

Generation of networks with prescribed degree-dependent clustering

Chrysanthos E. Gounaris ·
Karthikeyan Rajendran · Ioannis G. Kevrekidis ·
Christodoulos A. Floudas

Received: 8 March 2011 / Accepted: 25 March 2011 / Published online: 16 April 2011
© Springer-Verlag 2011

Abstract We propose a systematic, rigorous mathematical optimization methodology for the construction, “on demand,” of network structures that are guaranteed to possess a prescribed collective property: the degree-dependent clustering. The ability to generate such realizations of networks is important not only for creating artificial networks that can perform desired functions, but also to facilitate the study of networks as part of other algorithms. This problem exhibits large combinatorial complexity and is difficult to solve with off-the-shelf commercial optimization software. To that end, we also present a customized preprocessing algorithm that allows us to judiciously fix certain problem variables and, thus, significantly reduce computational times. Results from the application of the framework to data sets resulting from simulations of an acquaintance network formation model are presented.

Keywords Networks · Graph theory · Mathematical optimization · Clustering

C. E. Gounaris (✉) · K. Rajendran · I. G. Kevrekidis · C. A. Floudas
Department of Chemical and Biological Engineering,
Princeton University, Princeton, NJ 08544, USA
e-mail: gounaris@princeton.edu

K. Rajendran
e-mail: krajendr@princeton.edu

C. A. Floudas
e-mail: floudas@princeton.edu

I. G. Kevrekidis · C. A. Floudas
Program in Applied and Computational Mathematics,
Princeton University, Princeton, NJ 08544, USA
e-mail: yannis@princeton.edu

1 Introduction

Network science research (see [7]) has experienced significant growth in the last decade based on seminal work on random networks, scale-free networks and small-world networks (see [1, 14, 37, 38, 51] and references therein). To a large extent, this was driven by novel technological applications, such as the Internet and modern epidemiology, as well as by national security demands; however, it was also because network science is an integral part of analyzing and engineering many complex systems. The structural properties of networks as diverse as food webs [52], cellular and metabolic networks [28], the Internet backbone [23], coauthorship and citation networks of scientists [35], and social networks [54] to name but a few, have been explored and compared. Measuring the performance and vulnerability [40] of networks, enhancing their capacity for accommodating information or material flow (e.g., [42]), as well as using them in representing massive data sets [9] have all been active research areas. Finally, the effect of network structure on the dynamics of interactions on the network has also been the subject of extensive work—from epidemiology (see, e.g., [11] for the effect of social network structure on epidemic thresholds, or [4] for the effect of population mobility processes in the spreading of infectious diseases), to economy and finance (see, e.g., [15] for how network structure affects the dynamics of employment, or [39] for the containment of a financial crisis) and to opinion formation, culture and language dynamics [3, 46, 50]. A fundamental question arises: “*Why is network anatomy so important to characterize? Because structure always affects function*” [45].

Elucidating the relationship between the structure of a complex network and its behavior and performance is the crucial first step in analyzing, understanding, designing and controlling many complex systems; and to do this, we *must* be able to construct realizations of networks that have prescribed collective properties. For example, in order to assess how a disease spreading model is affected by a characteristic property of the population, such as its assortativity, we need to create representative population sets for various assortativity levels and apply the model to these sets. In order to simulate a large-scale empirical network such as the Internet, we need to scale it down into some workable size and, for that, we need to construct a test-bed network that is small in size but possesses the relevant features of its larger counterpart. In order to investigate which properties affect the evolution of networks, or be able to computationally accelerate this evolution (e.g., through Coarse Projective Integration [29, 30]), we need to have the ability to construct realizations of networks that adhere to prescriptions of some of their macroscopic characteristics.

Existing network generation techniques that give rise to networks with different statistical/structural properties include (but are not limited to): (i) the seminal Erdős–Rényi model [21] of random graphs, (ii) the Watts–Strogatz model [51] with rewiring probability p , for which the resulting graph degree distribution is Poisson (random) for $p = 1$, uniform (structured) for $p = 0$ and power-law (scale-free) for intermediate values of p , (iii) the Barabási–Albert preferential attachment model [6] which creates networks with a power law degree distribution, (iv) Fu and Liao’s [25] extension of the previous model where the probabilities are modified so as to obtain scale-free networks with a power law degree distribution in conjunction with upper bounded, tunable

clustering coefficient, and (v) the R-MAT recursive graph model [16], a procedural algorithm that creates directed graphs. The above techniques focus on the randomness of the results and their adherence to specific probability distributions. One also encounters several algorithms in the literature whose goal is precisely to impart prespecified properties to the graph generated. These include algorithms for (i) networks with given degree distribution (e.g., the Pairing or Configuration model [12,53], the probabilistic Chung–Lu algorithm [17] that creates a random graph with a given expected degree sequence, the deterministic Havel–Hakimi algorithm [26,27] which guarantees a simple graph if the given degree sequence is graphic, and the sequential importance sampling algorithm [8], shown to be amenable to the generation of uniformly distributed graphs with the given degree sequence), (ii) networks with specified degree–degree distribution for nearest neighbors [20], (iii) networks with specified degree distribution and level of clustering [5,31,36,49], and (iv) networks with specified (approximate) degree distribution and degree-dependent clustering [43].

