Multiple interaction strategies, parameter estimation, and clustering in networks

BY

Ahmed Elbanna

Ph.D. Thesis Outline

Thesis advisor: Dr. Bolla Marianna

Institute of Mathematics
Budapest University of Technology and Economics
Hungary

2017

Introduction

The thesis consists of three strongly intertwining chapters. Each of them uses optimization techniques, spectral clustering tools, and mixture models to assign scores or estimate parameters assigned to the vertices of large unweighted or weighted graphs, which correspond to networks.

In Chapter 1, the objective of the investigation is an interaction network, which is the collection of so-called agents with pairwise connections described by a weighted or unweighted graph. In the economics literature, the authors assign (usually non-negative) real-valued parameters, so-called strategies to the agents so that their common utility function is maximized. They distinguish between complement and substitute strategies, depending on whether the actions (goals) of neighboring (connected) agents are the same or opposite (the action of the neighbor stimulates or suppresses his/her action). In the classical setup, the main part of the objective function is quadratic, and maximized under linear constraints. In this case, the agents' actions are nonnegative real numbers, however, without exact meaning, the scaling and the actual values of them do not carry too much information for the physical or economic features of them. In fact, they are rather compared with respect to the agents, and in this way, give important information about agent groups that follow similar strategies, and hence about the overall structure of the network from the point of view of the underlying activity towards which the strategies are considered.

It has been proved that these strategies mainly depend on the structure of the underlying graph, the properties of which are captured by its eigenvalues and eigenvectors (if the number of agents is large, it suffices to consider only the significant ones). Therefore, we consider the so-called spectral relaxation of the maximization problem, and the quadratic objective functions will be maximized subject to certain orthogonality constraints, where the solutions can be rescaled, as if they were measured in different units. We investigate general social networks, given in the form of an undirected, weighted graph, and also multiple strategies that are real vectors. The maxima are given in terms of the top or bottom eigenvalues of the transformed edge-weight matrix, whereas the optimal multiple strategies are derived by means of the corresponding eigenvectors. The two extremes, corresponding to strategic complements or substitutes are unified into a multiway clustering problem, where we are looking for groups of agents pursuing similar strategies with respect to the other groups, and in this case, strategies can be assigned to the agents, depending on their group memberships. We also generalize the problem to the case, when the relationship between the agents differ under different strategies.

In the social networks literature, many parametric and nonparametric methods have been proposed for so-called community detection in networks. In the nonparametric scenario, hierarchical or spectral methods were applied to maximize the two- or multiway Newman–Girvan modularity; more generally, spectral clustering tools (**SC**), based on Laplacian or modularity spectra, proved to be feasible to find community, anticommunity, or regular structures in networks. In the parametric setup, certain model parameters are estimated,

usually via maximizing the likelihood function of the graph, i.e., the joint probability of our observations under the model equations. This so-called ML estimation is a promising method of statistical inference, has solid theoretical foundations, and also supports the common-sense goal of accepting parameter values based on which our sample is the most likely.

In Chapter 2, we estimate the parameters of two models. First we introduce a semiparametric block model for graphs, where the within- and between-cluster edge probabilities are not constants within the blocks, but are described by logistic type models, reminiscent of the 50 years old Rasch model and the newly introduced α–β models. Our purpose is to give a partition of the vertices of an observed graph so that the induced subgraphs and bipartite graphs obey these models, where their strongly interlaced parameters give multiscale evaluation of the vertices at the same time. In this way, a profoundly heterogeneous version of the stochastic block model is built via mixtures of the above submodels, while the parameters are estimated with a special EM iteration. In Theorem 1, we state the convergence of the iteration, while in Theorem 2, we formulate an interesting statement about the rank of the matrix taking part in the algorithm.

Second we introduce a directed, weighted random graph model, where the edge-weights are independent and beta-distributed with parameters depending on their endpoints. We show that in this model the row- and column-sums of the transformed edge-weight matrix are sufficient statistics for the parameters by the *Neyman–Fisher factorization theorem*. We use the theory of exponential families to prove that the ML estimate of the parameters exists and is unique. Then we present an algorithm that uses a fix point iteration to find this estimate and prove its convergence by the properties of the digamma function. Lastly we show simulation results and applications on that model.

