d_n$. Then we form the random graph by associating each edge in the matching with an edge between associated nodes. Clearly, in the configuration model, there are nontrivial dependencies among the edges.

F8: A useful tool in configuration model is a result of Molloy and Reed [MoRe95, MoRe98]. For a random graph with $(\gamma_i + o(1))n$ nodes of degree i, where γ_i are nonnegative values which sum to 1 and n is the number of nodes, the giant component emerges when $Q = \sum_{i\geq 1} i(i-2)\gamma_i > 0$, provided that the maximum degree is less than $n^{1/4-\epsilon}$ (where $\epsilon > 0$) and some "smoothness" conditions are satisfied. Also, there is asymptotically almost surely no giant component when $Q = \sum_{i\geq 1} i(i-2)\gamma_i < 0$ and the maximum degree is less than $n^{1/8-\epsilon}$.

F9: [AiChLu00] The evolution of a random power law graph with exponent β is as follows :

1. When $\beta > \beta_0 = 3.47875\ldots$, the random graph asymptotically almost surely has no giant component where the value $\beta_0 = 3.47875\ldots$ is a solution to

$$\zeta(\beta-2) - 2\zeta(\beta-1) = 0.$$

   When $\beta < \beta_0 = 3.47875\ldots$, there is asymptotically almost surely a unique giant component.

2. When $2 < \beta < \beta_0 = 3.47875\ldots$, the second largest component is asymptotically almost surely of size $\Theta(\log n)$. For any $2 \leq x < \Theta(\log n)$, there is asymptotically almost surely a component of size x.

3. When $\beta = 2$, asymptotically almost surely the second largest component is of size $\Theta(\frac{\log n}{\log\log n})$. For any $2 \leq x < \Theta(\frac{\log n}{\log\log n})$, there is asymptotically almost surely a component of size x.

4. When $1 < \beta < 2$, the second largest component is asymptotically almost surely of size $\Theta(1)$. The graph is asymptotically almost surely not connected.

5. When $0 < \beta < 1$, the graph is asymptotically almost surely connected.

6. When $\beta = \beta_0 = 3.47875\ldots$, the situation is complicated. It is similar to the double jump of the random graph $\mathcal{G}(n,p)$ with $p = \frac{1}{n}$. For $\beta = 1$, there is a nontrivial probability for either case that the graph is connected or disconnected.

1.1.4 On-line models of complex networks

Models of complex networks are usually (but not always) stochastic, and graphs evolve over time. As there are a large number of models, our survey is far from exhaustive. We focus on three models which have applications to modelling the web graph and on-line social networks. Other models which have been rigorously analyzed are **competition-based preferential attachment models** [Be-etal04], **forest fire models** [LeKlFa05], **ranking models** [FoFlMe06, JaPr09, LuPr06], **growth-deletion models** [ChLu04], **Kronecker graphs** [Le-etal05], **Multiplicative Attribute Graph model** [KiLe12], **planar power law graphs** [FrTs12], and models based on **algorithmic game theory** [Bo-etal12].

Preferential attachment model

Arguably the most commonly studied complex network models are ones incorporating some form of preferential attachment. The first evolving graph model explicitly designed to model the web graph was given by Barabási, Albert [BaAl99], and this model is now referred to as the **preferential attachment**. A rigorous analysis of a preferential attachment model for the case of $\beta = 3$ was given in [Bo-etal01]. Here we give a general definition for a preferential attachment model for $\beta > 2$ introduced and analyzed in [AiChLu01].

D9: The **preferential attachment model** $G(p)$ has one parameter $0 < p < 1$ and involves two possible operations:

- *Vertex-step*: Add a new node v, and add an edge uv from v by randomly and independently choosing u in proportion to the degree of u in the current graph.
- *Edge-step*: Add a new edge rs by independently choosing nodes r and s with probability proportional to their degrees.

We start with an initial graph G_0 (which can usually be taken to be the graph formed by one node having one loop). At time t, the graph G_t is formed by modifying G_{t-1} as follows: with probability p, take a vertex-step, otherwise take an edge-step.

F10: [AiChLu01] For the preferential attachment model $G(p)$, asymptotically almost surely the number of nodes with degree k at time t is

$$M_k t + O(2\sqrt{k^3 t \ln(t)}),$$

where $M_1 = \frac{2p}{4-p}$ and $M_k = \frac{2p}{4-p} \frac{\Gamma(k)\Gamma(1+\frac{2}{2-p})}{\Gamma(k+1+\frac{2}{2-p})} = O(k^{-(2+\frac{p}{2-p})})$, for $k \geq 2$. In other words, asymptotically almost surely the graphs generated by $G(p)$ have a power law degree distribution with the exponent $\beta = 2 + \frac{p}{2-p}$.

The copying model

The copying model was introduced in Kleinberg et al. [Kl-etal99].

