Degree sequence problems of partite, uniform hypergraphs

Research proposal, 2023 Summer

May 15, 2023

1 Problem description

A hypergraph $H = (V, E)$ is a generalization of graphs, where $E \subseteq 2^V \setminus \emptyset$, that is, a hyperedge is a non-epmpty subset of the vertices. Clearly, a simple graph is a hypergraph where each hyperedge is a subset of vertices of size 2. A hypergraph is a *k-uniform hypergraph* if each hyperedge is a subset of size k (that is, simple graphs are 2-uniform hypergraphs). A hypergraph $H = (V, E)$ is a *partite, k-uniform hypergraph* if V is a disjoint subset of vertex sets $V_1, V_2, \ldots, V_k$ and $E \subseteq V_1 \times V_2 \times \ldots \times V_k$. Partite, k-uniform hypergraphs are generalizations of bipartite graphs. Indeed, bipartite graphs are partite, 2-uniform hypergraphs.

The first interesting[1] partite, uniform hypergraphs are partite, 3-uniform hypergraphs, where the edges are "triangles", that is $(v_{1,i_1}, v_{2,i_2}, v_{3,i_3})$ triplets from $V_1 \times V_2 \times V_3$, where V_1, V_2 and V_3 are disjoint sets of vertices. Partite, 3-uniform hypergraphs naturally appear in data sience, especially in time series data. For example, $(patient, disease, timepoint)$ triplets can naturally be represented by partite, 3-uniform hypergraphs. Similarly, $(user, tweet\ type, timepoint)$ triplets also form a partite, 3-uniform hypergraphs. These time series can be subject of statistical analysis, and during statistical analysis, we would like to generate random partite, 3-uniform hypergraphs to empirically generate background distributions for hypothesis testing. When generating random hypergraphs, we would like to preserve the degrees of the hypergraph, that is, we would like to preserve for each vertex the number of incident hyperedges (triangles). This statistical approach has been a standard way in network science. For example, random ecological presence/absence matrices (adjacency matrices of $(species, habitat)$ pairs forming a bipartite graph) are generated to statistically test the hypothesis that there is a significant competition in an ecological community [1]. The distribution of local subgraphs in real networks can be compared with random networks, as well [2].

To conclude, the first algorithmic step in the above-mentioned statistical approach is to construct a partite, 3-uniform hypergraph whose degrees are some given degrees. We will

[1]Partite, 2-uniform hypergraphs are the bipartite graphs, and are not considered as truly hypergraphs. Partite, 1-uniform hypergraphs are simply the possible subsets of a set of vertices, and quite trivial in research on hypergraphs.

denote by $d(v)$ the degree of the vertex v, that is, the number of hyperedges incident to v. We can formulate the construction/existance problem formally as the following decision problem.

Problem 1 (Partite-3-uniform-degree-sequence).
INPUT *A 3-partite degree sequence,* $D := (d_{1,1}, d_{1,2}, \ldots, d_{1,n_1}), (d_{2,1}, d_{2,2}, \ldots, d_{2,n_2}), (d_{3,1}, d_{3,2}, \ldots, d_{3,n_3})$.
OUTPUT *"Yes" if there is partite, 3-uniform hypergraph* $H = (V_1 \cup V_2 \cup V_3, E)$ *such that for all* i, j, $d(v_{i,j}) = d_{i,j}$, *and "No" otherwise.*

When the answer is "Yes", such an H hypergraph is called the *realization* of the degree sequence D. If D has a realization, we say that D is a *graphic* degree sequence ("graphical degree sequence" is an obsolate definition).

Surprisingly, the problem whether or not there is a partite, 3-uniform hypergraph with prescribed degrees is already an NP-complete problem. However, there are special degree classes for which it is easy to decide if a partite, 3-uniform hypergraph exists with those given degreees.

The three main goals of the research class are:

1. Discover degree sequence classes for which it is easy to decide whether or not a hypergraph with those degrees exists.
2. Set up Erdős-Gallai type inequalities for such degree classes (see details in the qualifying problems).
3. For a given D, we can consider the set of realizations. On this set, we can define a topology by defining the neighbors of a realization as "small perturbations" of it (see details in the qualifying problems). We are interested in small perturbations that make the solution space connected.

2 Qualifying problems

Please, read Chapter 14 in this electronic notes: `https://www.renyi.hu/~miklosi/AlgorithmsOfBioinformatics.pdf`. Make sure that you understand the Havel-Hakimi theorem (Theorem 14.1) and its corollary, Theorem 14.3. Make sure you understand what Theorem 14.3. says. It says that swaps are sufficient small transformations that make the solution space of a (simple graph) degree sequence D connected.

Please, solve at least the first 5 of the following exercises. Exercises 6, 7 and 8 are hard, that's why they are marked with an asterisk.

1. Prove the Havel-Hakimi theorem for bipartite graphs, that is, prove the following. Let $D := (d_{1,1}, d_{1,2}, \ldots, d_{1,n_1}), (d_{2,1} \geq d_{2,2} \geq \ldots \geq d_{2,n_2})$ be a bipartite degree sequence. Prove that D is graphic if and only if $D' := ((d_{1,2}, d_{1,3}, \ldots, d_{1,n_1}), (d_{2,1} - 1, d_{2,2} - 1, \ldots, d_{2,d_{1,1}} - 1, d_{2,d_{1,1}+1}, \ldots, d_{2,n_2})$ is graphic.

Here the indexing is really horrible, but in narrative it is easy to explain: A bipartite degree sequence D is graphic if and only if the degree sequence is graphic that can be obtained from D by removing $d_{1,1}$ from the first sequence, and subtracting 1 from the $d_{1,1}$ largest degrees of the second sequence.

2. Prove the analogue theorem of Theorem 14.3. for bipartite graphs.

3. Let G_1 and G_2 be two realizations of the same bipartite degree sequence. Prove that both the adjacency matrix of G_1 and the adjacency matrix of G_2 contain at least one checkerboard unit.

4. In an $n \times n$ matrix of 0's and 1's, each row sum is the same, and there exists a column sum that is neither 0 nor n. Prove that the matrix contains a checkerboard unit.

5. The Erdős-Gallai theorem is the following: Let $D := d_1 \geq d_2 \geq \ldots \geq d_n$ be a degree sequence. Then G is graphic if and only if

   (a) $\sum_{i=1}^{n} d_i$ is even and

   (b) for all $k = 1, 2, \ldots n-1$,

$$\sum_{i=1}^{k} d_i \leq k(k-1) + \sum_{j=k+1}^{n} \min\{d_j, k\}.$$

   Prove the $\Rightarrow$ direction, that is, if G is graphic then the conditions necessarily hold.

6. * A degree sequence is k-regular if each degree is k. Prove that a k-regular 3-partite degree sequence on $n+n+n$ vertices has a partite, 3-uniform hypergraph realization if $k \leq n^2$.

7. * Find further degree sequence classes for which it is easy to decide if a partite, uniform hypergraph realization exists.

8. * In this short paper: `https://arxiv.org/pdf/1901.02272.pdf` there is a proof that constructing a 3-uniform hypergraph with a given degree sequence is hard. Extend the proof that the problem remains hard in case of partite, 3-uniform hypergraphs.

References

[1] Miklós, I., Podani, J. (2004) Randomization of presence/absence matrices: comments and new algorithms Ecology, 85:86–92.

[2] Orsini *et al.* (2015) Quantifying randomness in real networks. Nature Communications, 6:8627.