In Chapter 3, we discuss how graph based matrices are capable to find classification of the graph vertices with small within- and between-cluster discrepancies. The structural eigenvalues together with the corresponding spectral subspaces of the normalized modularity matrix are used to find a block-structure in the graph. We also investigate relations between spectral properties, multiway discrepancies, and degree distribution of generalized random and quasirandom graphs.

In the simulations, as the pure cases, we use the generalized random graphs. We will show that a generalized random graph with a convenient pattern matrix almost surely exhibits a k-partite k-regular graph. We prove that the between-cluster average degrees are highly concentrated on their expectations as the number of vertices tends to infinity, under some balancing conditions on the cluster sizes. We discuss the relation between the spectra and multiway discrepancy in the more general framework of rectangular matrices with nonnegative entries. As special (quadratic, but not symmetric cases), directed graphs are also treated. In Theorem 5, we give an upper estimate for the k-th non-trivial largest singular value of the normalized matrix by means of its k-way discrepancy.

Throughout the thesis, we use linear algebra, probability, and statistical methods. Simulation results and real-life examples are also presented. The chapters of this thesis correspond to four papers. Chapter 1 corresponds to [4], chapter 2 corresponds to [1, 5], and chapter 3 is based on [2].

Chapter 1

Multiple interaction strategies in networks related to graph spectra and dominant sets

An interaction network is a collection of agents with pairwise connections described by a graph. Our objective is to maximize the payoff of the agents simultaneously. In the classical strategic complements or substitutes setup, the objective function has a linear and a quadratic part, and maximized under linear constraints. To address this task, we use quadratic objective functions on linear or quadratic constraints. We will show how existing results of combinatorial graph theory and spectral clustering can be used to solve the optimization problems, where solutions are closely related to dominant sets or spectral clusters. Our primary focus is on the graph and show how certain model parameters can be built into the edge-weight matrix to get a new objective, thus modifying the interactions between the agents.

We mainly consider edge-weighted graphs and extend existing results on strategic interactions to them. In the classical papers on this topic there are unweighted interactions between the agents, and their actions – we are looking for – are nonnegative real numbers. However, without exact meaning, the scaling and the actual values of these actions do not carry too much information for the physical or economic features of them. In fact, they are rather compared with respect to the agents, and in this way, give important information about agent groups that pursue similar strategies, and hence about the overall structure of the network from the point of view of the underlying activity towards which the strategies are considered.

Here we rather investigate the problem from the point of the view of the graph. Based on the spectral properties of a graph based matrix, we are able to tell how many and what kind of strategies can be optimal for the agents, and find agent groups pursuing similar strategies. Since the agents form a social network, the optimal or nearly optimal strategies should inevitably be adopted to the structure of the underlying graph. Together with clustering, we also use evaluation of the vertices and edges, which give optima of potential functions, sometimes related to eigenvectors or weighted characteristic vectors of dominant sets.

We mainly consider quadratic objective function on linear constraints. If we optimize over the standard simplex, we can use the results of Motzkin and Straus [11] to unweighted and those of Pavan and Pelillo [13] to edge-weighted graphs. In this way, unweighted and weighted indicator vectors of dominant sets enter into the solution. In Section 1.3, quadratic constraints are considered, under which our quadratic optimization has an explicit solution based on eigenvalues and eigenvectors of graph based matrices. Here we use multiple strategies and spectral clustering tools of [3].

We also show that the existence of large positive eigenvalues makes rise to a comple-

mentary, whereas that of outstanding negative eigenvalues to a substitute strategy. Some coordinates of the multi-dimensional strategies of some agents can be negative here, but with appropriate rotations the strategy vectors can be substituted by vectors close to weighted indicator vectors of agent groups. Simulation results on generalized random graphs are also presented.

1.1 Basic notions and notation

For a weighted graph $G = (V, \boldsymbol{W})$ with vertex-set $V = \{1, \ldots, n\}$, the weighted adjacency matrix $\boldsymbol{W} = (w_{ij})$ is an $n \times n$ matrix with zero diagonal as there are no self-loops at the moment, and $w_{ij} = w_{ji} \geq 0$, where $w_{ij} = 0$ if vertices i and j are not connected, otherwise it reflects the strength of the connection between vertices i and j. The vertices correspond to the agents, while the weights represent their pairwise similarity or cohesiveness. The simple (unweighted) graph $G = (V, \boldsymbol{A})$ is a special case, where the weights are 0 or 1. In this chapter we deal with undirected graphs, so $\boldsymbol{W}$ and $\boldsymbol{A}$ are symmetric.

