A simple Havel-Hakimi type algorithm to realize graphical degree sequences of directed graphs*

Péter L. Erdős and István Miklós

A. Rényi Institute of Mathematics, Hungarian Academy of Sciences, Budapest, PO Box 127, H-1364, Hungary

{elp} {miklosi}@renyi.hu

Zoltán Toroczkai

Interdisciplinary Center for Network Science and Applications and Department of Physics University of Notre Dame

Notre Dame, IN, 46556, USA

toro@nd.edu

May 29, 2009

Abstract

One of the simplest ways to decide whether a given finite sequence of positive integers can arise as the degree sequence of a simple graph is the greedy algorithm of Havel and Hakimi. This note extends their approach to directed graphs. It also studies cases of some simple forbidden edge-sets. Finally, it proves a result which is useful to design an MCMC algorithm to find random realizations of prescribed directed degree sequences.

AMS subject classification[2000]. 05C07 05C20 90B10 90C35

Keywords. network modeling; directed graphs; degree sequences; greedy algorithm

1 Introduction

The systematic study of graph theory (or more precisely the theory of *linear graphs*, as it was called in that time) began sometimes in the late forties, through seminal works by P. Erdős, P. Turán, W.T. Tutte, and others. One problem which received considerable attention was the existence of certain subgraphs of a

*ELP was partly supported by OTKA (Hungarian NSF), under contract Nos. AT048826 and K 68262. IM was supported by a Bolyai postdoctoral stipend and OTKA (Hungarian NSF) grant F61730. ZT was supported in part by the NSF BCS-0826958, HDTRA 201473-35045 and by Hungarian Bioinformatics MTKD-CT-2006-042794 Marie Curie Host Fellowships for Transfer of Knowledge.

given graph. For example such subgraph could be a perfect matching in a (not necessarily bipartite) graph, or a Hamiltonian cycle through all vertices, etc. Generally these substructures are called *factors*. The first couple of important results of this kind are due to W.T. Tutte who gave necessary and sufficient conditions for the existence of 1-factors and f-factors.

In the case of complete graphs, the existence problem of such factors is considerably easier. In particular, the existence problem of (sometimes simple) undirected graphs with given degree sequences admits even simple greedy algorithms for its solution.

Subsequently, the theory was extended for factor problems of directed graphs as well, but the greedy type algorithm mentioned above is missing even today.

In this paper we fill this gap: after giving a short and comprehensive (but definitely not exhausting) history of the f-factor problem (Section 2), we describe a greedy algorithm to decide the existence of a directed simple graph possessing the prescribed degree sequence (Section 3). In Section 4 we prove a consequence of the previous existence theorem, which is a necessary ingredient for the construction of Markov Chain Monte Carlo (MCMC) methods to sample directed graphs with prescribed degree sequence. Finally in Section 5 we discuss a slightly harder existence problem of directed graphs with prescribed degree sequences where some vertex-pairs are excluded from the constructions. This result can help to efficiently generate all possible directed graphs with a given degree sequence.

2 A brief history (of f-factors)

For a given function $f : V(G) \to \mathbb{N} \cup \{0\}$, an *$f$-factor* of a given simple graph $G(V, E)$ is a subgraph H such that $d_H(v) = f(v)$ for all $v \in V$. One of the very first key results of modern graph theory is due to W.T. Tutte: in 1947 he gave a complete characterization of simple graphs with f-factor in case of $f \equiv 1$ (Tutte's 1-factor theorem, [14]). Tutte later solved the problem of the existence of f-factors for general f's (Tutte's f-factor theorem, [15]). In 1954 he also found a beautiful graph transformation to handle f-factor problems via perfect matchings in bipartite graphs [16]. This also gave a clearly polynomial time algorithm for finding f-factors.

In cases where G is a complete graph, the f-factor problem becomes easier: then we are simply interested in the existence of a graph with a given degree sequence (the exact definitions will come in Section 3). In 1955 P. Havel developed a simple greedy algorithm to solve the *degree sequence problem* for simple undirected graphs ([8]). In 1960 P. Erdős and T. Gallai studied the f-factor problem for the case of a complete graph G, and proved a simpler Tutte-type result for the degree sequence problem (see [3]). As they already pointed out, the result can be derived directly form the original f-factor theorem, taking into consideration the special properties of the complete graph G, but their proof was independent of Tutte's proof and they referred to Havel's theorem.

