Chapter 15.

Graphical degree sequences

The research on networks is a rapidly developing, new interdisciplinary science. Networks emerge everywhere in life, to restrict it only to biological sciences, we mention here the network of biochemical reactions, the network of neurons in the brain, interaction networks of individuals in which some epidemic might break out, etc. Below we give two important problems that looks quite different, however, they might be answered in the same way.

- Researchers measured the neural activity between the different areas of the macaque brain. The measurement can be described with a directed graph, $G(V,E)$, where the vertices are the different areas of the macaque brain, and an edge is going from *u* to *v* if neurons are going from the area represented by *u* to the area represented by *v*. They found that there are some main processing centers, which are areas with many incoming neurons, from where outgoing neurons go to other areas that have many outgoing neurons. They can define a function quantifying the pattern in this way:

  $$R(G) = \sum_{(u,v)\in E} d_u^{in} d_v^{out} \tag{15.1}$$

  where d_u^{in} and d_v^{out} represents the incoming degree of *u* and outgoing degree of *v*, respectively. It is easy to count this number, but what this value means? How can it be decided if it is a large value or a low value? We should compare it with values coming from random networks. Obviously, the value depends on the incoming and outgoing degrees, so we would like to generate random networks with prescribed incoming and outgoing degrees. Namely, we would like to generate random macaque brains, in which the different areas have the same amount of incoming and outgoing neurons than in the real macaque brain, but otherwise the areas are randomly connected. If the majority (or all) of these networks have a smaller value than we get from the experiment, we can conclude that the macaque brain is far from randomness, and the observed pattern did not emerge by chance for in random networks we rarely see such high values.
- The Vanuatu islands are famous for its very colorful and diverse bird fauna. Ecologists monitored the bird fauna, and they summarized it with a so-called presence/absence matrix. The rows of the matrix represent the species and the columns represent the islands. If a species can be found at an island, it is denoted by writing a 1 into the matrix, otherwise we write a 0. If a species A lives at place X but not at a place Y, on the other hand, species B lives at place Y but not at place X, then species A and B are suspicious to be competitors. It is only suspicious: they can avoid each other also by chance. We can count the number of so-called checkerboard units in this matrix, namely, two, not necessarily consecutive rows and columns with $\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}$ or $\begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix}$ pattern, but again, the question emerges: is it a low or a high value? Namely, how much competition can be found in the Vanuatu bird fauna? We would like to compare the number of checkerboard units in the Vanuatu presence/absence matrix with that in some random matrix. However, we would like to generate random matrices with the same row and column sums, since the number of checkerboard units depends on it. Namely, we want to generate random presence/absence matrices in which the species are such widespread than in the Vanuatu fauna, and the places are as rich in species as on the Vanuatu islands, but otherwise the species are randomly distributed. If the number of checkerboard units is typically smaller in the random matrices than in the Vanuatu matrix, then we can support the hypothesis that there is significant competition of birds on the Vanuatu islands.