Let $d_i = \sum_{j=1}^{n} w_{ij}$ be the *generalized degree* of vertex i; the degrees are sometimes collected in the *degree-vector* $\mathbf{d} = (d_1, \ldots, d_n)^T$ or in the diagonal *degree-matrix* $\boldsymbol{D} = \mathrm{diag}(\mathbf{d})$. In the edge-weighted case we assume that $\sum_{i=1}^{n} d_i = 1$, since the normalized edge-weight matrix, $\boldsymbol{D}^{-1/2}\boldsymbol{W}\boldsymbol{D}^{-1/2}$, is not affected by this normalization. $\boldsymbol{W}$ is a Frobenius type matrix. Denoting its eigenvalues by $\lambda_1 \geq \cdots \geq \lambda_n$, $\sum_{i=1}^{n} \lambda_i = 0$, λ_1 is strictly positive, $d_{min} \leq \lambda_1 \leq d_{max}$, and $\lambda_1 > |\lambda_n|$, where $\lambda_n < 0$.

The *Laplacian matrix* is defined as $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$. The Laplacian matrix is positive semidefinite, and the multiplicity of its 0 eigenvalue is equal to the number of connected components of G.

The *normalized Laplacian matrix* is defined as

$$\boldsymbol{L_D} = \boldsymbol{D}^{-1/2}\boldsymbol{L}\boldsymbol{D}^{-1/2} = \boldsymbol{I}_n - \boldsymbol{D}^{-1/2}\boldsymbol{W}\boldsymbol{D}^{-1/2}.$$

Its spectrum lies in the interval $[0, 2]$, and again, the multiplicity of its 0 eigenvalue is equal to the number of connected components of G.

With the convention $\sum_{i=1}^{n} d_i = 1$, the *modularity matrix* is defined as

$$\boldsymbol{M} = \boldsymbol{W} - \mathbf{d}\mathbf{d}^T.$$

The (i, j) entry of $\boldsymbol{M}$ just measures the deviation of w_{ij} (actual connection of vertices i and j) from $d_i d_j$ (their connection under independent attachment with the vertex-degrees as probabilities).

The *normalized modularity matrix* of G is the following.

$$\boldsymbol{M_D} = \boldsymbol{D}^{-1/2}\boldsymbol{M}\boldsymbol{D}^{-1/2} = \boldsymbol{W}_D - \sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^T = \boldsymbol{D}^{-1/2}\boldsymbol{W}\boldsymbol{D}^{-1/2} - \sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^T \qquad (1.1)$$

Let $\mathbf{u}_1, \ldots, \mathbf{u}_k \in \mathbb{R}^n$ be unit-norm, pairwise orthogonal eigenvectors corresponding to the k largest absolute value eigenvalues of $\boldsymbol{W}$ and let

$$(\mathbf{u}_1, \ldots, \mathbf{u}_k) = \begin{pmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \vdots \\ \mathbf{r}_n^T \end{pmatrix},$$

where $\mathbf{r}_1, \ldots, \mathbf{r}_n \in \mathbb{R}^k$ are k-dimensional vertex representatives. The k-variance of them over the k-partition (clustering) $\mathcal{P}_k = (C_1, \ldots, C_k)$ of the vertices is

$$S_k^2 = \min_{(C_1,\ldots,C_k)} \sum_{u=1}^{k} \sum_{j \in C_u} \|\mathbf{r}_j - \mathbf{c}_u\|^2, \tag{1.2}$$

where $\mathbf{c}_u = \frac{1}{|C_u|} \sum_{j \in C_u} \mathbf{r}_j$ is the center of the cluster C_u. This minimum can be found by the k-means algorithm.

In case of the normalized modularity matrix $\boldsymbol{M}_D$, the representatives $\mathbf{r}_1, \ldots, \mathbf{r}_n \in \mathbb{R}^{k-1}$ are row vectors of the $n \times (k-1)$ matrix of column vectors $\boldsymbol{D}^{-1/2}\mathbf{u}_1, \ldots, \boldsymbol{D}^{-1/2}\mathbf{u}_{k-1}$, where $\mathbf{u}_i$'s are unit-norm, pairwise orthogonal eigenvectors corresponding to the eigenvalues (in decreasing absolute values) of $\boldsymbol{M}_D$. The *weighted k-variance* of these representatives is defined as

$$\tilde{S}_k^2 = \min_{(C_1,\ldots,C_k)} \sum_{u=1}^{k} \sum_{j \in C_u} d_j \|\mathbf{r}_j - \mathbf{c}_u\|^2, \tag{1.3}$$

where $\mathbf{c}_u = \frac{1}{\mathrm{Vol}(C_u)} \sum_{j \in C_u} d_j \mathbf{r}_j$ is the weighted center of cluster C_u. It is the weighted k-means algorithm that gives this minimum.