D10: The **copying model** has parameters $p \in (0, 1)$, $d \in \mathbb{N}^+$, and an initial directed graph H with constant out-degree d. Assume that the graph G_t at time t has constant out-degree d. At time $t + 1$, we add a new node v_{t+1} and the d out-link of v_{t+1} are generated as follows: We choose a "prototype" node u_t, from the set of all existing

nodes at random. For each of the d out-neighbours w of u_t with probability p, add a directed edge (v_{t+1}, z), where z is chosen uniformly at random from $V(G_t)$, and with the remaining probability $1-p$ add the directed edge (v_{t+1}, w).

F11: [Ku-etal00] If $k > 0$, then in the copying model with parameter p, asymptotically almost surely

$$\frac{N_{k,t}^{in}}{t} = \Theta\left(k^{-\frac{2-p}{1-p}}\right).$$

In particular, the copying model asymptotically almost surely generates directed graphs G_t whose in-degree distribution follow a power law with exponent

$$\beta = \frac{2-p}{1-p} \in (2, \infty).$$

F12: [Ku-etal00] Let $N_{t,i,j}$ denote the expected number of distinct $K_{i,j}$'s which are subgraphs of G_t. Asymptotically almost surely in the copying model with constant outdegree d, for $i \leq \log t$,

$$N_{t,i,j} = \Omega(t \exp(-i)).$$

This abundance of bipartite cliques, which model community structure, is provably absent from the preferential attachment model.

Iterated Local Transitivity model

We describe a recent model for on-line social networks based on transitivity in social networks (that is, friends of friends are often friends). The model is deterministic and simple to describe, but leads to complex behaviour over time.

D11: The **Iterated Local Transitivity** model [Bo-etal11], given some initial graph as a starting point, nodes are repeatedly added over time which clone each node, so that the new nodes form an independent set. The only parameter of the model is the initial graph G_0, which is any fixed finite connected graph. Assume that for a fixed $t \geq 0$, the graph G_t has been constructed. To form G_{t+1}, for each node $x \in V(G_t)$, add its **clone** x', such that x' is joined to x and all of its neighbours at time t. Note that the set of new nodes at time $t+1$ form an independent set of cardinality $|V(G_t)|$.

D12: We write $\deg_t(x)$ for the degree of a node at time t, n_t for the order of G_t, and e_t for its number of edges. Define the **volume** of G_t by

$$\mathrm{vol}(G_t) = \sum_{x \in V(G_t)} \deg_t(x) = 2e_t.$$

F13: [Bo-etal11] For $t > 0$, the average degree of G_t equals

$$\left(\frac{3}{2}\right)^t \left(\frac{\mathrm{vol}(G_0)}{n_0} + 2\right) - 2.$$

Note that the average degree tends to infinity with n; that is, the model generates graphs satisfying a **densification power law**. In [LeKlFa05], densification power laws were reported in several real-world networks such as the physics citation graph and the internet graph at the level of autonomous systems.

D13: Define the **Wiener index** of G_t as

$$W(G_t) = \frac{1}{2} \sum_{x,y \in V(G_t)} d(x,y).$$

F14: [Bo-etal11] For $t > 0$,

$$L(G_t) = \frac{4^t \left(W(G_0) + (e_0 + n_0) \left(1 - \left(\frac{3}{4} \right)^t \right) \right)}{4^t n_0^2 - 2^t n_0}.$$

Note that the average distance of G_t is bounded above by $\mathrm{diam}(G_0) + 1$ (in fact, by $\mathrm{diam}(G_0)$ in all cases except cliques). Further, for many initial graphs G_0 (such as large cycles) the average distance decreases.

F15: [Bo-etal11] If we let $n_t = n$ (so $t \sim \log_2 n$), then

$$C(G_t) = n^{\log_2(7/8) + o(1)}.$$

In contrast, for a random graph $G(n,p)$ with comparable average degree

$$pn = \Theta((3/2)^{\log_2 n}) = \Theta(n^{\log_2(3/2)})$$

F16: [Bo-etal11] Define $\lambda(G_t)$ to be the **spectral gap** of the normalized Laplacian of G_t. For $t \geq 1$, $\lambda(G_t) > \frac{1}{2}$. This represents a drastic departure from the good expansion found in random graphs, where $\lambda = o(1)$; see [Ch97].

1.1.5 Geometric models for complex networks

In **geometric random graph models**, nodes are identified with points in some metric space $\mathcal{S}$, and edges are determined via a mixture of probabilistic rules and distance between nodes in $\mathcal{S}$. Complex networks do not usually live in physical space, but live in some **feature space**, where nodes are associated to vectors of features, and nodes with similar features are more likely to be adjacent. For example, we may view the web graph as residing in **topic-space**, where pages are close if they have common topics. On-line social networks are embedded in **social-space** (sometimes called **Blau space**; see [Mi83]), where nodes with similar social attributes are close.

Several geometric models for complex networks have been proposed , such as **geometric preferential attachment** [FlFrVe07], **random dot product graphs** [ScYo08], and **spatially preferred preferential attachment** [Ai-etal09, CoFrPr12]. We focus on the geometric Protean model which was introduced as a model of on-line social network, in particular, as it generates dense graphs.