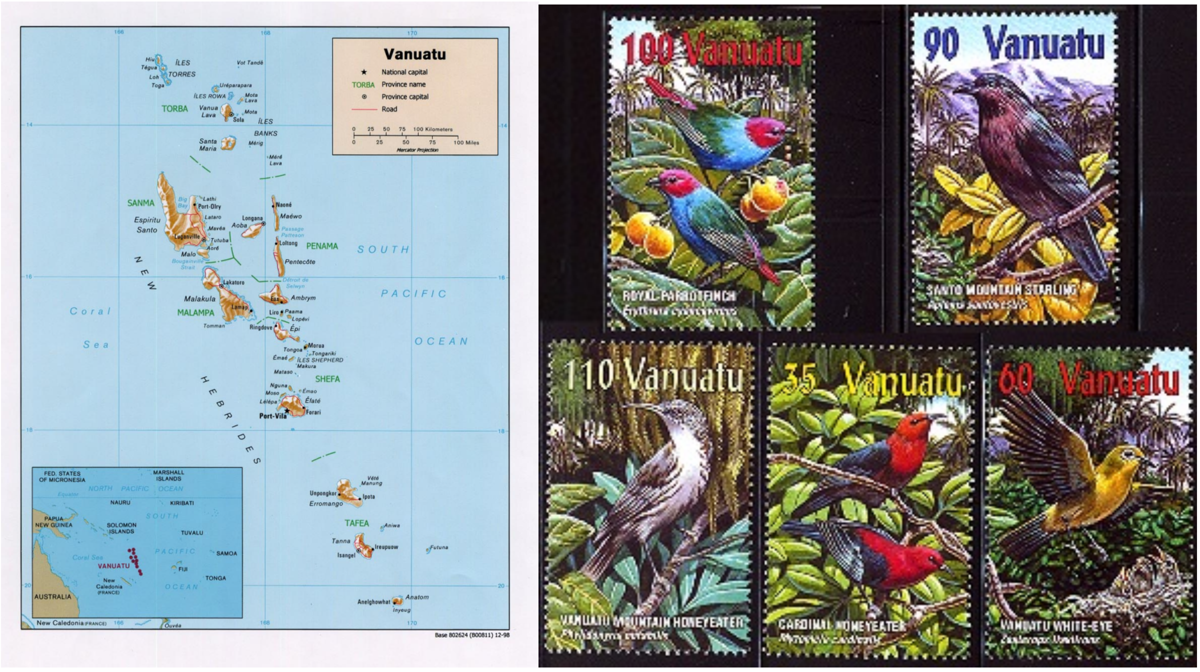


Figure 15.1. The Vanuatu islands in the Pacific Ocean and some birds from Vanuatu pictured on postal stamps.

Although the two problems seem to be far from each other, they are quite similar. In the first case, we want to generate directed graphs with prescribed in and out degrees. In the second case, we want to generate 0-1 matrices with prescribed row and column sums. However, any 0-1 matrix can be viewed as the adjacency matrix of a bipartite graph, namely, generating a matrix with prescribed row and column sums is equivalent with generating a bipartite graph with prescribed degrees. Below we first give an algorithm how to decide if a graph with prescribed degrees exists and how to construct one of them. After this, we introduce the state-of-the-art of uniform generation of graphs with prescribed degree sequences.

15.1. The Havel-Hakimi theorem

Definition A *degree sequence* is a sequence of positive integers $d_1 \geq d_2 \geq \ldots \geq d_n$. A degree sequence is *graphical* if a simple graph exists whose degrees are exactly the degree sequence. For such a graph, we say that the graph is a *realization* of the degree sequence.

Theorem 15.1. (Havel-Hakimi) A degree sequence $d_1 \geq d_2 \geq \ldots \geq d_n$ is graphical if and only if the degree sequence $d_2 - 1, d_3 - 1, \ldots d_{d_1+1} - 1, d_{d_1+2}, \ldots, d_n$ (with some possible reordering) is graphical.

Proof: The backward direction is trivial: if $d_2 - 1, d_3 - 1, \ldots d_{d_1+1} - 1, d_{d_1+2}, \ldots, d_n$ is graphical, take a realization of it, and extend it with one vertex, call it *v*, and *v* should be connected with the first d_1 vertices. Then we get a graph whose degrees are $d_1 \geq d_2 \geq \ldots \geq d_n$, thus this degree sequence is also graphical.

Proving the forward direction is done in an iterative way. Let the vertices be indexed by their degree indices, namely, v_i is the vertex with degree d_i. We show if $d_1 \geq d_2 \geq \ldots \geq d_n$ is graphical then such a realization also exists in which the vertex v_1 is connected with the vertices $v_2, v_3, \ldots v_{d_1+1}$. Assume that in a realization of $d_1 \geq d_2 \geq \ldots \geq d_n$, there is an index *i* such that v_1 is not connected to v_i, although $i \leq d_1 + 1$. Let *i* be the smallest such index. Then there must be an index *j* such that $j > i$, and v_1 is connected to v_j. We know that $d_i \geq d_j$, therefore amongst the neighbor of v_i, there must be a vertex which is not a neighbor of v_j. Let this vertex be v_k. Then edges (v_1,v_j) and (v_i,v_k) exist in the realization, and (v_1,v_i) and (v_i,v_k) do not exist. If we delete the before mentioned existing edges and add the not existing edges, we get a realization of $d_1 \geq d_2 \geq \ldots \geq d_n$ in which v_1 is connected to v_i, thus the first index i' for which v_1 is not connected to $v_{i'}$ is greater than i. We can repeat this alteration such that eventually v_1 is connected to $v_2, v_3, \ldots v_{d_1+1}$. Then deleting v_1 and its edges leads to a realization of $d_2 - 1, d_3 - 1, \ldots d_{d_1+1} - 1, d_{d_1+2}, \ldots, d_n$.

□

The proof is constructive, namely, it is also possible to construct a realization if such exists by following the proof: take n vertices, index it with v_1, v_2 ... v_n. Connect v_1 to $v_2, v_3, \ldots v_{d_1+1}$. Then take the sequence $d_2 - 1, d_3 - 1, \ldots d_{d_1+1} - 1, d_{d_1+2}, \ldots, d_n$, reorder it, moving the vertices together with the degrees, so we get another degree sequence $d'_1 \geq d'_2 \geq \ldots \geq d'_{n-1}$. Take the corresponding v'_1, connect it to the next d'_1 vertices, modify the degrees accordingly, rearrange them, etc. In this way, either we construct a graph with the prescribed sequence or at some point, d_1 will be greater than the number of remaining vertices, and thus, the degree sequence is not graphical.