1.2 Optimizing over the unit simplex

Strategic interaction is a game in which groups of players are looking for a common goal such as profit. This process is happening by using some sort of communication between players, and this communication can be represented by a graph (network), an edge between two players in which represent a direct communication between these two players. The interaction between agents in a game consisting of n agents is described in two main cases; games with strategic complements, in which the payoff (benefit) of an agent increases as a result of the actions of the other agents, and games with strategic substitutes, in which the payoff of an agent decreases due to the actions of the other agents. The main goal is to reach the equilibrium state that no player has benefit to change his action taking in consideration the actions of the other agents.

Motzkin and Straus were the first to consider maximizing $\mathbf{x}^T \boldsymbol{A} \mathbf{x}$ over the unit simplex $S = \{x_i \geq 0 \, (i = 1, \ldots, n), \quad \sum_{i=1}^{n} x_i = 1\} = \{\mathbf{x} \geq \mathbf{0}, \quad \mathbf{x}^T \mathbf{1} = 1\}$, and introduced the solution in terms of maximal and maximum clique and their corresponding characteristic vectors which are defined as follows.

Definition 1 *Let $G = (V, E)$ be an undirected graph. A subset $U \subseteq V$ is said to be a clique if and only if U represents a complete graph: every pair of vertices is connected by an edge.*

A *maximal clique* is a clique that cannot be extended by including one more adjacent vertex, in other words, it is not a subset of a larger clique. A maximal clique C is *strictly maximal* if no vertex i external to C has the property that the enlarged set $C \cup \{i\}$ contains a clique of the size $|C|$. A clique is called *maximum clique* in G if it has maximum cardinality among all cliques in G. The characteristic vector of a vertex-subset $U \subset V$ is denoted $\mathbf{x}^U$ and is defined with the following coordinates: $x_i^U = \frac{1}{|U|}$ if $i \in U$ and 0 otherwise.

Now we show how Pavan and Pelillo generalized the problem from unweighted graphs to weighted ones by generalizing the notion of maximal cliques into *dominant sets*, and the characteristic vectors into weighted characteristic vectors.

1.2.1 Interactions and dominant sets

Let $G = (V, \boldsymbol{W})$ be an edge-weighted graph. We will use the notion of a dominant set as introduced by Pavan and Pelillo [13] as follows. Let $U \subset V$ and $j \notin U$. Let

$$\varphi_U(i,j) = w_{ij} - \frac{1}{|U|}\sum_{l\in U} w_{il}, \quad i \in U$$

be the relative similarity between vertices i and j with respect to the average similarity between vertex i and its neighbors in U. Using this relative similarity, the weight of vertex i with respect to U is defined by the following recursive formula:

$$\mathbf{w}_U(i) = \begin{cases} 1, & \text{if } |U| = 1 \\ \sum_{l\in U\setminus\{i\}} \varphi_{U\setminus\{i\}}(l,i)\mathbf{w}_{U\setminus\{i\}}(l), & \text{otherwise.} \end{cases}$$

The total weight of U is $W(U) = \sum_{i\in U} \mathbf{w}_U(i)$. The function $\mathbf{w}_U(i)$ measures the relative similarity between vertex i and the vertices of $U \setminus \{i\}$ with respect to the overall similarity among the vertices in $U \setminus \{i\}$.

Definition 2 *If $W(T) > 0$ for any nonempty $T \subseteq U$, $U \subseteq V$, then U is a dominant set if*

- $\mathbf{w}_U(i) > 0$, *for all* $i \in U$,
- $\mathbf{w}_{U\cup\{i\}}(i) < 0$, *for all* $i \notin U$.