While all these techniques and algorithms—some deterministic, some stochastic, some with convergence guarantees and some without—address several individual aspects of the Network Generation Problem, there is an obvious and important gap in our ability to systematically perform this task, and in hierarchically being able to prescribe several/different collective properties with explicit guarantees of achieving the prescriptions exactly or within bounded deviations. The aim of this paper is to address this gap by proposing a new methodology based on mathematical optimization that can be used to solve the Network Generation Problem with prespecified degree-dependent clustering (see Sect. 3.1 for definition). The degree-dependent clustering is an important property of a network that elucidates its inherent hierarchical topology [41,47,48]. Therefore, capturing this property correctly during the derivation of network models is integral for the latter to exhibit dynamics that are consistent with empirical measurements.

In Sect. 2 we present the general outline of our approach and discuss its advantages, its limitations, and potential remedies for overcoming the latter. In Sect. 3 we present in detail the mathematical formulation and a customized preprocessing algorithm required for its efficient solution. Finally, in Sect. 4 we provide computational results for a collection of test data resulting from simulations of social network formation.

2 Proposed framework

Mathematical optimization has been successfully used in the past to solve many combinatorial problems. In the area of biological networks [13], it has been applied to identify critical nodes in protein–protein interaction networks [10] as well as motifs in DNA sequences [32,55]. In the context of Network Generation, mathematical optimization approaches offer a number of advantages when compared to traditional methods, as a number of additional goals may be pursued. For example, in addition to identifying a network that possesses an ensemble of desired properties, an optimization-based framework would be suitable to: (i) certify if no such network exists, (ii) identify multiple, or even all, non-isomorphic such networks, (iii) identify the network that is most similar—under a suitable metric—to a given network, (iv) identify the network

that minimizes/maximizes some objective, (v) determine the smallest size N_0 for a network to admit all properties, (vi) construct networks that approximately satisfy the properties within given tolerances.

A mathematical optimization framework is versatile enough to accommodate prescriptions that involve various types of network properties and characteristic features of networks. These include among others the degree distribution, global and average clustering coefficient, distribution of local clustering coefficients, absolute number of distinctive patterns, or “motifs,” such as triplets, triangles, squares, stars, etc. It may be required that the above properties be imposed on a network individually, or in any suitable combination. They may also be collectively and independently imposed on subsets of nodes, called *classes*. An example of this concept is the requirement to impose degree-dependent clustering coefficients, that is, to identify a network where the total number of triangles involving nodes of a given degree (the *degree class*) is pre-specified.

In general, the Network Generation Problem calls for the identification of the connections (arcs) that need to exist between its N nodes. To that end, we define a collection of binary variables x_{ij} , $i < j$ to indicate the existence or not of these connections:

$$x_{ij} = \begin{cases} 1 & : \text{ if arc } (i, j) \text{ exists in the solution} \\ 0 & : \text{ otherwise.} \end{cases} \quad (1)$$

For computational tractability, the aim is to always express the problem as a Mixed-Integer Linear Problem (MILP) [24, 34], that is, through a formulation of the general form (2–3).

$$\min_{x, x'} \alpha^T x + \beta^T x' \quad (2)$$

$$s.t. \quad Ax + Bx' \leq c \quad (3)$$

where x' is a generic set of auxiliary variables that could be either binary or continuous.

The two major challenges of the Network Generation Problem are its large combinatorial complexity and the potential isomorphism of the solutions. Both of these challenges make instances with even a small number of nodes to be intractable for off-the-shelf commercial optimization software. In order to address such challenges, the proposed mathematical optimization approach is based on efficient representation (modeling) of the problem and effective model reduction procedures.

Typically, we aim at formulating the various requirements of the prescription using few variables and involving expressions that have as tight LP-relaxations as possible. We also focus on modularity of the model, that is, we attempt to describe each property with a stand-alone set of equations that can be included—like a “building block”—in an overall formulation, whenever the property is part of the prescription. This allows us to easily move from one formulation to another, making heavy use of results at hand (e.g., use a network that satisfies property A to initialize the search process for a network that satisfies both properties A and B).

Furthermore, since most of the times one is interested in only feasible solutions (not necessarily optimal in terms of an objective function), we can reduce the combinatorial

search space by a priori postulating the existence of certain arcs as part of the solution, e.g., by postulating some sub-network as an unnegotiable part of the network under construction. In principle, this requires thoughtful analysis of the particular prescription at hand, as one needs to rank-order the set of possible network arcs in terms of their likelihood to participate in feasible solutions. However, if performed well, such a step would reduce the size of the MILP formulation and possibly bring the remaining problem—which can be of significantly smaller complexity than the original—within the capabilities of available optimization software. By providing the possibility to relax some of the postulated arcs in an iterative fashion, the overall scheme inherits the exactness of the approach and can still guarantee the identification of a solution, when one exists, in all cases.