Similar theorem is true for bipartite graphs and it is left as an exercise.

Similar theorem exists for directed graphs. First we need the definition of bi-degree sequences.

Definition A sequence of non-negative integer pairs $\left(d_1^{in}, d_1^{out}\right), \left(d_2^{in}, d_2^{out}\right), \ldots \left(d_n^{in}, d_n^{out}\right)$ is called *bi-degree sequence*. Such a sequence is called *graphical* if a simple, directed graph exists whose in and out degrees are the given pairs.

Theorem 15.2. (Havel-Hakimi for directed graphs) Let $\left(d_1^{in}, d_1^{out}\right), \left(d_2^{in}, d_2^{out}\right), \ldots \left(d_n^{in}, d_n^{out}\right)$ be a bi-degree sequence. Take any pair $\left(d_i^{in}, d_i^{out}\right)$ such that $d_i^{out} > 0$ and rearrange the remaining pairs into lexicographically decreasing order $\left(d'^{in}_1, d'^{out}_1\right), \left(d'^{in}_2, d'^{out}_2\right), \ldots \left(d'^{in}_{n-1}, d'^{out}_{n-1}\right)$, namely, for each $1 \leq i < n-1$, $d_i^{in} \geq d_{i+1}^{in}$ and $d_i^{out} \geq d_{i+1}^{out}$ if $d_i^{in} = d_{i+1}^{in}$. Then $\left(d_1^{in}, d_1^{out}\right), \left(d_2^{in}, d_2^{out}\right), \ldots \left(d_n^{in}, d_n^{out}\right)$ is graphical if and only if

$$\left(d_i^{in}, 0\right)\left(d'^{in}_1 - 1, d'^{out}_1\right), \left(d'^{in}_2 - 1, d'^{out}_2\right), \ldots \left(d'^{in}_{d_i^{out}} - 1, d'^{out}_{d_i^{out}}\right), \left(d'^{in}_{d_i^{out}+1}, d'^{out}_{d_i^{out}+1}\right) \ldots \left(d'^{in}_{n-1}, d'^{out}_{n-1}\right) \qquad (15.2)$$

is also graphical.

Proof: Again, the backward direction is trivial: if the degree sequence in (15.2) is graphical, then take a realization of it, take the vertex with degree $\left(d_i^{in}, 0\right)$, and connect it with the first d_i^{out} vertices. Then we get a realization of $\left(d_1^{in}, d_1^{out}\right), \left(d_2^{in}, d_2^{out}\right), \ldots \left(d_n^{in}, d_n^{out}\right)$.

The forward way is also proved in an analogous way to the proof of Theorem 15.1. We prove if a realization exists for the bi-degree sequence $\left(d_1^{in}, d_1^{out}\right), \left(d_2^{in}, d_2^{out}\right), \ldots \left(d_n^{in}, d_n^{out}\right)$ then also a realization exists in which the outgoing edges of v_i are going to $v'_1, v'_2 \ldots v'_{d_i^{out}}$. Assume that this is not the case, then take the smallest index j such that v_i does not have an outgoing edge towards v'_j. Then there exists a $k > j$ such that v_i does have an outgoing edge towards v'_k. Since $d_j^{in} \geq d_k^{in}$ there must be a vertex v'_l such that there is an edge going from v'_l to v'_j but not to v_k. If l is not k, then we can delete edges (v_i, v'_k,) and (v'_l, v'_j,) and add edges (v_i, v'_j,) and (v'_l, v'_k,). If l is k but $d_j^{in} > d_k^{in}$ or there is an edge going from v'_j to v'_k, then there still is another l which is not k and there is an edge going from v'_l to v'_j but not to v_k. If l is k, $d_j^{in} = d_k^{in}$ and there is no edge going from v'_j to v'_k, then we can use the fact that $d_j^{out} \geq d_k^{out}$ since the degree pairs are in lexicographically decreasing order, and we must be able to find a vertex v'_m such that there is an edge going from v'_j to v'_m, but there is no edge from v'_k to v'_m. Then we can delete edges (v_i, v'_k), (v'_j, v'_m) and (v'_k, v'_j) and add edges (v_i, v'_j), (v'_j, v'_k) and (v'_k, v'_m) without changing the bi-degree sequence. Thus, the smallest index j' for which no edge goint from v_i to $v'_{j'}$ will be greater than j, and eventually, the outgoing edges from v_i will go to $v'_1, v'_2 \ldots v'_{d_i^{out}}$. Then we can remove these vertices to get a realization of the bi-degree sequence in Equation 15.2.