The first condition ensures that vertices in U are strongly connected to each other, while the second condition ensures that the set U induces the most strongly connected subset in G. This definition shows that in a dominant set, the overall similarity among its vertices is higher than the similarity between its vertices and the rest of the vertices in V. The quadratic programming task

$$\begin{aligned} &\text{maximize} \quad P(\mathbf{x}) = \mathbf{x}^T\boldsymbol{W}\mathbf{x} \\ &\text{subject to} \quad \mathbf{x} \in S, \end{aligned} \tag{1.4}$$

favors pairs of vertices with similar coordinates in $\mathbf{x}$ that also have strong connection in $\boldsymbol{W}$. Pavan and Pelillo [13] characterized the strict local maxima of the above task by means of weighted characteristic vectors.

Definition 3 *The weighted characteristic vector of a set U, also denoted by $\mathbf{x}^U$, has the following coordinates:*

$$x_i^U = \begin{cases} \frac{\mathbf{w}_U(i)}{W(U)}, & \text{if } i \in U \\ 0, & \text{otherwise.} \end{cases}$$

Let us consider the simplest case when the agents have individual costs and mutual benefit based on complementarities between them. The connections between the agents is described by the edge-weighted graph $G = (V, \boldsymbol{W})$. For maximizing the utility over the simplex S, the following quadratic programming task arises:

$$\begin{aligned} &\text{maximize} \quad P(\mathbf{x}) = \frac{1}{2}\beta\mathbf{x}^T\boldsymbol{W}\mathbf{x} - \frac{1}{2}\alpha\mathbf{x}^T\boldsymbol{I}\mathbf{x} = \frac{1}{2}\mathbf{x}^T(\beta\boldsymbol{W} - \alpha\boldsymbol{I})\mathbf{x} \\ &\text{subject to} \quad \mathbf{x} \in S. \end{aligned} \tag{1.5}$$

with positive constants α and β.

Proposition 1 *The programming task (1.5) can be solved by replacing the matrix* $\beta\boldsymbol{W}-\alpha\boldsymbol{I}$ *with* $\beta\boldsymbol{W}-\alpha\boldsymbol{I}+\kappa\mathbf{1}\mathbf{1}^T$*, where* κ *is an arbitrary real number*[1]*. If we put* $\kappa=\alpha$*, then the resulting matrix has nonnegative entries and zero diagonal and -as Pavan and Pelillo proved for that form- the strict local maxima of (1.5) are weighted characteristic vectors of dominant sets for the scaled edge-weight matrix* $\beta\boldsymbol{W}+\alpha(\mathbf{1}\mathbf{1}^T-\boldsymbol{I})$ *having zero diagonal and off-diagonal entries equal to* $\beta w_{ij}+\alpha\geq 0$ $(i\neq j)$*.*

Pavan and Pelillo adopted the following iteration to maximize (1.5) over S. Starting with $\mathbf{x}(0)$,

$$x_i(t+1)=x_i(t)\frac{(\beta\boldsymbol{W}\mathbf{x}(t))_i-\alpha x_i(t)}{\mathbf{x}(t)^T(\beta\boldsymbol{W}-\alpha\boldsymbol{I})\mathbf{x}(t)} \tag{1.6}$$

for $i=1,\ldots,n$ and $t=0,1,2,\ldots$, until convergence.

However, α could basically change the scale that would result in excluding dominant sets under a certain size. When α is large, namely $\alpha>\beta\lambda_{max}(\boldsymbol{W})$, then the regularization term dominates, and the only solution is an $\mathbf{x}$ having all positive coordinates, and hence, being the weighted characteristic vector of the whole V. If α gets smaller, but $\alpha>\beta\lambda_{max}(\boldsymbol{W}_U)$, where $\boldsymbol{W}_U$ is the edge-weight matrix of the induced subgraph of G on the vertex-set $U\subset V$, then there is no maximizing $\mathbf{x}$ with support which is the subset or equal to U. Therefore, if one wants to avoid too small clusters, we select an α according to this rule. Starting with $\alpha=\beta(n-1)\geq\beta\lambda_{max}(\boldsymbol{W})$, we can decrease α one by one to obtain smaller and smaller clusters, which support the weighted characteristic vector of the solution. However, if $\alpha>\beta(m-1)$, then we exclude characteristic vectors of dominant sets with $|U|\leq m$. Nonetheless, if α is very small, the effect of regularization becomes negligible and dominant sets of $G=(V,\beta\boldsymbol{W})$, or equivalently, those of $G=(V,\boldsymbol{W})$ will enter into the solution.

1.3 Optimizing over spheres and ellipsoids

From now on, we consider *multiple strategies*. The k-dimensional strategies of the agents can be thought of as intensities of buying/selling k different stocks or borrowing/lending k different goods (they may have negative coordinates).