3 Degree-dependent clustering

We begin by providing a few definitions:

Definition 1 The degree of a node equals the number of nodes with which it is connected.

Definition 2 Node A forms a triplet with nodes B and C if the connections (A, B) and (A, C) exist.

Definition 3 Node A forms a triangle with nodes B and C if the connections (A, B), (A, C) and (B, C) exist.

Definition 4 The local clustering coefficient of a node equals the number of triangles divided by the number of triplets it forms with other nodes.

Definition 5 A degree class is the collection of all nodes in a network that have the same degree.

3.1 Optimization formulation

Consider a set of N nodes and a prescription $d_i, i = 1, 2, \dots, N$, for the degree sequence. We aim at constructing a network where each node i attains degree d_i . Let index q loop over all the degree classes thus defined. Degree-dependent clustering calls for imposing—in addition to the degree sequence—an allowable range on the number of triangles, $T_q \in [T_q^L, T_q^U], \forall q$, that can be collectively formed by all nodes of degree class q . This is equivalent to imposing an allowable range on the average local clustering coefficient across all nodes of the same degree. The problem is formulated as a feasibility problem. Formulation (4–8) applies.

$$\min_{\substack{x_{ij} \in \{0,1\} \\ 0 \leq r_{ijk} \leq 1 \\ s d_i^+, s d_i^-, s t_q^+, s t_q^- \geq 0}} \sum_i (s d_i^+ + s d_i^-) + \sum_q (s t_q^+ + s t_q^-) \quad (4)$$

$$s.t. \quad \sum_{j>i} x_{ij} + \sum_{j<i} x_{ji} = d_i + sd_i^+ - sd_i^- \quad \forall i \quad (5)$$

$$\sum_{i \in q} \left\{ \sum_{j>i} \sum_{k>j} tr_{ijk} + \sum_{j>i} \sum_{k<i} tr_{kij} + \sum_{j<i} \sum_{k<j} tr_{jki} \right\} \geq T_d^L - st_q^- \quad \forall q \quad (6)$$

$$\sum_{i \in q} \left\{ \sum_{j>i} \sum_{k>j} tr_{ijk} + \sum_{j>i} \sum_{k<i} tr_{kij} + \sum_{j<i} \sum_{k<j} tr_{jki} \right\} \leq T_d^U + st_q^+ \quad \forall q \quad (7)$$

$$\left. \begin{array}{l} tr_{ijk} \leq x_{ij} ; tr_{ijk} \leq x_{jk} ; tr_{ijk} \leq x_{ik} \\ tr_{ijk} \geq x_{ij} + x_{jk} + x_{ik} - 2 \end{array} \right\} \forall i < j < k \quad (8)$$

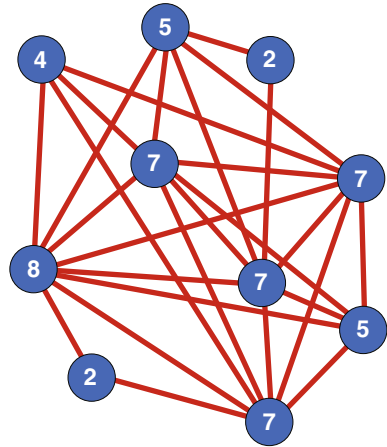
In addition to the binary variables x_{ij} introduced in (1), the formulation uses two sets of non-negative slack variables, sd_i^+ and sd_i^- , to express the deviation of node i 's actual degree ($\sum_{j>i} x_{ij} + \sum_{j<i} x_{ji}$) from the prescribed degree d_i (Eq. 5). Two additional sets of slack variables, st_q^+ and st_q^- , are used to penalize the deviation from the prescribed bounds of the number of triangles formed by the members of each degree class (Eqs. 6,7). We also use continuous variables $tr_{ijk} \in [0, 1]$, $\forall i < j < k$, to denote the existence or not of a triangle between nodes i , j and k . We define these through products of variables, $tr_{ijk} = x_{ij}x_{jk}x_{ik}$, and use their convex/concave envelopes [33,2] in the formulation (Eq. 8). Since at optimality the x_{ij} variables attain only binary values, the envelopes are exact representations of the products. The objective (Eq. 4) calls for minimizing the sum of all slack variables so as to reduce the total deviation of the solution from the prescription.

Note that if the optimal objective value of this formulation is $z^* > 0$, then the desired prescription is not attainable; this could be the result of the degree sequence being not *graphic*¹ or the triangle requirement being too restrictive. However, if the prescription is indeed attainable, the above formulation ought to have an optimal solution $z^* = 0$ and will yield at optimality a solution set of arcs x_{ij} that correspond to a network with the desired properties.