D14: [BoJaPr12] The **geometric Protean model**, written GEO-P(α, β, m, p), produces a sequence $(G_t : t \geq 0)$ of undirected graphs on n nodes, where t denotes time. We write $G_t = (V_t, E_t)$. There are four parameters: the *attachment strength* $\alpha \in (0,1)$, the *density parameter* $\beta \in (0, 1-\alpha)$, the *dimension* $m \in \mathbb{N}$, and the *link probability* $p \in (0,1]$. Each node $v \in V_t$ has rank $r(v,t) \in [n] = \{1, 2, \ldots, n\}$. The rank function $r(\cdot, t) : V_t \to [n]$ is a bijection for all t, so every node has a unique rank. The highest ranked node has rank equal to 1; the lowest ranked node has rank n. The initialization

and update of the ranking is done so that the node added at time t obtains an initial rank R_t which is randomly chosen from $[n]$ according to a prescribed distribution. Ranks of all nodes are adjusted accordingly. Let S be the unit hypercube in $\mathbb{R}^m$, with the torus metric $d(\cdot,\cdot)$ derived from the L_∞ metric. More precisely, for any two points x and y in $\mathbb{R}^m$, their distance is given by

$$d(x,y) = \min\{||x-y+u||_\infty \,:\, u \in \{-1,0,1\}^m\}.$$

To initialize the model, let $G_0 = (V_0, E_0)$ be any graph on n nodes that are chosen from S. We define the *influence region* of node v at time $t \geq 0$, written $R(v,t)$, to be the ball around v with volume

$$|R(v,t)| = r(v,t)^{-\alpha} n^{-\beta}\ .$$

For $t \geq 1$, we form G_t from G_{t-1} according to the following rules.

1. Add a new node v that is chosen uniformly at random from S. Next, independently, for each node $u \in V_{t-1}$ such that $v \in R(u, t-1)$, an edge vu is created with probability p. Note that the probability that u receives an edge is proportional to $p\, r(u,t-1)^{-\alpha}$. The negative exponent guarantees that nodes with higher ranks ($r(u,t-1)$ close to 1) are more likely to receive new edges than lower ranks.
2. Choose uniformly at random a node $u \in V_{t-1}$, delete u and all edges incident to u.
3. Vertex v obtains an initial rank $r(v,t) = R_t$ which is randomly chosen from $[n]$ according to a prescribed distribution.
4. Update the ranking function $r(\cdot,t) : V_t \to [n]$.

F17: Since the geometric Protean process is an ergodic Markov chain, it will converge to a stationary distribution.

F18: [BoJaPr12] The geometric Protean model asymptotically almost surely generates power law graphs. More precisely, let $N_k = N_k(n,p,\alpha,\beta)$ denote the number of nodes of degree k, and $N_{\geq k} = \sum_{l \geq k} N_l$. Let $\alpha \in (0,1)$, $\beta \in (0, 1-\alpha)$, $m \in \mathbb{N}$, $p \in (0,1]$, and

$$n^{1-\alpha-\beta} \log^{1/2} n \leq k \leq n^{1-\alpha/2-\beta} \log^{-2\alpha-1} n.$$

Then asymptotically almost surely GEO-P(α,β,m,p) satisfies

$$N_{\geq k} = \big(1 + O(\log^{-1/3} n)\big) \frac{\alpha}{\alpha+1} p^{1/\alpha} n^{(1-\beta)/\alpha} k^{-1/\alpha}.$$

F19: [BoJaPr12] Asymptotically almost surely the average degree of GEO-P(α,β,m,p) is

$$d = (1+o(1)) \frac{p}{1-\alpha} n^{1-\alpha-\beta}.$$

Note that the average degree tends to infinity with n; that is, the model generates graphs satisfying a **densification power law**.

F20: [BoJaPr12] Let $\alpha \in (0,1)$, $\beta \in (0, 1-\alpha)$, $m \in \mathbb{N}$, and $p \in (0,1]$. Then asymptotically almost surely the diameter D of GEO-P(α,β,m,p) satisfies

$$D = O(n^{\frac{\beta}{(1-\alpha)m}} \log^{\frac{2\alpha}{(1-\alpha)m}} n). \tag{1.1}$$

F21: We note that in a geometric model where regions of influence have constant volume and possessing the same average degree as the geo-protean model, the diameter is $\Theta(n^{\frac{\alpha+\beta}{m}})$. This is a larger diameter than in the geometric Protean model. If $m = C \log n$, for some constant $C > 0$, then by (1.1) asymptotically almost surely we obtain a diameter bounded above by a constant.

F22: The dimensions of social space quantify user traits such as interests or geography; for instance, nodes representing users from the same city or in the same profession would likely be closer in social space. The **Logarithmic Dimension Hypothesis** [BoJaPr12] conjectures that the dimension m of an on-line social network is best fit by about $\log n$, where n is the number of users in the on-line social network. The motivation for the conjecture comes from both the geometric protean and multiplicative attribute models (see [KiLe12] for more on the latter model).