In 1962 S.L. Hakimi studied the degree sequence problem in undirected graphs with multiple edges ([6]). He developed an Erdős-Gallai type result for this much simpler case, and for the case of simple graphs he rediscovered the greedy algorithm of Havel. Since then this algorithm is referred to as the *Havel–Hakimi algorithm.*

For directed graphs the analogous question of recognizability of a *bi-graphical-sequence* comes naturally. In this case we are given two n-element vectors $\mathbf{d}^+, \mathbf{d}^-$ of non-negative integers. The problem is the existence of a directed graph on n vertices, such that the first vector represents the out-degrees and the second one the in-degrees of the vertices in this graph. In 1957 D. Gale and H. J. Ryser independently solved this problem for simple directed graphs (there are no parallel edges, but loops are allowed), see [5, 13]. In 1958 C. Berge generalized these results for p-graphs where at most p parallel edges are allowed ([1]). (Berge calls the out-degree and in-degree together the *demi-degrees.*) Finally in 1973, the revised version of his book *Graphs* ([2]) gives a solution for the p-graph problem, loops excluded. To show some of the afterlife of these results: D. West in his renowned recent textbook ([17]), discusses the case of simple directed graphs with loops allowed.

The analog of f-factor problems for directed graphs has a sparser history. Øystein Ore started the systematic study of that question in 1956 (see [11, 12]). His method is rather algebraic, and the finite and infinite cases - more or less - are discussed together. The first part developed the tools and proved the directly analog result of Tutte's f-factor problem for finite directed graphs (with loops), while the second part dealt with the infinite case.

In 1962 L.R. Ford and D.R. Fulkerson studied, generalized and solved the "original" f-factor problem for a directed graph $\vec{G}$. Here lower and upper bounds were given for both demi-degrees of the desired subgraph (no parallel edges, no loops) with the original question naturally corresponding to equal lower and upper bounds. The solutions (as well as in Berge's cases) are based on network flow theory.

Finally, in a later paper Hakimi also prove results for bi-graphical sequences, but, he did not present a directed version of his original greedy algorithm (see [7]).

3 A greedy algorithm to realize bi-graphical sequences

A sequence $\mathbf{d} = \{d_1, d_2, \ldots, d_n\}$ of nonnegative integers is called a *graphical sequence* if a simple graph $G(V, E)$ exists on n nodes, $V = \{v_1, v_2, \ldots, v_n\}$, whose degree sequence is $\mathbf{d}$. In this case we say that G *realizes* the sequence $\mathbf{d}$. For simplicity of the notation we will consider only sequences of strictly positive integers ($d_n > 0$) to avoid isolated points. The following, well-known result, was proved independently by P. Havel and S.L. Hakimi.

Theorem 1 (Havel [8], Hakimi [6]) *There exists a simple graph with degree sequence $d_1 > 0$, $d_2 \geq \ldots \geq d_n > 0$ $(n \geq 3)$ if and only if there exists one* *with degree sequence $d_2 - 1, \ldots, d_{d_1+1} - 1, d_{d_1+2}, \ldots, d_n$. (Note that there is no prescribed ordering relation between d_1 and the other degrees.)*

This can be proved using a recursive procedure, which transforms any realization of the degree sequence into the form described in the Theorem, by a sequence of two-edge swaps.

A *bi-degree-sequence* (or *BDS* for short) $(\mathbf{d}^+, \mathbf{d}^-) = (\{d_1^+, d_2^+, \ldots, d_n^+\}, \{d_1^-, d_2^-, \ldots, d_n^-\})$ of nonnegative integers is called a *bi-graphical sequence* if there exists a simple directed graph (digraph) $\vec{G}(V, \vec{E})$ on n nodes, $V = \{v_1, v_2, \ldots, v_n\}$, such that the out-degree and in-degree sequences together form $(\mathbf{d}^+, \mathbf{d}^-)$. (That is the in-degree of vertex v_j is d_j^+ and its out-degree is d_j^-.) In this case we say that $\vec{G}$ *realizes* our BDS. For simplicity, we will consider only sequences of strictly positive integer BDS's, that is each degree is ≥ 0 and $d_j^+ + d_j^- > 0$, to avoid isolated points.