Note that a given prescription often admits multiple $z^* = 0$ solutions and this is true even for very small problem sizes (e.g., the 10-node instance of Ex. 1). It should be noted, though, that not all of the solutions will be truly unique solutions. This is due to the inherent isomorphism of the problem; the indices of two or more nodes in the same degree class can be interchanged without changing the structure of the network. Typically, two remedies can be applied: (i) differentiation between the nodes in each degree class by adding a few explicit constraints that relate to a property beyond the degree or (ii) designation of some solution features as “nominal” and dynamic rejection during branch-and-cut of other symmetric solutions that do not possess those features.

¹ In the case of degree sequence it is easy to infer a priori, through application of the “Handshake Theorem” and the “Erdős–Gallai Criterion” [22,44], whether the prescription will be graphic or not. However, this is not the case when additional requirements, beyond the degree sequence are imposed.

Fig. 1 Representative solution for a network that satisfies the prescription of Example 1. The node labels correspond to their degree



Example 1 Let $N = 10$ nodes, and let $d_i = \{2, 2, 4, 5, 5, 7, 7, 7, 7, 8\}$ be a prescription for the degree sequence. The set of degree classes is then $q = \{2, 4, 5, 7, 8\}$. Let also $T_q^L = \{0, 2, 4, 20, 6\}$ and $T_q^U = \{2, 6, 20, 84, 26\}$ be a prescription in regards to the allowable intervals for the cumulative number of triangles of each class. Despite the small size of this problem, there exist 72 different (that is, non-isomorphic) solutions. Fig. 1 illustratively depicts one of them. \square

3.2 Preprocessing algorithm

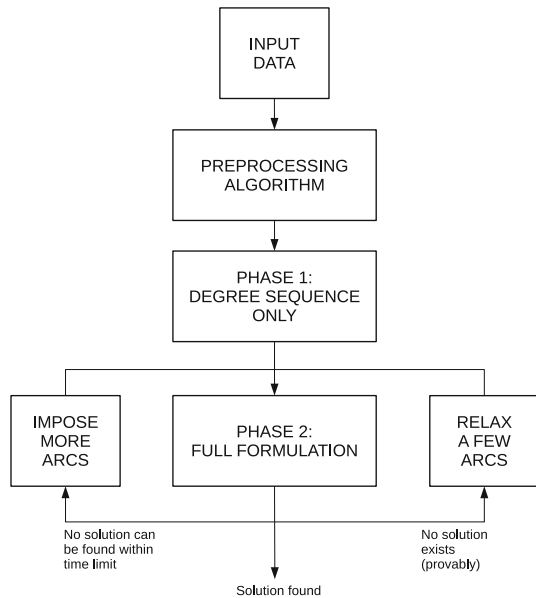
Overall, the degree-dependent clustering formulation cannot be addressed efficiently with uncustomized commercial MILP software. For reference, we attempted to solve for multiple test prescriptions and for network sizes of as low as 20 nodes. CPLEX v11.0 [18] with various parameter configurations² had not been able to attain a feasible solution after 10 hours.³ In order to address this limitation, we need to somehow reduce the combinatorial complexity of the problem and, to that end, we have developed a customized preprocessing algorithm. The basic idea behind this algorithm is to rank-order the set of possible network arcs in terms of their likelihood to participate in feasible solutions, and enforce the existence of the most likely arcs as part of the solution (through fixing the corresponding variables x_{ij}).

The preprocessing algorithm works in an iterative greedy fashion. At each step, we select—out of the pool of non-fixed arcs—the one that appears at the given moment to be the most attractive to be appended to the set of fixed arcs, that is, the one that (i) addresses the triangle prescription the most, i.e., forms new triangles that involve the degree-classes q that deviate the most from their upper allowable limit, T_q^U , and at the

² We attempted various options in regards to upper bounding heuristics, LP solver, search emphasis, and application of cuts.

³ All computational experiments in this paper were performed on an Intel Core2 Quad 2.83 GHz (single-processor runs).

Fig. 2 High-level algorithm flowsheet for solving the Network Generation Problem with prescribed degree-dependent clustering



same time (ii) does not over-constrain the remaining problem, rendering it infeasible. Such a preprocessing step can significantly reduce the combinatorial complexity of the problem and, as presented later in Sect. 4, allow us to address instances with 100 nodes in as little as a few seconds.