1.1.6 Percolation in a general host graph

Most information networks that we observe are subgraphs of certain host graphs for which only partial and incomplete information is available. A natural question is to deduce properties of the host graph from its random subgraphs and vice versa. It is of interest to understand the connections between a graph and its subgraphs. What graph invariants of the host graph are preserved in its subgraphs? Under what conditions can we predict the behavior of the host graph from some samples of subgraphs?

D15: One way to deal with a random subgraph of a given graph G is a type of (bond) **percolation** problem. For a positive value $p \leq 1$, we consider G_p, which is formed by retaining each edge independently with probability p and discarding the edge with probability $1 - p$. A fundamental problem of interest is to determine the **critical probability** for which G_p contains a giant connected component. As an example, if the host graph is the complete graph K_n, G_p is the same as the random graph model $G(n, p)$.

F23: In the applications of epidemics, the host graph is taken to be a **contact graph**, consisting of edges formed by pairs of people with possible contact. The question of interest is to determine the critical probability which corresponds to the problem of finding epidemic threshold for the spreading of diseases.

F24: Percolation problems have long been studied in theoretical physics, especially for the case of the host graph taken to be the lattice graphs $\mathbb{Z}^k$ or special families of Cayley graphs [Bo-etal05, Bo-etal05a, Bo-etal06, MaPa02], and dense graphs [Bp-etal10].

F25: For a general host graph G, it was shown in [ChLu02] that under some (mild) conditions depending on its spectral gap and higher moments of its degree sequence of G, for any $\epsilon > 0$, if $p > (1 + \epsilon)/\tilde{d}$, then asymptotically almost surely the percolated subgraph G_p has a giant component. In the other direction, if $p < (1 - \epsilon)/\tilde{d}$, then asymptotically almost surely the percolated subgraph G_p contains no giant component. We note that the second order average degree $\tilde{d}$ is $\tilde{d} = \sum_v d_v^2/(\sum_v d_v)$, where d_v denotes the degree of v. The volume of the giant component in G_p is given in [ChLu06].

F26: Suppose the host graph G has a large spectral gap λ, defined by

$$\lambda = \min\{\lambda_1, 2 - \lambda_{n-1}\},$$

where $0 = \lambda_0 \leq \lambda_1 \leq \ldots \leq \lambda_{n-1}$ are eigenvalues of the normalized Laplacian of G. It was shown in [ChHo07] that the random subgraph H in G_p preserves the spectral gap as follows:

$$\lambda_H = \lambda - O\Big(\sqrt{\frac{\log n}{pd_{\min}}} + \frac{(\log n)^{3/2}}{pd_{\min}(\log\log n)^{3/2}}\Big).$$

F27: In a paper by Liben-Nowell and Kleinberg [LiKl08] it was observed that the tree-like subgraphs derived from some chain-letter data seem to have relatively large diameter. In the study of the Erdős-Rényi graph model $G(n,p)$, it was shown that the diameter of a random spanning tree is of order $\sqrt{n}$, in contrast with the fact that the diameter of the host graph K_n is 1. Aldous [Al90] proved that in a regular graph G with a certain spectral bound σ, the expected diameter of a random spanning tree T of G, denoted by $\mathrm{diam}(T)$, has expected value satisfying

$$\frac{c\sigma\sqrt{n}}{\log n} \leq \mathrm{E}(\mathrm{diam}(T)) \leq \frac{c\sqrt{n}\log n}{\sqrt{\sigma}}$$

for some absolute constant c. In [ChHoLu12], it was shown that for a general host graph G, with high probability the diameter of a random spanning tree of G is between $c\sqrt{n}$ and $c'\sqrt{n}\log n$, where c and c' depend on the spectral gap of G and the ratio of the moments of the degree sequence.

1.1.7 PageRank for ranking nodes

The notion of PageRank, first introduced by Brin and Page [BrPa98], forms the basis for Google's web search algorithms. Furthermore, PageRank can be used to measure the global importance of nodes, and capture quantitative correlations between pairs or subsets of nodes. Roughly speaking, PageRank is the stationary distribution of some random walk on a given directed graph. In other words, the PageRank value of a node is the probability that, at any given moment, a random surfer is visiting this node following a random walk whose diffusion is under control (by a specified constant α which we will define).

D16: In a directed graph G without vertices with zero outdegree, the transition probability matrix P for a typical random walk on G is defined

$$P(u,v) = \begin{cases} 1/\deg^+(u) & \text{if } (u,v) \in E(G) \\ 0 & \text{otherwise} \end{cases}$$

where $\deg^+(u)$ denotes the out-degree of vertex u.

D17: For a **teleportation constant** $\alpha \in (0,1]$, the **PageRank matrix** of G is defined by

$$P_\alpha = \alpha J_n/n + (1-\alpha)P$$

where J_n is the all 1's matrix of size $n \times n$ and n is the number of vertices in G.

F28: [BrPa98] For a teleportation constant α, the PageRank matrix P_α is **stochastic** (that is, its row sums are all 1) and hence, is a transition probability matrix of a Markov chain. The teleportation constant α is sometimes called the **damping** or **diffusion constant**.