Our goal is to prove a Havel-Hakimi type algorithm to realize bi-graphical sequences. To that end we need to introduce the notion of normal order: we say that the BDS is in *normal order* if the entries satisfy the following properties: for each $i = 1, ..., n-2$ we either have $d_i^- > d_{i+1}^-$ or $d_i^- = d_{i+1}^-$ and $d_i^+ \geq d_{i+1}^+$. Clearly, all BDS-s can be arranged into normal order. Note that we made no ordering assumption about node v_n (the pair d_n^+, d_n^-).

Theorem 2 *Assume that the BDS $(\mathbf{d}^+, \mathbf{d}^-)$ with $d_j^+ + d_j^- > 0$, $j \in [1, n]$ is in normal order and $d_n^- > 0$. Then $(\mathbf{d}^+, \mathbf{d}^-)$ is bi-graphical if and only if the BDS*

$$\Delta_k^+ = \begin{cases} d_k^+ & \text{if} \quad k \neq n \\ 0 & \text{if} \quad k = n \,, \end{cases} \tag{1}$$

$$\Delta_k^- = \begin{cases} d_k^- - 1 & \text{if} \quad k \leq d_n^+ \\ d_k^- & \text{if} \quad k > d_n^+ \end{cases} \,, \tag{2}$$

with zero elements removed (those j for which $\Delta_j^+ = \Delta_j^- = 0$) is bi-graphical.

Before we start the proof, we want to emphasize the similarity between this result and the original HH-algorithm. As in the undirected case, using Theorem 2, we can find a proper realization of graphical bi-degree sequences, greedily.

Indeed: choose any vertex v_n with non-zero out-degree from the sequence, arrange the rest in normal order, then make d_n^- connections from v_n to nodes with largest in-degrees, thus constructing the out-neighborhood of v_n in the final realization. Next, remove the vertices (if any) from the remaining sequence that have lost both their in- and out- degrees in the process, pick a node with non-zero out-degree, then arrange the rest in normal order. Applying Theorem 2 again, we find the final out-neighborhood of our second pick vertex. Step by step we find this way the out-neighborhood of all vertices, while their in-neighborhoods get defined eventually (being exhausted by incoming edges). Note, that every vertex in this process is picked at most once, namely, when its out-neighborhood is determined by the Theorem, and never again after that.

Our forthcoming proof is not the simplest, however, we use a more general setup to shorten the proofs of later results.

First, we define the partial order $\preceq$ among k-element vectors of increasing positive integers: we say $\mathbf{a} \preceq \mathbf{b}$ iff for each $j = 1, ..., k$ we have $a_j \le b_j$.

A *possible out-neighborhood* (or *PON* for short) of vertex v_n is a d_n^+-element subset of $V \setminus \{v_n\}$ which is a candidate for an out-neighborhood of v_n in some graphical representation. (In essence, a PON can be any d_n^+-element subset of $V \setminus \{v_n\}$ but later on we may consider some restrictions on it.) Let A be a PON of v_n. Then denote by $\boldsymbol{i}(A)$ the vector of the increasingly ordered subscripts of the elements of A. (For example, if $A = \{v_2, v_4, v_9\}$, then $\boldsymbol{i}(A) = (2, 4, 9)$.) Let A and B two PONs of v_n. We write:

$$B \preceq A \quad \Leftrightarrow \quad \boldsymbol{i}_B \preceq \boldsymbol{i}_A \ . \tag{3}$$

In this case we also say that B is *to the left* of A. (For example, $B = \{v_1, v_2, v_6, v_7\}$ is to the left of $A = \{v_2, v_4, v_6, v_9\}$.)

Definition 3 *Consider a bi-graphical BDS sequence* $(\mathbf{d}^+, \mathbf{d}^-)$ *and let* A *be a PON of* v_n*. The* A-reduced BDS $\left(\mathbf{d}^+\big|_A, \mathbf{d}^-\big|_A\right)$ *is defined as:*

$$d_k^+\big|_A \quad = \quad \begin{cases} d_k^+ & \text{if} \quad k \neq n \\ 0 & \text{if} \quad k = n \ , \end{cases} \tag{4}$$

$$d_k^-\big|_A \quad = \quad \begin{cases} d_k^- - 1 & \text{if} \quad k \in \boldsymbol{i}(A) \\ d_k^- & \text{if} \quad k \notin \boldsymbol{i}(A). \end{cases} \tag{5}$$

In other words, if A is a PON in a BDS, then the reduced degree sequence $\left(\mathbf{d}^+\big|_A, \mathbf{d}^-\big|_A\right)$ is obtained by removing the out-edges of node v_n (according to the possible out-neighborhood A). As usual, if for one subscript k in the A-reduced BDS we have $d_k^+\big|_A = d_k^-\big|_A = 0$ then the vertex with this index is to be removed from the bi-degree sequence.