Now the quadratic objective function or its multidimensional extension will be maximized with respect to quadratic constraints. Here we have explicit solutions: the maxima are given in terms of the bottom or top eigenvalues of the transformed edge-weight matrix, whereas the optimal multiple strategies are derived by means of the corresponding eigenvectors. The two extremes, corresponding to strategic complements or substitutes are unified into a multiway clustering problem, where we are looking for groups of agents following similar strategies with respect to the other groups, and in this case, strategies can be assigned to the agents, depending on their group memberships.

1.3.1 When there are complementarities between the agents

The k-dimensional strategies $\mathbf{s}_1,\ldots,\mathbf{s}_n\in\mathbb{R}^k$ of the agents are collected as row vectors in the $n\times k$ matrix $\mathbf{X}$. The coordinate $s_{i\ell}$ of $\mathbf{s}_i$ denotes the strategy of agent i towards the ℓ-th subject $(\ell=1,\ldots,k)$. The utility function of agent i is defined by

$$u_i(\mathbf{X})=\sum_{\ell=1}^{k}\alpha s_{i\ell}^2-\frac{1}{2}\sum_{\ell=1}^{k}s_{i\ell}^2+\phi\sum_{\ell=1}^{k}\sum_{j=1}^{n}w_{ij}s_{i\ell}s_{j\ell} \tag{1.7}$$

[1]the solutions of (1.5) remain the same if the matrix $\beta\boldsymbol{W}-\alpha\boldsymbol{I}$ is replaced with $\beta\boldsymbol{W}-\alpha\boldsymbol{I}+\kappa\mathbf{1}\mathbf{1}^T$, where κ is an arbitrary real number. Indeed, $\kappa\mathbf{x}^T\mathbf{1}\mathbf{1}^T\mathbf{x}=\kappa(\mathbf{x}^T\mathbf{1})^2=\kappa$, since $\mathbf{x}^T\mathbf{1}=1$ due to $\mathbf{x}\in S$. In particular, if $\kappa=\alpha$, the resulting matrix has nonnegative entries and zero diagonal.

where α and ϕ are given positive parameters. The first term is the benefit of agent i using strategy x_i, the second is the cost of agent i, and the last one is the utility

The simultaneous maximization of $u_i(\mathbf{X})$'s with respect to $\mathbf{s}_1, \ldots, \mathbf{s}_n$ subject to $\sum_{i=1}^n \mathbf{s}_i\mathbf{s}_i^T = \mathbf{X}^T\mathbf{X} = \boldsymbol{I}_k$ is equivalent to have the following task:

$$\begin{aligned} &\text{maximize} \quad \text{tr}\mathbf{X}^T[(2\alpha-1)\boldsymbol{I} + \phi\boldsymbol{W}]\mathbf{X} \\ &\text{subject to} \quad \mathbf{X}^T\mathbf{X} = \boldsymbol{I}_k. \end{aligned}$$

Irrespective of the definiteness of the matrix in brackets, the maximum is attained by an $\mathbf{X}^*$ which contains pairwise orthogonal, unit-norm eigenvectors, corresponding to the k largest eigenvalues of $(2\alpha-1)\boldsymbol{I} + \phi\boldsymbol{W}$ in its columns, and the maximum is $\sum_{l=1}^k (2\alpha - 1 + \phi\lambda_l)$, where $\lambda_1 \geq \cdots \geq \lambda_n$ are the eigenvalues of $\boldsymbol{W}$, and it is attained by the corresponding eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ as columns of $\mathbf{X}^*$. These may contain negative coordinates, but they can be approximated by stepwise constant vectors of mainly nonnegative coordinates if the following condition is met: the subspace of these partition-vectors is close to the subspace spanned by $\mathbf{u}_1, \ldots, \mathbf{u}_k$. This is the case if there is a gap between λ_k and λ_{k+1}, i.e. $|\lambda_k - \lambda_{k+1}| > |\lambda_i - \lambda_{i+1}|, i = 0, 1, \cdots, k-1$. In this case,it was proved that the squared distance between these two subspaces is the k-variance of the clusters (see [3]), which is the minimum of the objective function of the k-means algorithm. Hence, the clusters of agents following similar strategies are obtained by applying the k-means algorithm to the optimum strategy vectors, row vectors of $\mathbf{X}^*$. Note, that the representatives can as well be rotated so that the column vectors of the matrix $\mathbf{X}^*$ are near to characteristic vectors of the optimizing vertex clusters, giving the same representation, but resulting in near zero or positive strategies. In this way, a k-partition of the vertices is obtained, so that each cluster of the partition is specialized to a particular strategy out of the k ones. Members of the same cluster pursue the same strategy with the same (positive) intensity, and the others do almost nothing.