F29: [BrPa98] The PageRank Markov chain with transition probability matrix P_α converges to a stationary distribution $\mathbf{pr}_\alpha$ which is called the **PageRank vector** (with teleportation constant α) and the entry associated with vertex v is the **PageRank** of the v in G.

D18: [Be06, JeWi03] A useful generalization of PageRank is defined as follows. Given a digraph G, consider a typical random walk on G with the transition probability matrix defined by $W = D^{-1}A$, where D is the diagonal degree matrix and A is the adjacency matrix. Personalized PageRank vectors are based on random walks with two governing parameters: a **seed vector s**, representing a probability distribution over $V(G)$, and a teleportation constant α, controlling the rate of diffusion. The **personalized PageRank** $\mathrm{pr}(\alpha, \mathbf{s})$ is defined to be the solution to the following recurrence relation:

$$\mathrm{pr}(\alpha, \mathbf{s}) = \alpha\mathbf{s} + (1-\alpha)\mathrm{pr}(\alpha, \mathbf{s})W$$

where the PageRank is taken to be a row vector.

F30: The original PageRank $\mathbf{pr}_\alpha$ is the special case where the seed vector $\mathbf{s}$ is the uniform distribution.

F31: Personalized PageRank has found a number of applications, such as finding graph clusters and drawing graphs [ChTs10], devising nearly linear time local partitioning algorithms [AnChLa06, AnCh07], and to graph sparsification and graph partitioning [ChZh10].

F32: PageRank has lead to the discovery of a number of ranking algorithms in various complex networks (beyond web search) such as **TwitterRank** [We-etal10], **Social-PageRank** [Ba-etal07], and **ProteinRank** [Fr07].

F33: Because of the close connection of PageRank with random walks, there are very efficient and robust algorithms for computing and approximating PageRank [AnChLa06, Be06, JeWi03].

F34: The usage of PageRank leads to numerous applications including the basic problem of finding a "good" cut in a graph with small Cheeger ratio. Some of the most widely used approximation algorithms are spectral partitioning algorithms with their performance guaranteed by the **Cheeger inequality**:

$$2h_G \geq \lambda_1 \geq \frac{h_f^2}{2} \geq \frac{h_G^2}{2},$$

where h_f is the minimum Cheeger ratio among subsets which are initial segments in the order determined by the eigenvector f associated with the spectral gap λ. Instead of using eigenvector (which requires expensive computation), one can use PageRank vector to derive local partitioning algorithm [AnChLa06], based on the **local Cheeger inequality** for a subset S of vertices in a graph G:

$$h_S \geq \lambda_S \geq \frac{h_g^2}{8\log \mathrm{vol}(S)} \geq \frac{h_S^2}{8\log \mathrm{vol}(S)},$$

where λ_S is the Dirichlet eigenvalue of the induced subgraph on S, h_S is the local Cheeger constant of S defined by $h_S = \min_{T\subseteq S} h(T)$, h_g is the minimum Cheeger ratio over initial segments determined by PageRank vectors g with the seed as a vertex in S, and α is appropriately chosen depending only on the volume of S.

1.1.8 Network games

In morning traffic, drivers choose the most convenient way to get to work without necessarily paying attention to the consequences of their decisions on others. The Internet network can be viewed as a similar macrocosm which functions neither by the control of a central authority nor by coordinated rules. The basic motivation for each individual can only be deduced by greed and selfishness. Every player chooses the most convenient route and use strategies to maximize possible payoff. In other words, we face a combination of **game theory** and graph theory for dealing with large networks both in quantitative analysis and algorithm design. There has been a great deal of progress in algorithmic game theory; see [Ni-etal07]. Of special interests are games on networks that involve graph theory as well as internet economics.

D19: Classical **chromatic graph theory** can be examined from the perspective of game theory. The **chromatic number** of a graph G, written $\chi(G)$, is the minimum number of colors needed to properly color the nodes of G so that adjacent nodes have different colors. Suppose there is a payoff of 1 unit for each player (represented by a node) if its color is different from all its neighbors. A proper coloring is a **Nash equilibrium**, while no player has an incentive to change their color.

F35: Kearns et. al [KeSuMo06] conducted an experimental study of several coloring games on specified networks. A multiple round model of graph coloring games was analyzed in [ChChJa08].

D20: The analysis of **selfish routing** arises naturally in network management. The "**price of anarchy**" refers to the ratio of the average travel time of the decentralized selfish routing versus the coordinated routing respecting the collective welfare.

D21: Braess's paradox is the counter-intuitive observation that, if the travelers were behaving selfishly it was possible to improve everyone's travel time by removing roads.

F36: Valiant and Roughgarden [VaRo10] showed that in a random graph $G(n,p)$ with edge density $p \geq n^{-1/2+\epsilon}$, Braess paradox occurs asymptotically almost surely. Chung and Young proved that Braess' paradox occurs in sparse random graph with $p \geq c \log n/n$. In fact, it was shown [ChYoZh12] that Braess' paradox is ubiquitous in expander graphs.