Lemma 4 *Let* $(\mathbf{d}^+, \mathbf{d}^-)$ *be a BDS, and let* A *be a possible out-neighborhood of* v_n*. Furthermore let* B *be another PON with* $B = A \setminus \{v_k\} \cup \{v_i\}$ *where* $d_i^- \ge d_k^-$ *and in case of* $d_i^- = d_k^-$ *we have* $d_i^+ \ge d_k^+$*. Then if* $(\mathbf{D}^+, \mathbf{D}^-) := \left(\mathbf{d}^+\big|_A, \mathbf{d}^-\big|_A\right)$ *is bi-graphical, so is* $\left(\mathbf{d}^+\big|_B, \mathbf{d}^-\big|_B\right)$*.*

Proof. Since our A-reduced BDS $(\mathbf{D}^+, \mathbf{D}^-)$ is bi-graphical, there exists a directed graph $\vec{G}$ which realizes the bi-degree sequence $(\mathbf{D}^+, \mathbf{D}^-)$. We are going to show that in this case there exists a directed graph $\vec{G}'$ which realizes the BDS $\left(\mathbf{d}^+\big|_B, \mathbf{d}^-\big|_B\right)$. In the following, $v_a v_b$ will always mean a directed edge from node v_a to node v_b. Let us now construct the directed graph $\vec{G}_1$ by adding $v_n v$ directed edges for each $v \in A$. (Since according to (4), in $(\mathbf{D}^+, \mathbf{D}^-)$ the out-degree of v_n is equal to zero, no parallel edges are created.) The bi-degree-sequence of $\vec{G}_1$ is $(\mathbf{d}^+, \mathbf{d}^-)$. Our goal is to construct another realization $\vec{G}_1'$ of $(\mathbf{d}^+, \mathbf{d}^-)$ such that the deletion of the out-edges of v_n in the latter produces the BDS $\left(\mathbf{d}^+\big|_B, \mathbf{d}^-\big|_B\right)$.

By definition we have $v_nv_k \in \vec{E}_1$, (the edge set of $\vec{G}_1$) but $v_nv_i \notin \vec{E}_1$. At first assume that there exists a vertex v_ℓ ($\ell \neq i,k,n$), such that $v_\ell v_i \in \vec{E}_1$ but $v_\ell v_k \notin \vec{E}_1$. (When $d_i^- > d_k^-$ then this happens automatically, however if $d_i^- = d_k^-$ and $v_kv_i \in \vec{E}_1$ then it is possible that the in-neighborhood of v_i and v_k are the same - except of course v_k, v_i themselves and v_n.) This means that now we can *swap* the edges v_nv_k and $v_\ell v_i$ into v_nv_i and $v_\ell v_k$. (Formally we create the new graph $\vec{G}_1' = (V, \vec{E}_1')$ such that $\vec{E}_1' = \vec{E}_1 \setminus \{v_nv_k, v_\ell v_i\} \cup \{v_nv_i, v_\ell v_k\}$.) This achieves our wanted realization.

Our second case is when $d_i^- = d_k^-$, $v_kv_i \in \vec{E}_1$, and furthermore

$$\text{for each } \ell \neq i,k,n \quad \text{we have} \quad v_\ell v_i \in \vec{E}_1 \Leftrightarrow v_\ell v_k \in \vec{E}_1. \tag{6}$$

It is important to observe that in this case $v_iv_k \notin \vec{E}_1$: otherwise some v_ℓ would not satisfy (6) (in order to keep $d_i^- = d_k^-$).

Now, if there exists a subscript m (different from k,i,n) such that $v_iv_m \in \vec{E}_1$ but $v_kv_m \notin \vec{E}_1$, then we create the required new graph $\vec{G}_1'$ by applying the following triple swap (or three-edge swap): we exchange the directed edges v_nv_k, v_kv_i and v_iv_m into v_nv_i, v_iv_k and v_kv_m.

By our our assumption we have $d_i^+ \geq d_k^+$. On one hand side if $d_i^+ > d_k^+$ holds then due to the properties $v_kv_i \in \vec{E}$ and $v_iv_k \notin \vec{E}$, there exist at least two subscripts $m_1, m_2 \neq i,k$ such that $v_iv_{m_j} \in \vec{E}$ but $v_kv_{m_j} \notin \vec{E}$ and at least one of them differs from n. Thus, when $d_i^+ > d_k^+$, we do find such an m for which the triple swap above can be performed.