We introduced further possibilities of generalization for that task, either by making the scalar of collaboration term depend on the strategy, or taking into consideration the vertex degrees in G.

1.3.2 When there are substitutes between the agents

The utility function of agent i is now defined by

$$u_i(\mathbf{X}) = \sum_{\ell=1}^k \alpha s_{i\ell}^2 - \frac{1}{2}\sum_{\ell=1}^k s_{i\ell}^2 - \delta \sum_{\ell=1}^k \sum_{j=1}^n a_{ij}s_{i\ell}s_{j\ell} \tag{1.8}$$

with parameters $0 < \alpha \leq \frac{1}{2}$ and $\delta > 0$ to regulate the effect of substitutes. The simultaneous maximization of $u_i(\mathbf{X})$'s subject to $\sum_{i=1}^n \mathbf{s}_i\mathbf{s}_i^T = \mathbf{X}^T\mathbf{X} = \boldsymbol{I}_k$ is equivalent to maximizing the following potential function under the same constraint:

$$\begin{aligned} P(\mathbf{X}) &= \sum_{i=1}^n u_i(\mathbf{X}) + \frac{\delta}{2}\sum_{i=1}^n \sum_{\ell=1}^k \sum_{j=1}^n a_{ij}s_{i\ell}s_{j\ell} \\ &= -\frac{1}{2}\text{tr}\mathbf{X}^T[(1-2\alpha)\boldsymbol{I} + \delta\boldsymbol{W}]\mathbf{X}. \end{aligned}$$

Its maximum subject to $\mathbf{X}^T\mathbf{X} = \boldsymbol{I}_k$ is attained at the same $\mathbf{X}$ that gives the minimum of

$$\text{tr}\mathbf{X}^T[(1-2\alpha)\boldsymbol{I} + \delta\boldsymbol{W}]\mathbf{X}$$

on the same constraint. Irrespective of the definiteness of the matrix in brackets, the minimum is attained at an $\mathbf{X}^*$ which contains pairwise orthogonal, unit-norm eigenvectors, corresponding to the k smallest eigenvalues of $(1-2\alpha)\boldsymbol{I}+\delta\boldsymbol{W}$ in its columns, and the minimum is $\sum_{\ell=1}^{k}(1-2\alpha+\delta\lambda_{n-\ell+1})$, where $\lambda_1 \geq \cdots \geq \lambda_n$ are the eigenvalues of $\boldsymbol{G}$, and it is attained by the corresponding eigenvectors $\mathbf{u}_n, \ldots, \mathbf{u}_{n-k+1}$ as columns of $\mathbf{X}$. These may contain negative coordinates, but they can be approximated by stepwise constant vectors of nonnegative coordinates. The subspace of these partition-vectors is close to the subspace spanned by $\mathbf{u}_n, \ldots, \mathbf{u}_{n-k+1}$ if there is a gap between λ_{n-k+1} and λ_{n-k}. In this case, the clusters of agents following similar strategies are obtained by applying the k-means algorithm to the optimum strategy vectors, row vectors of the optimum $\mathbf{X}^*$.

Chapter 2

Estimating parameters of a probabilistic heterogeneous block model and a directed weighted graph model with beta-distributed edge-weights

Parameter estimation is a process that involves estimating the values of the parameters of some distribution based on measured or observed empirical data. For such purpose, several methods came to exist to suit different settings such that maximum likelihood, Bayes, moments, etc. Through our study in this chapter we use the method of maximum likelihood (ML) in the context of estimating the parameters of two different models: a probabilistic heterogeneous block model and a directed weighted graph model. For the block model, we use two models for its building blocks. For the subgraphs we use the α-β model after we refer to already proved facts about the existence of the ML estimate, algorithm and its convergence; while for the bipartite subgraphs, we use the β-γ model for which we introduce a novel algorithm and prove its convergence. We use both of these algorithms for our sample graph, and we connect them together in the framework of the EM algorithm. We apply the algorithm to randomly generated and real-world data. In the second part of this chapter we introduce a directed, weighted random graph model, where the edge-weights are independent and beta-distributed with parameters depending on their endpoints and prove that the ML estimate of the parameters exists and is unique. Then we define an algorithm that finds this estimate and prove its convergence. Simulation results and applications are also included in the Thesis.