F37: There has been extensive research done on selfish routing [RoTa02]. The reader is referred to several surveys and some recent books on this topic [Ro06].

References

[AdBuAd03] L.A. Adamic, O. Buyukkokten, E. Adar, A social network caught in the web, *First Monday* **8** (2003).

[Ah-etal07] Y. Ahn, S. Han, H. Kwak, S. Moon, H. Jeong, Analysis of topological characteristics of huge on-line social networking services, In: *Proceedings of the 16th International Conference on World Wide Web*, 2007.

[AiChLu00] W. Aiello, F.R.K. Chung, L. Lu, A random graph model for massive graphs, In: *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, 2000.

[AiChLu01] W. Aiello, F.R.K. Chung, L. Lu, Random evolution of massive graphs, In: *Proceedings of FOCS 2001.*

[Ai-etal09] W. Aiello, A. Bonato, C. Cooper, J. Janssen, P. Prałat, A spatial web graph model with local influence regions, *Internet Mathematics* **5** (2009) 175–196.

[Al90] D. Aldous, The random walk construction of uniform spanning trees, *SIAM J. Discrete Math.* **3** (1990) 450-465.

[Ba-etal12] L. Backstrom, P. Boldi, M. Rosa, J. Ugander, S. Vigna, Four degrees of separation, Preprint 2012.

[AlJeBa99] R. Albert, H. Jeong, A. Barabási, Diameter of the world-wide web, *Nature* **401** (1999) 130.

[AnChLa06] R. Andersen, F.R.K. Chung, K. Lang, Local graph partitioning using pagerank vectors, In: *Proceedings of FOCS'06*, 2006.

[AnCh07] R. Andersen, F.R.K. Chung, Detecting sharp drops in PageRank and a simplified local partitioning algorithm Theory and Applications of Models of Computation, In: *Proceedings of TAMC'07*, 2007.

[Ba-etal07] S. Bao, G. Xue, X. Wu, Y. Yu, B. Fei, Z. Su, Optimizing web search using social annotations, In *Proc. of 16th Int. World Wide Web Conference*, 2007.

[BaAl99] A. Barabási, R. Albert, Emergence of scaling in random networks, *Science* **286** (1999) 509–512.

[Be-etal04] N. Berger, C. Borgs, J.T. Chayes, R. D'Souza, R.D. Kleinberg, Competition-induced preferential attachment, In: *Proceedings of the 31st International Colloquium on Automata, Languages and Programming (ICALP)*, 2004.

[Be06] P. Berkhin, Bookmark-coloring approach to personalized pagerank computing, *Internet Math.*, **3**, (2006) 41–62.

[Bo01] B. Bollobás, *Random graphs, Second edition,* Cambridge Studies in Advanced Mathematics **73** Cambridge University Press, Cambridge, 2001.

[Bp-etal10] B. Bollobás, C. Borgs, J. Chayes, O. Riordan, Percolation on dense graph sequences, *Ann. Probab.* **38** (2010) 150–183.

[Bo-etal01] B. Bollobás, O. Riordan, J. Spencer, G. Tusnády, The degree sequence of a scale-free random graph process, *Random Structures and Algorithms* **18** (2001) 279–290.

[BoRi04] B. Bollobás, O. Riordan, The diameter of a scale-free random graph, *Combinatorica* **24** (2004) 5–34.

[Bo08] A. Bonato, *A Course on the Web Graph*, American Mathematical Society Graduate Studies Series in Mathematics, Providence, Rhode Island, 2008.

[Bo-etal11] A. Bonato, N. Hadi, P. Horn, P. Prałat, C. Wang, Models of on-line social networks, *Internet Mathematics* **6** (2011) 285-313.

[BoJaPr12] A. Bonato, J. Janssen, P. Prałat, Geometric protean graphs, *Internet Mathematics* **8** (2012) 2–28.

[BoTi12] A. Bonato, A. Tian, Complex networks and social networks, book chapter in: *Social Networks*, editor E. Kranakis, Springer, Mathematics in Industry series, 2012.

[Bo-etal05] C. Borgs, J. Chayes, R. van der Hofstad, G. Slade, J. Spencer, Random subgraphs of finite graphs: I. The scaling window under the triangle condition, *Random Structures and Algorithms*, **27** (2005) 137–184.

[Bo-etal05a] C. Borgs, J. Chayes, R. van der Hofstad, G. Slade, J. Spencer, Random subgraphs of finite graphs: II. The lace expansion and the triangle condition, *Annals of Probability*, **33** (2005) 1886–1944.

[Bo-etal06] C. Borgs, J. Chayes, R. van der Hofstad, G. Slade, J. Spencer, Random subgraphs of finite graphs: III. The phase transition for the n-cube, *Combinatorica*, **26** (2006) 395–410.

[Bo-etal12] C. Borgs, J.T. Chayes, J. Ding, B. Lucier, The hitchhiker's guide to affiliation networks: A game-theoretic approach, In: *Proceedings of the 2nd Symposium on Innovations in Computer Science*, 2011.