The final case is when $d_i^- = d_k^-$ and $d_i^+ = d_k^+$. If vertex v_m does not exist, then we must have $v_iv_n \in \vec{E}_1$ (to keep $d_i^+ = d_k^+$), and in this case clearly, $v_kv_n \notin \vec{E}_1$. Therefore, in this (final) case the graphical realization $\vec{G}_1$ has the properties $v_nv_k, v_kv_i, v_iv_n \in \vec{E}_1$ and $v_nv_i, v_iv_k, v_kv_n \notin \vec{E}_1$. Then the triple swap

$$\vec{E}_1' := \vec{E}_1 \setminus \{v_nv_k, v_kv_i, v_iv_n\} \cup \{v_nv_i, v_iv_k, v_kv_n\} \tag{7}$$

will produce the required new graphical realization $\vec{G}_1'$. □

Observation 5 *For later reference it is important to recognize that in all cases above, the transformations from one realization to the next one happened with the use of two-edge or three-edge swaps.*

Lemma 6 *Let* $(\mathbf{d}^+, \mathbf{d}^-)$ *be an BDS and let* A *and* C *be two possible out-neighborhoods of* v_n. *Furthermore assume that* $C \preceq A$, *that is* C *is to the left of* A. *Finally assume that vertices in* $A \cup C$ *are in normal order. Then if* $\left(\mathbf{d}^+\big|_A, \mathbf{d}^-\big|_A\right)$ *is bi-graphical, so is* $\left(\mathbf{d}^+\big|_C, \mathbf{d}^-\big|_C\right)$.

Proof. Since C is to the left of A therefore, there is a (unique) bijection $\phi : C \setminus A \to A \setminus C$ such that $\forall c \in C \setminus A : \boldsymbol{i}(\{c\}) < \boldsymbol{i}(\{\phi(c)\})$ (the subscript of vertex c is smaller than the subscript of vertex $\phi(c)$). (For example, if $A = \{v_4, v_5, v_6, v_7, v_8, v_9\}$ and $C = \{v_1, v_2, v_3, v_5, v_7, v_8\}$, then $C \setminus A = \{v_1, v_2, v_3\}$, $A \setminus C = \{v_4, v_6, v_9\}$, and ϕ is the map $\{v_1 \leftrightarrow v_4, v_2 \leftrightarrow v_6, v_3 \leftrightarrow v_9\}$).

To prove Lemma 6 we apply Lemma 4 recursively for each $c \in C \setminus A$ (in arbitrary order) to exchange $\phi(c) \in A$ with $c \in C$, preserving the graphical character at every step. After the last step we find that the sequence reduced by C is graphical. $\square$

Proof of Theorem 2: We can easily achieve now the required graphical realization of $(\mathbf{d}^+, \mathbf{d}^-)$ if we use Lemma 6 with the current A, and $C = \{v_1, ..., v_{d_n^+}\}$. We can do that since $(\mathbf{d}^+, \mathbf{d}^-)$ is in normal order, therefore the assumptions of Lemma 6 always hold. $\square$

4 A simple prerequisite for MCMC algorithms to sample directed graphs with given BDS

In practice it is often useful to choose uniformly a random element from a set of objects. A frequently used tool for that task is a well-chosen Markov-Chain Monte-Carlo method (MCMC for short). To that end, a graph is established on the objects and random walks are generated on it. The edges represent operations which can transfer one object to the other. If the Markov chain can steps from an object x to object y with non-zero probability, then it must be able to jump to x from y with non-zero probability (reversibility). If the graph is connected, then applying the well-known Metropolis-Hastings algorithm, it will yield a random walk converging to the uniform distribution starting from an arbitrary (even fixed) object.