After the preprocessing step has been applied and a possible model reduction has occurred, the overall algorithm for the Network Generation Problem with degree-dependent clustering prescriptions proceeds through a number of phases, as depicted in Fig. 2. The first phase corresponds to imposing the degree sequence prescription by itself, that is, minimizing the sum of degree slack variables, $\sum_i (sd_i^+ + sd_i^-)$, subject only to Eq. (5). Due to the near-total-unimodularity of this problem, solving it is a fast⁴ operation and can be used to obtain a good initialization for the full problem. The algorithm then moves to the second phase, which is the application of Formulation (4–8). If no solution has been found within some predefined time limit, we augment the set of imposed arcs with some additional selections and rerun this phase (which now corresponds to a more restricted formulation and should, thus, be easier to solve). On the other hand, if no solution with objective value $z^* = 0$ exists,⁵ we relax some of the imposed arcs before attempting to rerun the second phase (which now corresponds to a less restricted formulation).

⁴ 1,000-node instances can be solved in 5 s by CPLEX v11.0 (default options). Available machine memory was the limiting factor to solving larger instances.

⁵ This can be inferred when the overall lower bound during the search process rises to a value above $z^{LB} = 0$.

4 Computational experiments

We illustrate our MILP framework with prescriptions resulting from the acquaintance network formation model proposed by Davidsen et al. [19]. The model demonstrates the emergence of small world behavior in social networks, where the nodes represent people and the edges represent some sort of relationship between them. We first initialize the simulation with a random Erdős–Rényi graph (0.1 edge probability). The model evolves at each time step according to the dynamics presented in detail in [19]. For reference, we used a $p = 0.04$ removal probability. In each experiment, we evolve 100 different 100-node networks. After a given number of steps, we stop the simulations and obtain a degree-dependent clustering prescription by averaging over the 100 resulting networks. We then use the approach presented in this paper to construct a single 100-node network that possesses this prescription. The actual prescriptions used as input data to our framework are presented in the Appendix.

Table 1 presents the results of our methodology for 28 independent experiments. The simulation stoppage time (in terms of simulation steps) is denoted by t_S , while $|q|$ denotes the number of the different degree-classes.⁶ The total number of arcs in the required graph is denoted by a_T , while a_P denotes the arcs that were fixed by the preprocessing algorithm. The latter took t_P time and all the optimization phases took cumulatively t_O time, for a total computation time of $t_P + t_O$. We were able to generate a network satisfying the required prescription in all 28 instances, with 20 of those instances requiring less than 1 min of computation. The preprocessing algorithm, on average, fixed $\sim 90\%$ of the total arcs; yet in the majority (24/28) of the instances, that was not over-restrictive and they were solved at first pass. The remaining 4 instances (denoted with an asterisk) required 1 round of arc removal where 10% of the fixed arcs (a_P) were removed. The total computation time for these 4 problems was significantly larger.

5 Conclusions

A mathematical optimization-based framework was introduced for the solution of the Network Generation Problem with degree-dependent clustering prescriptions. This was the first time in the open literature that a framework with theoretical guarantees of achieving the required prescription, when feasible, was presented for this problem. However, the optimization of the resulting formulation presents computational challenges and is beyond the capabilities of off-the-shelf commercial optimization software. To that end, we developed a preprocessing scheme to judiciously prefix a subset of the optimization variables and reduce the model's complexity. We were then able to address the construction of 100-node networks for test data sets generated through a social network evolution method. We conclude by noting that this approach can be readily generalized to other prescriptions of network collective properties and this will be the subject of a forthcoming publication.

⁶ This is not necessarily equal to the maximum required degree, as some degree levels may be missing from the sequence

Table 1 Computational results for a suite of 100-node test-problems

Prob ID	t_S	$ q $	Arcs		Time (sec)	
			a_T	a_P	t_P	t_O
A	0	29	496	461	10	26
B	0	16	573	328	7	1,144
C	500	33	579	550	12	5
D	1,000	34	713	682	15	2
E	1,000	30	721	691	13	4
F	1,000	30	812	764	15	11
G	1,000	33	703	665	13	7,810*
H	1,000	31	755	650	14	135
I	2,000	35	832	777	16	17,175*
J	2,000	36	824	803	18	2
K	2,000	37	725	696	13	3
L	2,000	32	655	632	14	2
M	2,000	32	834	636	13	6,499*
N	4,000	31	637	616	13	2
O	4,000	30	572	537	11	4
P	4,000	35	723	700	14	3,101*
Q	4,000	31	605	555	11	27
R	6,000	35	700	675	14	3
S	6,000	31	602	423	11	1,215
T	6,000	33	633	543	12	5
U	6,000	34	622	518	12	15
V	6,000	37	740	607	14	365
W	10,000	33	632	598	12	5
X	10,000	35	650	609	12	4
Y	10,000	34	606	582	12	2
Z	10,000	37	712	674	15	6
AA	10,000	27	749	703	14	9
AB	10,000	32	606	508	14	25

Acknowledgments The work of I.G.K and K.R. was partially supported by the US DOE (DE-SC0002097) and DTRA (HDTRA1-07-1-0005). C.A.F. acknowledges support from the National Science Foundation (CBET-0827907).