[BrPa98] S. Brin, L. Page, Anatomy of a large-scale hypertextual web search engine, In: *Proceedings of the 7th International World Wide Web Conference*, 1998.

[Br-etal00] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, J. Wiener, Graph structure in the web, *Computer Networks* **33** (2000) 309–320.

[ChChJa08] K. Chaudhuri, F.R.K. Chung, M.S. Jamall, A network game, In: *Proceedings of WINE*, 2008.

[Ch97] F.R.K. Chung, *Spectral Graph Theory*, American Mathematical Society, Providence, Rhode Island, 1997.

[Ch10] F. Chung, Graph theory in the information age, *Notices of AMS*, **56**, no. 6, (2010) 726–732.

[ChLu02] F.R.K. Chung, L. Lu, Connected components in random graphs with given degree sequences, *Annals of Combinatorics* **6** (2002) 125–145.

[ChLu03] F.R.K. Chung, L. Lu, The average distances in random graphs with given expected degrees, *Internet Mathematics* **1** (2003) 91–113.

[ChLu04] F.R.K. Chung, L. Lu, Coupling on-line and on-line analyses for random power law graphs, *Internet Mathematics* **1** (2004) 409–461.

[ChLu04] F.R.K. Chung, L. Lu, *Complex Graphs and Networks*, American Mathematical Society, U.S.A., 2004.

[ChLu06] F.R.K. Chung, L. Lu, The volume of the giant component of a random graph with given expected degrees, *SIAM J. Discrete Math.* **20** (2006) 395–411

[ChHo07] F.R.K. Chung, P. Horn, The spectral gap of a random subgraph of a graph, *Internet Mathematics* **4** (2007) 225–244.

[ChHoLu12] F.R.K. Chung, P. Horn, L. Lu, Diameter of random spanning trees in a given graph, *Journal of Graph Theory,* **69** (2012) 223–240.

[ChTs10] F.R.K. Chung, A. Tsiatas, Finding and visualizing graph clusters using PageRank optimization, In: *Proceedings of WAW'10*, 2010.

[ChYo10] F.R.K. Chung, S. Young, Braess's paradox in large sparse graphs, In: *Proceedings of WINE 2010.*

[ChYoZh12] F.R.K. Chung, S. Young, W. Zhao, Braess's paradox in expander graphs, Random Structures and Algorithms, to appear.

[ChZh10] F.R.K. Chung, W. Zhao, PageRank algorithm with applications to edge ranking and graph sparsification, In: *Proceedings of WAW'10*, 2010.

[CoFrPr12] C. Cooper, A. Frieze, P. Prałat, Some typical properties of the Spatial Preferred Attachment model, In: *Proceedings of WAW'12*, 2012.

[ErRe59] P. Erdős, A. Rényi, On random graphs I, *Publicationes Mathematicae Debrecen* **6** (1959) 290-297.

[ErRe60] P. Erdös, A. Rényi, On the evolution of random graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* **5** (1960) 17-61.

[Es06] E. Estrada, Spectral scaling and good expansion properties in complex networks, *Europhys. Lett.* **73** (2006) 649–655.

[Fa12] Facebook: statistics. Accessed September 21, 2012. `http://www.facebook.com/press/info.php?statistics`.

[FaFaFa99] M. Faloutsos, P. Faloutsos, C. Faloutsos, Michalis Faloutsos On power-law relationships of the Internet topology, In: *Proceedings of SIGCOMM*, 1999.

[FlFrVe07] A. Flaxman, A. Frieze, J. Vera, A geometric preferential attachment model of networks, *Internet Mathematics* **3** (2007) 187–205.

[FoFlMe06] S. Fortunato, A. Flammini, F. Menczer, Scale-free network growth by ranking, *Phys. Rev. Lett.* **96** 218701 (2006).

[Fr07] V. Freschi, Protein function prediction from interaction networks using random walk ranking algorithm, In: *Proceedings of the IEEE-BIBE07 7th International Symposium on Bioinformatics and Bioengineering*, 2007.

[FrTs12] A. Frieze, C.E. Tsourakakis, On certain properties of random Apollonian networks, In: *Proceedings of WAW'12*, 2012.

[Gr23] J. Grossman, The Erdős number project. Accessed September 21, 2012. `http://www.oakland.edu/enp/`.

[JaPr09] J. Janssen, P. Prałat, Protean graphs with a variety of ranking schemes, *Theoretical Computer Science* **410** (2009) 5491–5504.

[Ja-etal07] A. Java, X. Song, T. Finin, B. Tseng, Why we twitter: understanding microblogging usage and communities, In: *Proceedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007*, 2007.

[JeWi03] G. Jeh, J. Widom. Scaling personalized Web search, In: *Proceedings of the 12th International Conference on World Wide Web*, 2003.

[KiLe12] M. Kim, J. Leskovec, Multiplicative attribute graph model of real-world networks, In: *Proceedings of the 7th Workshop on Algorithms and Models for the Web Graph (WAW 2010)*, 2010.

[KeSuMo06] M. Kearns, S. Suri, N. Montfort, An experimental study of the coloring problem on human subject networks, *Science*, **322** (2006) 824–827.