To be able to apply this technique we have to define our graph (the Markov chain) $\mathcal{G}(\mathbf{d}^+, \mathbf{d}^-) = (\mathcal{V}, \mathcal{E})$. The vertices are the different possible realizations of the bi-graphical sequence $(\mathbf{d}^+, \mathbf{d}^-)$. An edge represents an operation consisting of a two or three-edge swap which transforms the first realization into the second one. (For simplicity, sometimes we just say *swap* for any of them.) We will show:

Theorem 7 *Let $\vec{G}, \vec{G}^*$ be two realizations of the same bi-graphical sequence $(\mathbf{d}^+, \mathbf{d}^-)$. Then there is a sequence of graphs $\vec{G} = \vec{G}_0, \vec{G}_1, ..., \vec{G}_m = \vec{G}^*$ where all graphs are graphical realizations of $(\mathbf{d}^+, \mathbf{d}^-)$ and for each $i = 0, ..., m-1$ there is a two- or three-edge swap, which transform G_i into G_{i+1}.*

Remark: In the case of undirected graphs the (original) analogous observation (needing only two-edges swaps) was proved by H.J. Ryser ([13]).

Proof. At first it easy to see that if a swap transforms a graph $\vec{H}$ into $\vec{H}'$ then the "inverse swap" transforms $\vec{H}'$ into $\vec{H}$. Therefore, our plan is this: at first we fix a certain *reference realization* $\vec{\mathbb{G}}$, then we take the realization $\vec{H}$ and step by step we transform into $\vec{\mathbb{G}}$ in such a way that an increasing number of vertices of the current realization will have the same out-neighbors as in the reference realization $\vec{\mathbb{G}}$.

In order to do that, we consider the original BDS $(\mathbf{d}^+, \mathbf{d}^-)$ and take its *standard normal order* $(\mathbf{\Delta}^+, \mathbf{\Delta}^-)$ defined as the normal order which has the following extra property: whenever for two subscripts $i < j$ we have $\Delta_i^+ = \Delta_j^+$

and $\Delta_i^- = \Delta_j^-$ then $\pi^{-1}(i) < \pi^{-1}(j)$ where the permutation $\pi : [n] \to [n]$ gives the correspondence among the subscripts in the BDS's $(\mathbf{d}^+, \mathbf{d}^-)$ and $(\mathbf{\Delta}^+, \mathbf{\Delta}^-)$.

Now the reference realization $\vec{\mathbb{G}}$ will be constructed as follows: take the standard normal order $(\mathbf{\Delta}^+, \mathbf{\Delta}^-)$ and apply our greedy algorithm for the last subscript j for which $\Delta_j^- > 0$. Next, reduce the BDS with the out-neighbors $A_{\pi^{-1}(j)}$ (the $k = \pi^{-1}(j)$ element of the set A) of v_j and repeat the whole process again. (Here we will keep the $(0, 0)$ degree-pairs to make the construction of the current permutation π easier.) The graph emerging in the end is $\vec{\mathbb{G}}$. Now it is enough to prove the following statement:

Lemma 8 *For any realization $\vec{H}$ of $(\mathbf{d}^+, \mathbf{d}^-)$ there is a series of swaps that transform $\vec{H}$ into $\vec{\mathbb{G}}$.*

Proof. We will generate our realization (swap) sequence starting from $\vec{H}$ towards $\vec{\mathbb{G}}$ in the same order as the reference realization was originally created. Take the standard normal order of the original BDS and pick the last vertex v_j with non-zero out-degree. We know its out-neighborhood in (the current) $\vec{H}$. We also know the out-neighborhood of the same vertex in $\vec{\mathbb{G}}$: that will serve as the set C in Lemma 6. Due to the construction of $\vec{\mathbb{G}}$ it must be to the left of A, therefore, the lemma applies. It creates a realization $\vec{H}_1$, where v_j (and, by the way, all vertices with larger subscripts in the standard normal order) has the same out-neighborhood as in $\vec{\mathbb{G}}$.

Next, reduce the degree sequence with this new out-neighborhood C and continue recursively: repeat the process with this new BDS. Since we proceed in the same order of the vertices as in the creation of the original reference graph $\vec{\mathbb{G}}$, consequently, the current C out-neighborhood will be always to the left of A, and therefore Lemma 6 will always apply and this finishes the proof. $\square$

Proof of Theorem 7: Now we can easily finish the proof by finding the swap chains $\mathcal{C}$ and $\mathcal{C}^*$ from $\vec{G}$ to $\vec{\mathbb{G}}$ and from $\vec{G}^*$ to $\vec{\mathbb{G}}$ and then the concatenated chain $\mathcal{C} \circ \overline{\mathcal{C}^*}$ (where the over-line denotes the chain of "inverse" swaps) will transform $\vec{G}$ into $\vec{G}^*$. $\square$