Appendix: Prescriptions used

Each row represents a different degree class q . The row format is $\{D_q, n_q, T_q^L, T_q^U\}$, where D_q is the degree of the class, n_q is the number of nodes in the class, and (T_q^L, T_q^U) is the allowable range for the number of triangles that can be formed collectively by the nodes in this class.

A	B			C			D			E			F			G		
1	1	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	1	0
2	9	9	5	1	0	6	1	0	10	4	10	2	4	2	4	2	11	5
3	9	18	27	6	2	3	8	3	8	13	24	3	4	6	12	3	7	11
4	9	27	54	7	7	14	40	3	6	16	36	4	6	15	36	4	7	18
5	14	70	126	8	11	33	74	4	10	28	61	5	9	27	81	5	3	9
6	8	40	104	9	8	36	60	5	9	45	81	6	8	35	36	7	4	20
7	3	24	51	10	7	40	67	6	4	24	52	7	3	12	48	8	1	4
8	5	35	110	11	14	95	150	7	3	24	52	8	1	18	39	10	3	27
9	4	60	124	12	9	68	121	8	4	46	100	9	3	27	69	11	2	18
10	5	70	170	13	13	117	185	9	3	61	136	10	2	28	62	12	1	46
11	4	72	184	14	11	113	189	10	7	98	235	11	1	28	62	12	1	31
12	4	84	212	15	3	33	60	11	1	19	48	12	2	14	30	13	1	30
13	1	23	60	16	4	48	100	12	4	42	88	13	4	78	160	14	7	238
14	3	72	204	18	1	15	29	13	2	45	156	14	2	54	110	15	3	59
15	2	50	144	19	4	60	144	14	3	26	60	15	4	119	200	16	4	119
16	2	66	158	20	3	51	111	15	2	126	275	16	12	422	768	17	3	106
17	4	168	388					16	3	141	297	17	2	74	132	18	7	132
18	1	42	95					17	2	104	213	18	3	108	212	19	5	174
21	2	84	324					19	1	98	238	19	4	172	292	20	7	132
22	1	66	137					20	1	51	137	20	5	255	430	21	3	106
25	1	74	202					21	2	256	600	21	4	215	356	22	6	161
27	1	95	198					22	1	146	291	23	4	248	416	24	2	375
28	1	89	216					23	2	317	668	24	4	263	432	25	3	588
29	1	90	225					24	4	180	334	25	1	73	121	26	3	216
31	1	122	249					25	4	94	185	26	4	329	512	27	1	363
33	1	130	293					27	2	270	591	27	2	134	306	28	6	23
35	1	147	286					28	2	206	408	28	2	166	296	29	4	17
38	1	187	285					29	1	342	633	29	1	89	153	30	1	88
40	1	156	350					30	1	128	228	30	1	120	189	31	1	159
								33	1	125	266	31	1	127	200	32	2	106
								35	1	146	286	34	3	134	218	33	1	171
								39	1	169	329	35	1	157	283	34	2	180
								41	1	198	351	36	1	152	327	35	1	189
								47	1	257	404	40	1	164	397	38	2	400
												48	1	300	456		1	236

H	I	J	K	L	M	N	
1	1	0	0	0	0	0	1
2	7	3	11	19	33	7	2
3	12	19	36	4	8	24	45
4	3	8	18	5	6	27	51
5	9	27	81	6	2	6	26
6	2	10	24	7	3	23	50
7	1	4	16	8	2	20	40
9	2	18	46	9	2	12	55
11	3	42	90	10	3	41	102
13	6	117	240	11	2	35	84
14	2	54	110	12	1	20	46
15	4	119	200	13	3	74	180
16	3	106	192	14	1	30	61
17	2	74	132	16	1	44	83
18	4	144	282	17	1	47	92
19	2	86	146	18	3	123	290
20	7	357	602	20	3	201	347
21	3	161	267	21	1	65	123
22	1	63	98	22	1	78	134
23	3	186	312	23	1	88	154
24	2	131	216	24	3	243	443
25	2	146	242	25	2	200	330
26	1	82	128	26	4	376	706
27	6	402	918	27	2	197	377
28	3	249	444	28	2	232	408
29	1	89	153	29	3	375	696
31	4	426	684	30	3	396	711
32	1	103	180	31	3	351	726
33	1	120	189	32	3	378	748
42	1	188	276	33	3	501	773
50	1	260	363	34	3	513	846
				35	3	516	881
				36	1	170	294
				37	2	382	614
							40
							1
							214
							357
							37
							3
							574
							921
							46
							1
							254
							428
							38
							1
							159
							228
							47
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2
							432
							926
							395
							173
							40
							240
							380
							42
							2

O	P			Q		R		S		T		U																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				</