[Kl99] J.M. Kleinberg, Authoritative sources in a hyperlinked environment, *Journal for the Association of Computing Machinery* **46** (1999) 604–632.

[Kl00] J. Kleinberg, The small-world phenomenon: An algorithmic perspective, In: *Proceedings of the 32nd ACM Symposium on Theory of Computing*, 2000.

[Kl-etal99] J. Kleinberg, S.R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins, The web as a graph: Measurements, models and methods, In: *Proceedings of the International Conference on Combinatorics and Computing*, 1999.

[Ku-etal99] R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins, Trawling the web for emerging cyber-communities, In: *Proceedings of the 8th WWW Conference*, 1999.

[Ku-etal00] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, E. Upfal, Stochastic models for the web graph, In: *Proceedings of the 41th IEEE Symposium on Foundations of Computer Science*, 2000.

[KuNoTo06] R. Kumar, J. Novak, A. Tomkins, Structure and evolution of on-line social networks, In: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2006.

[Kw-etal10] H. Kwak, C. Lee, H. Park, S. Moon, What is Twitter, a social network or a news media?, In: *Proceedings of the 19th International World Wide Web Conference*, 2010.

[LeKlFa05] J. Leskovec, J. Kleinberg, C. Faloutsos, Graphs over time: densification Laws, shrinking diameters and possible explanations, In: *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2005.

[Le-etal05] J. Leskovec, D. Chakrabarti, J. Kleinberg, C. Faloutsos, Realistic, mathematically tractable graph generation and evolution, using Kronecker multiplication, In: *Proceedings of European Conference on Principles and Practice of Knowledge Discovery in Databases*, 2005.

[Li-etal05] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, A. Tomkins, Geographic routing in social networks, In: *Proceedings of the National Academy of Sciences* **102** (2005) 11623–11628.

[LiKl08] D. Liben-Nowell, J. Kleinberg, Tracing information flow on a global scale using Internet, *PNAS*, **105** (2008) 4633–4638.

[LuPr06] T. Łuczak, P. Prałat, Protean graphs, *Internet Mathematics* **3** (2006) 21–40.

[MaPa02] C. Malon, I. Pak, Percolation on finite Cayley graphs, *Lecture Notes in Comput. Sci.* **2483**, Springer, Berlin, (2002) 91–104.

[MiPa02] M. Mihail, C.H. Papadimitriou, On the eigenvalue power law, In: *Proceedings of 6th Int. Workshop on Randomization and Approximation Techniques*, 2002.

[Mi83] M. Miller, Ecology of affiliation, *American Sociological Review* **48** (1983) 519–532.

[Mi-etal07] A. Mislove, M. Marcon, K. Gummadi, P. Druschel, B. Bhattacharjee, Measurement and analysis of on-line social networks, In: *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, 2007.

[Mi04] M. Mitzenmacher, A brief history of generative models for power law and lognormal distributions, *Internet Mathematics* **1** (2004) 226–251.

[MoRe95] M. Molloy, B. Reed, A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms* **6** (1995) 161–179.

[MoRe98] M. Molloy, B. Reed, The size of the giant component of a random graph with a given degree sequence, *Combin. Probab. Comput.* **7** (1998) 295–305.

[NePa03] M.E.J. Newman, J. Park, Why social networks are different from other types of networks, *Phys. Rev. E* **68** 036122 (2003).

[Ni-etal07] N. Nisan, T. Roughgarden, E. Tardos, V.V. Vazirani, *Algorithmic Game Theory*, Cambridge University Press, 2007.

[Pr05] N. Przulj, *Graph theory analysis of protein-protein interactions*, In: Knowledge Discovery in Proteomics, (I. Jurisica and D. Wigle eds.), CRC Press, 2005.

[Ro06] T. Roughgarden, *Selfish Routing and the Price of Anarchy*, MIT Press, 2006.

[RoTa02] T. Roughgarden, E. Tardos, How bad is selfish routing? *Journal of the ACM*, **49** (2002) 236–259.

[VaRo10] G. Valiant, T. Roughgarden, Braess's paradox in large random graphs, *Random Structures and Algorithms* **37** (2010) 495–515.

[ScYo08] E.R. Scheinerman, S. Young, Directed random dot product graphs, *Internet Mathematics* **5** (2008) 91–111.

[Tw12] Twitaholic. Accessed September 21, 2012. `http://twitaholic.com/`.

[WaDoNe02] D.J. Watts, P.S. Dodds, M.E.J. Newman, Identity and search in social networks, *Science* **296** (2002) 1302–1305.

[WaStr98] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, *Nature* **393** (1998) 440–442.

[We-etal10] J. Weng, E. Lim, J. Jiang, Q. He, TwitterRank: finding topic-sensitive influential twitterers, In: *Proceedings of the 3rd ACM International Conference on Web Search and Data Mining*, 2010.

[Wi12] Wikipedia: List of social networking websites. Accessed September 21, 2012. `http://en.wikipedia.org/wiki/List of social networking websites`.