5 Is a BDS bi-graphical when one of its vertex's out-neighborhood is constrained?

In network modeling of complex systems (for a rather general reference see [10]) one usually defines a (di)graph with components of the system being represented by the nodes, and the interactions (usually directed) amongst the components being represented as the edges of this digraph. Typical cases include biological networks, such as the metabolic network, signal transduction networks, gene transcription networks, etc. The graph is usually inferred from empirical observations of the system and it is uniquely determined if one can specify all the connections in the graph. Frequently, however, the data available from the system is incomplete, and one cannot uniquely determine this graph. In this case there will be a *set* $\mathcal{D}$ of (di)graphs satisfying the existing data, and one can be faced with:

(i) finding a typical element of the class $\mathcal{D}$,

(ii) or generating *all* elements of the class $\mathcal{D}$.

(A more complete analysis of this phenomenon can be found in [9].) In Section 4 we already touched upon problem (i) when $\mathcal{D}$ is the class of all directed graphs of a given BDS. The analogous Problem (ii) for undirected graphs was recently addressed in [9] which provides an economical way of constructing all elements from $\mathcal{D}$. In this Section we give a prescription based on the method from [9], to solve (ii) for the case of all directed graphs with prescribed BDS. This is particularly useful from the point of view of studying the abundance of motifs in real-world networks: one needs to know first *all* the (small) subgraphs, or motifs, before we study their statistics from the data.

Before we give the details, it is perhaps worth making the following remark: Clearly, one way to solve problem (i) would be to first solve problem (ii), then choose uniformly from $\mathcal{D}$. However, in (those very small) cases when reasonable answers can be expected for problem (ii), problem (i) is rather uninteresting. In general, however, (i) cannot be solved efficiently by the use of (ii).

We start the discussion of problem (ii) with pointing out that our new, directed Havel- Hakimi type algorithm is unable to generate all realization of a prescribed DBS (see Figure 1).

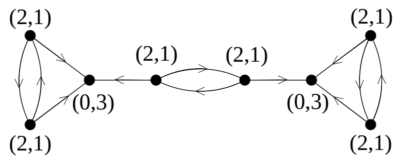

Figure 1: *This graph cannot be obtained by the directed Havel-Hakimi procedure. The integers indicate node degrees.*

The situation is very similar to the non-directed case, see [9]. The directed HH-algorithm must start with a vertex with degree-pair $(2, 1)$, therefore the two vertices of degree-pair $(0, 3)$ must be out-neighbors of the same vertex - not for the graph in the Figure.

One possible way to overstep this shortage is to discover systematically all possible out-connections from a given vertex v in all realizations of the prescribed graphical BDS. We do not know a greedy algorithm to achieve this. The next best thing we can do is to develop a greedy algorithm to *decide* whether a given (sub)set of prescribed out-neighbors of v would prevent to find a realization of the BDS containing those prescribed out-neighbors. In the following, we describe such a greedy algorithm. (It is perhaps interesting to note that this latter problem can be considered as a very special directed f-factor problem.)

For start, we consider a $(\mathbf{d}^+, \mathbf{d}^-)$ bi-degree sequence together with a forbidden vertex set F whose elements are not allowed to be out-neighbors of vertex v_n. (Or, just oppositely, we can imagine that we already have decided that those vertices will become out-neighbors of v_n and the BDS is already updated accordingly. The forbidden vertex set governs only the out-neighbors, since in the process the in-neighbors are born "automatically".) It is clear that $|F| + 1 + d_n^- \leq n$ must hold for the existence of a graphical realization of this F*-restricted* BDS.

Assume that the vertices are enumerated such a way that subset F consists of vertices $v_{n-|F|}, ..., v_{n-1}$ and vertices $V' = \{v_1, ..., v_{n-|F|-1}\}$ are in normal order. (We can also say that we apply a permutation on the subscripts accordingly.) Then we say that the BDS is in F*-normal order.*

Definition 9 *Consider a bi-graphical BDS sequence* $(\mathbf{d}^+, \mathbf{d}^-)$ *in* F*-normal order, and let* A *be a PON. The* A-reduced BDS $\left(\mathbf{d}^+\big|_A, \mathbf{d}^-\big|_A\right)$ *is defined as in (4) and (5), while keeping in mind the existence of an* F *set to the right of* A.