V	W			X			Y			Z			AA			AB									
30	3	362	660	37	1	160	305	34	2	242	564	34	1	121	282	32	1	125	246			39	1	199	355
32	1	131	259	38	1	126	317	35	1	146	308	35	2	293	616	33	1	112	265			41	1	221	385
33	2	298	526	39	1	177	316	37	1	160	305	36	1	154	287	34	3	363	846			45	1	274	433
35	2	310	598	41	1	193	344	38	1	126	317	37	1	160	305	35	1	146	308			50	1	334	501
38	1	191	319	45	1	228	381	39	1	177	316	39	1	177	316	36	2	308	574						
39	2	405	702					41	2	387	688	42	1	195	350	38	1	126	317						
40	1	213	373					47	1	245	404					39	1	177	316						
41	1	191	364													43	1	164	359						
51	1	343	495													54	1	313	484						

References

1. Albert, R., Barabási, A.L.: Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002)
2. Al-Khayyal, F.A., Falk, J.E.: Jointly constrained biconvex programming. *Math. Oper. Res.* **8**, 273–286 (1983)
3. Axelrod, R.: *The Complexity of Cooperation*. Princeton University Press, Princeton (1997)
4. Balcan, D., Colizza, V., Gonçalves, H., Hu, H., Ramasco, J.J., Vespignani, A.: Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci. USA* **106**, 21484–21489 (2009)
5. Bansal, S., Khandelwal, S., Meyers, L.A.: Exploring biological network structure with clustered random networks. *BMC Bioinforma.* **10**, 405 (2009)
6. Barabási, A.L., Albert, R.: Emergence of scaling in random networks. *Science* **286**, 509–512 (1999)
7. Barrat, A., Barthélemy, M., Vespignani, A.: *Dynamical processes on complex networks*. Cambridge University Press, New York (2008)
8. Blitzstein, J., Diaconis, P.: A sequential importance sampling algorithm for generating random graphs with prescribed degrees. Unpublished paper (2006). <http://nrs.harvard.edu/urn-3:HUL.InstRepos:2757225>
9. Boginski, V., Butenko, S., Pardalos, P.M.: Modeling and optimization in massive graphs. In: Pardalos, P.M., Wolkowitz, H. (eds.) *Novel Approaches to Hard Discrete Optimization*, vol. 37, pp. 17–39. American Mathematical Society, Fields Institute Communications, Providence (2003)
10. Boginski, V., Commander, C.W.: Identifying critical nodes in protein–protein interaction networks. In: Butenko, S., Chaovalitwongse, W.A., Pardalos, P.M. (eds.) *Clustering Challenges in Biological Networks*, pp. 153–167. World Scientific, Singapore (2009)
11. Boguñá, M., Pastor-Satorras, R.: Epidemic spreading in correlated complex networks. *Phys. Rev. E* **66**, 047104 (2002)
12. Bollobas, B.: A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *Eur. J. Combin.* **1**, 311–316 (1980)
13. Butenko, S., Chaovalitwongse, W.A., Pardalos, P.M.: *Clustering Challenges in Biological Networks*. World Scientific, Singapore (2009)
14. Callaway, D.S., Hopcroft, J.E., Kleinberg, J.M., Newman, M.E.J., Strogatz, S.H.: Are randomly grown graphs really random. *Phys. Rev. E* **64**, 041902 (2001)
15. Calvo-Armengol, A., Jackson, M.O.: The effects of social networks on employment and inequality. *Am. Econ. Rev.* **94**, 426–454 (2004)
16. Chakrabarti, D., Zhan, Y., Faloutsos, C.: R-MAT: a recursive model for graph mining. In: *Fourth SIAM International Conference on Data Mining* (2004)
17. Chung, F., Lu, L.: Connected components in random graphs with given expected degree sequences. *Ann. Comb.* **6**, 125–145 (2002)
18. IBM Corp.: *IBM ILOG CPLEX V11.0. User's Manual for CPLEX* (2009)
19. Davidsen, J., Ebel, H., Bornholdt, S.: Emergence of a small world from local interactions: modeling acquaintance networks. *Phys. Rev. Lett.* **88**, 128701 (2002)
20. Dorogovtsev, S.N., Mendes, J.F.F., Samukhin, A.N.: How to construct a correlated net. Unpublished paper (2002) Available from: ArXiv Condensed Matter e-prints
21. Erdős, P., Rényi, A.: On random graphs, I. *Publ. Math-Debrecen* **6**, 290–297 (1959)
22. Erdős, P., Gallai, T.: Graphs with prescribed degrees of vertices. *Mat. Lapok* **11**, 264–274 (1960)
23. Faloutsos, M., Faloutsos, P., Faloutsos, C.: On power-law relationships of the Internet topology. *Comp. Comm. R.* **29**, 251–262 (1999)
24. Floudas, C.A.: *Nonlinear and Mixed-Integer Optimization: Fundamentals and Applications*. Oxford University Press, New York (1995)
25. Fu, P., Liao, K.: An evolving scale-free network with large clustering coefficient. *Proc. Natl. Acad. Sci. USA* **99**, 7821–7826 (2002)
26. Hakimi, S.L.: On realizability of a set of integers as degrees of the vertices of a linear graph. *I. J. Soc. Ind. Appl. Math.* **10**, 496–506 (1962)
27. Havel, V.: A remark on the existence of finite graphs (in Czech). *Casopis Pest. Mat.* **80**, 477–480 (1955)
28. Jeong, H., Tombor, B., Albert, R., Oltavi, Z.N., Barabási, A.L.: The large-scale organization of metabolic networks. *Nature* **407**, 651–654 (2000)