□

15.2. The swap Markov chain

Definition: A swap in a graph $G(V,E)$ takes four vertices *a, b, c, d*, for which $(a,b) \in E, (c,d) \in E$ and $(a,d) \notin E, (c,b) \notin E$ and changes the edge set such that the new edge set will be $E \setminus \{(a,b),(c,d)\} \cup \{(a,d),(b,c)\}$. If the graph is a bipartite graph, then it is required that *a* and *c* be in one of the vertex set, and *c* and *d* be in the other vertex set. If the graph is directed then the edges must be directed in an order as indicated here (namely, the edge is going from *a* to *b*, etc.)

It is obvious that a swap do not change the degree sequence, and in case of directed graphs, it does not change the bi-degree sequence. A swap on a bipartite graph is equivalent with changing a $\begin{matrix} 1 & 0 \\ 0 & 1 \end{matrix}$ checkerboard unit to a $\begin{matrix} 0 & 1 \\ 1 & 0 \end{matrix}$ checkerboard unit or vice versa.

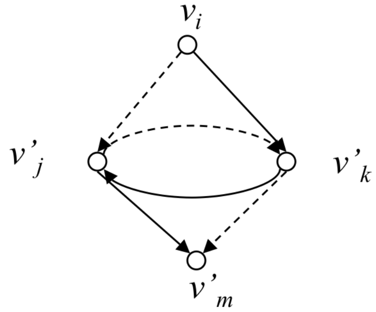
Theorem 15.3. Let *G* and *H* be two graphs realizing the same degree sequence. Then there is a finite series of swaps that transforms *G* into *H*.
Proof: From the proof of Theorem 15.1, it follows that both *G* and *H* can be transformed into the Havel-Hakimi realization. The inverse of a swap is also a swap, so *G* can be transformed into *H* such that it first transformed into the Havel-Hakimi realization, then the Havel-Hakimi realization is transformed back to *H*.

□

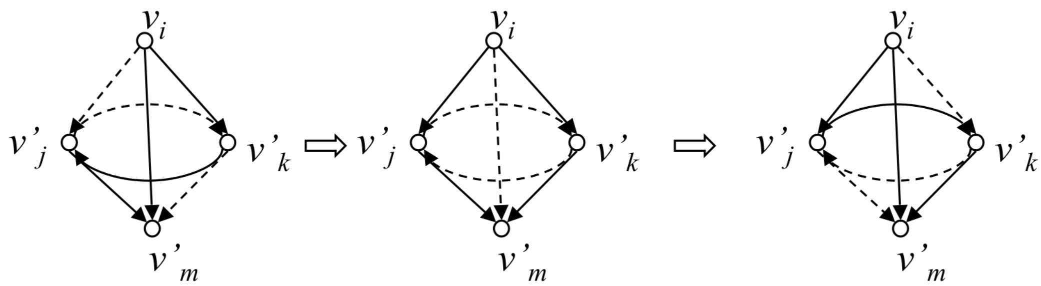
Definition: A triangular C_3 swap takes 3 vertices, *a, b* and *c* from a directed graph $\vec{G}(V,E)$ such that $(a,b) \in E, (b,c) \in E, (c,a) \in E$ and $(a,c) \notin E, (b,a) \notin E, (c,b) \notin E$, then it removes the existing edges and adds the non-existing edges.

Again, it is obvious that a triangular C_3 swap does not change the bi-degree sequence.

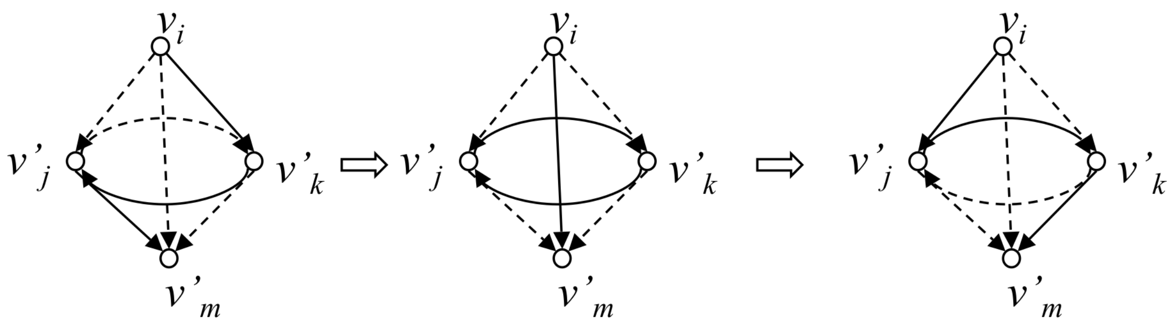
Theorem 15.4. Let $\vec{G}$ and $\vec{H}$ be two directed graphs, both of them realizing the same bi-degree sequence. Then there is a finite series of swaps and triangular C_3 swaps that transform $\vec{G}$ into $\vec{H}$.
Proof: From the proof of Theorem 15.2, it follows that both $\vec{G}$ and $\vec{H}$ can be transformed into the Havel-Hakimi realization using swaps and alterations that affect at most 4 vertices. If v_i equals to v'_m then it is a triangular C_3 swap, otherwise the case can be pictured in the following way:



Now if there is an edge going from v_i to v'_m, then there is a swap removing edges (v_i, v'_m) and (v'_k, v'_j) and adding edges (v_i, v'_j) and (v'_k, v'_m), then after this swap, another swap is available removing edges (v_i, v'_k) and (v'_j, v'_m) and adding edges (v_i, v'_m) and (v'_j, v'_k). The following picture shows these two steps:



The effect of the two swaps is the same than the alteration in the proof of the Havel-Hakimi theorem for directed graphs. Finally, if there is no edge going from v_i to v'_m, then there is a swap removing edges (v_i, v'_k) and (v'_j, v'_m) and adding edges (v_i, v'_m) and (v'_j, v'_k), then after this swap, another swap is available removing edges (v_i, v'_m) and (v'_k, v'_j) and adding edges (v_i, v'_j) and (v'_k, v'_m). The following picture shows these two steps:



Again, the effect of the two swaps is the same than the alteration in the proof of the Havel-Hakimi theorem for directed graphs. In this way, we can transform $\vec{G}$ into the Havel-Hakimi realization with swaps and triangular C_3 swaps, then the Havel-Hakimi realization can be transformed back to $\vec{H}$ with swaps and triangular C_3 swaps since the inverse of a triangular C_3 swap is also a triangular C_3 swap. □

The swaps, and in case of directed graphs, the swaps and triangular C_3 swaps are the basis of a so-called Markov chain Monte Carlo algorithm, that sample from the (almost) uniform distribution of the realizations of degree and bi-degree sequences. A Markov chain is a random walk, and the swap Markov chain is a random walk that walks on the realizations of degree and bi-degree sequences. In each step, a random swap (or triangular C_3 swap) is taken and applied on the current realization to get a new realization as the next step in the random walk. With some mild conditions on how to chose randomly the next swap, it is possible to achieve that the Markov chain converge to the uniform distribution of all realizations. This means that after sufficiently many number of steps, the walk will be in a random realization that is very close to the uniform distribution. The key point in this approach is that the walk can reach any realization from any other realization, and essentially, this is what Theorems 15.3 and 15.4 state.

The central and still open question is how fast the convergence of the Markov chain, namely, in practice, how many steps are necessary to get close to the uniform distribution. It is a generally accepted conjecture that the necessary number of steps grows only polynomial with the length of the degree (or bi-degree) sequence, but it is proved only for some special cases, when the degree sequence is regular or the bi-degree sequence is half-regular, it is when the in-degrees are the same, and the out degrees are arbitrary or the out-degrees are the same and the in-degrees are arbitrary.

Exercises

Exercise 15.1. Prove that the function in Equation 15.1 is the number of directed 3 long paths in the directed graph.

Exercise 15.2. Let *G* and *H* be two bipartite graphs with the same degree sequence. Show that the adjacency matrices of *G* and *H* both contain at least one checkerboard unit.

Exercise 15.3. State and prove the Havel-Hakimi theorem for bipartite graphs.

Exercise 15.4. Give a realization of the degree sequence 5, 5, 4, 4, 4, 4, 1, 1, 1, 1.

Exercise 15.5.* Which are the 0/1 matrices that do not contain any checkerboard unit?

Exercise 15.6.* Give an example that the triangular C_3 swaps are necessary to transform a directed graph into another one.

Exercise 15.7.* Show that in the Havel-Hakimi algorithm an arbitrary vertex can be chosen which is connected to the maximal degree vertices. In each step, we can chose such arbitrary vertex, and thus, we can get several realizations. On the other hand, show that not all realizations of a degree sequence can be constructed in this way.

Exercise 15.8** Prove that in case of regular bi-degree sequences, swaps are sufficient to transform any realization into any another realization.