In other words, if A is a PON in an F-restricted BDS, then the reduced degree sequence $\left(\mathbf{d}^+\big|_A, \mathbf{d}^-\big|_A\right)$ is still obtained by removing the out-edges of node v_n (according to the possible out-neighborhood A).

Finally, one more notation: let $(\mathbf{d}^+, \mathbf{d}^-)$ be a BDS, F a forbidden vertex subset of V and denote by $F[k]$ the set of the first k vertices in the F-normal order.

Theorem 10 *Let* A *be any PON in the* F*-restricted* $(\mathbf{d}^+, \mathbf{d}^-)$ *BDS, which is in* F*-normal order. Then if the* A*-reduced BDS* $\left(\mathbf{d}^+\big|_A, \mathbf{d}^-\big|_A\right)$ *is graphical, then the* $F[d_n^+]$*-reduced BDS* $\left(\mathbf{d}^+\big|_{F[d_n^+]}, \mathbf{d}^-\big|_{F[d_n^+]}\right)$ *is graphical as well.*

Proof. It is trivial: Lemma 6 applies. $\square$

This statement gives us indeed a greedy way to check whether there exists a graphical realization of the F-restricted bi-degree sequence $(\mathbf{d}^+, \mathbf{d}^-)$: all we have to do is to check only whether the $F[d_n^+]$-reduced BDS $\left(\mathbf{d}^+\big|_{F[d_n^+]}, \mathbf{d}^-\big|_{F[d_n^+]}\right)$ is graphical.

Finally, we want to remark that, similarly to the indirected case, Theorem 10 is suitable to speed up the generation of all possible graphical realizations of a BDS. The details can be found in [9] which is a joint work of these authors with Hyunju Kim and László A. Székely.

Acknowledgements

The authors acknowledge useful discussions with Gábor Tusnády, Éva Czabarka and László A. Székely and Hyunju Kim. ZT would also like to thank for the kind hospitality extended to him at the Alfréd Rényi Institute of Mathematics, where this work was completed.

References

[1] C. Berge: *The Theory of Graphs*, Methuen & Co Ltd. London (1962), Chapter 9.

[2] C. Berge: Graphs and Hypergraphs, North Holland Pub. Company, Amsterdam (1973), Chapter 6.

[3] P. Erdős - T. Gallai: Gráfok előírt fokú pontokkal (Graphs with prescribed degree of vertices), *Mat. Lapok*, **11** (1960), 264–274. (in Hungarian)

[4] L.R. Ford - D.R. Fulkerson: Flows in Networks, *RAND Corporation* **R-375-PR** (1962) Chapter 2 Section 11.

[5] D. Gale: A theorem on flows in networks, *Pacific J. Math.* **7** (2) (1957),1073–1082.

[6] S.L. Hakimi: On the realizability of a set of integers as degrees of the vertices of a simple graph. *J. SIAM Appl. Math.* **10** (1962), 496–506.

[7] S.L. Hakimi: On the degrees of the vertices of a directed graph, *J. Franklin Institute* **279** (4) (1965), 290–308.

[8] V. Havel: A remark on the existence of finite graphs. (Czech), *Časopis Pěst. Mat.* **80** (1955), 477–480.

[9] Hyunju Kim - Z. Toroczkai - P.L. Erdős - I. Miklós - L.A. Székely: Degree-based graph construction, submitted *Journal of Physics A* (2009), 1–12.

[10] M.E.J. Newman - A.L. Barabasi - D.J. Watts: *The Structure and Dynamics of Networks* (Princeton Studies in Complexity, Princeton UP) (2006), pp.

[11] O. Ore: Studies on Directed Graphs, I, *The Annals of Math.* (Second Series) **63** (3) (1956), 383–406.

[12] O. Ore: Studies on Directed Graphs, II, *The Annals of Math.* (Second Series) **64** (3) (1956), 142–153.

[13] H. J. Ryser: Combinatorial Properties of Matrices of Zeros and Ones, *Canad. J. Math.* **9** (1957), 371–377.

[14] W.T. Tutte: The factorization of linear graphs, *J. London Math. Soc.* **22** (1947), 107–111.

[15] W.T. Tutte: The factors of graphs, Canadian J. of Mathematics **4** (1952), 314-328.

[16] W.T. Tutte: A short proof of the factors theorem for finite graphs, Canadian J. Of Mathematics **6** (1954), 347–352.

[17] D.B. West: *Introduction to Graph Theory* Prentice Hall, Upper Sadle River, US, (2001), Section 1.4.