29. Kevrekidis, I.G., Gear, C.W., Hummer, G.: Equation-free: the computer-aided analysis of complex multiscale systems. *AIChE J.* **50**, 1346–1355 (2004)
30. Kevrekidis, I.G., Gear, C.W., Hyman, J.M., Kevrekidis, P.G., Runborg, O., Theodoropoulos, C.: Equation-free, coarse-grained multiscale computation: enabling microscopic simulators to perform system-level analysis. *Commun. Math. Sci.* **1**, 715–762 (2003)
31. Kim, B.J.: Performance of networks of artificial neurons: the role of clustering. *Phys. Rev. E* **69**, 045101 (2004)
32. Kingsford, C., Zaslavsky, E., Singh, M.: A compact mathematical programming formulation for DNA motif finding. In: *Proceedings of the Combinatorial Pattern Matching. Lecture Notes in Computer Science*, vol. 4009, pp. 233–245 (2006)
33. McCormick, G.P.: Computability of global solutions to factorable nonconvex programs. Part I. Convex underestimating problems. *Math. Program.* **10**, 147–175 (1976)
34. Nemhauser, G.L., Wolsey, L.A.: *Integer and Combinatorial Optimization*. Wiley, New York (1998)
35. Newman, M.E.J.: The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* **98**, 404–409 (2001)
36. Newman, M.E.J.: Properties of highly clustered networks. *Phys. Rev. E* **68**, 026121 (2003)
37. Newman, M.E.J.: The structure and function of complex networks. *SIAM Rev.* **45**, 167–256 (2003)
38. Newman, M.E.J., Watts, D.J., Strogatz, S.H.: Random graph models of social networks. *Proc. Natl. Acad. Sci. USA* **99**, 2566–2572 (2002)
39. Nier, E., Yang, J., Yorulmazer, T., Alentorn, A.: Network models and financial stability. *J. Econ. Dyn. Control* **31**, 2033–2060 (2007)
40. Qiang, Q., Nagurney, A.: A unified network performance measure with importance identification and the ranking of network components. *Optim. Lett.* **2**, 127–142 (2008)
41. Ravasz, E., Barabási, A.L.: Hierarchical organization in complex networks. *Phys. Rev. E* **67**, 026112 (2003)
42. Qiang, Q., Nagurney, A., Rebennack, S., Arulselvan, A., Eleftheriadou, L., Pardalos, P.M.: Complexity analysis for maximum flow problems with arc reversals. *J. Comb. Optim.* **19**, 200–216 (2010)
43. Serrano, M.A., Boguñá, M.: Tuning clustering in random networks with arbitrary degree distributions. *Phys. Rev. E* **72**, 036133 (2005)
44. Sierskma, G., Hoogeveen, H.: Seven criteria for integer sequences being graphic. *J. Graph Theor.* **15**, 223–231 (1991)
45. Strogatz, S.H.: Exploring complex networks. *Nature* **410**, 268–276 (2001)
46. Valente, T.W.: Social network thresholds in the diffusion of innovations. *Soc. Netw.* **18**, 441–458 (1996)
47. Vazquez, A., Boguñá, M., Moreno, Y., Pastor-Satorras, R., Vespignani, A.: Topology and correlations in structured scale-free networks. *Phys. Rev. E* **67**, 046111 (2003)
48. Vazquez, A., Pastor-Satorras, R., Vespignani, A.: Large-scale topological and dynamical properties of the Internet. *Phys. Rev. E* **65**, 066130 (2002)
49. Volz, E.: Random networks with tunable degree distribution and clustering. *Phys. Rev. E* **70**, 056115 (2004)
50. Watts, D.J., Dodds, P.S.: Influentials, networks, and public opinion formation. *J. Consum. Res.* **34**, 441–458 (2007)
51. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998)
52. Williams, R.J., Martinez, N.D.: Simple rules yield complex food webs. *Nature* **404**, 180–183 (2000)
53. Wormald, N.C.: Some problems in the enumeration of labelled graphs. *B. Aust. Math. Soc.* **21**, 159–160 (1980)
54. Xanthopoulos, P., Arulselvan, A., Boginski, V., Pardalos, P.M.: A retrospective review of social networks. In: *Advances in Social Networks Analysis and Mining (ASONAM-2009)*, pp. 300–305. IEEE Computer Society, USA (2009)
55. Zaslavsky, E., Singh, M.: A combinatorial complex optimization approach for diverse motif finding applications. *Algorithm. Mol. Biol.* 1–13 (2006)