2.1 Estimating parameters of a probabilistic heterogeneous block model via the EM algorithm

Here we propose a profoundly heterogeneous block model by carrying on the Rasch model developed more than 50 years ago for evaluating psychological tests. We will call it Logistic Block Model (**LBM**). Given the number of clusters and a classification of the vertices, we will use the Rasch model for the bipartite subgraphs, whereas the α-β models for the subgraphs themselves, and process an iteration (inner cycle) to find the ML estimate of their parameters. Then, based on their contributions to the overall likelihood, we find a new classification of the vertices via taking conditional expectation and using the Bayes rule. Eventually, the two steps are alternated, giving the outer cycle of the iteration.

α-β models for undirected random graphs

This random graph model was introduced before, where the degree sequence is a sufficient statistic. We have an unweighted, undirected random graph on n vertices without loops, such that edges between distinct vertices come into existence independently, but not with the same probability. This random graph can uniquely be characterized by its $n \times n$ symmetric adjacency matrix $\boldsymbol{A} = (A_{ij})$ which has zero diagonal and the entries above the main diagonal are independent Bernoulli random variables whose parameters $p_{ij} = \mathbb{P}(A_{ij} = 1)$ obey the following rule.

$$\frac{p_{ij}}{1 - p_{ij}} = \alpha_i \alpha_j \quad (1 \le i < j \le n), \tag{2.1}$$

where the parameters $\alpha_1, \dots, \alpha_n$ are positive reals. In [9], an algorithm was recommended for the parameter estimation in this model, and its convergence to the unique solution of the maximum likelihood equation was proved.

β-γ model for bipartite graphs

This bipartite graph model was applied for psychological and educational measurements, later market research. According to the Rasch model, the entries of an $m \times n$ binary table $\boldsymbol{A}$ are independent Bernoulli random variables, where for the parameter p_{ij} of the entry A_{ij} the following holds:

$$\frac{p_{ij}}{1 - p_{ij}} = b_i g_j \quad (i = 1, \dots, m,\ j = 1, \dots, n) \tag{2.2}$$

where $b_1, \dots, b_m$ and $g_1, \dots, g_n$ are positive real parameters based on the original ones β and γ. This also holds with the transformed parameters

$$b'_i = b_i \kappa \quad \text{and} \quad g'_j = \frac{g_j}{\kappa} \tag{2.3}$$

with some $\kappa > 0$. Therefore, the parameters b_i and g_j are arbitrary within a multiplicative constant. That yields a dependence between the system of likelihood equations, and hence, the solution is unique only for equivalence classes of the parameters.

We define an algorithm that converges to the unique (up to the above equivalence) solution of the maximum likelihood equation.

Theorem 1 *Under certain condition, the following algorithm gives a unique equivalence class of the parameter vectors as the fixed point of the iteration, which therefore provides the ML estimate of the parameters.*

Starting with positive parameter values $b_i^{(0)}$ $(i = 1, \dots, m)$ *and* $g_j^{(0)}$ $(j = 1, \dots, n)$ *and using the observed row- and column-sums, the iteration is as follows:*

$$I. \quad b_i^{(t)} = \frac{r_i}{\sum_{j=1}^{n} \frac{1}{\frac{1}{g_j^{(t-1)}} + b_i^{(t-1)}}}, \quad i = 1, \dots m$$

$$II. \quad g_j^{(t)} = \frac{c_j}{\sum_{i=1}^{m} \frac{1}{\frac{1}{b_i^{(t)}} + g_j^{(t-1)}}}, \quad j = 1, \dots n$$

for $t = 1, 2, \dots,$ *until convergence.*

2.1.1 Parameter estimation in the LBM

In the several clusters case, we are putting the bricks together. The above discussed α-β and β-γ models will be the building blocks of the **LBM** to be introduced. Here the degree sequences are not any more sufficient for the whole graph, only for the building blocks of the subgraphs.

Given $1 \leq k \leq n$, we are looking for k-partition, in other words, clusters $C_1, \ldots, C_k$ of the vertices such that

- different vertices are independently assigned to a cluster C_u with probability π_u ($u = 1, \ldots, k$), where $\sum_{u=1}^{k} \pi_u = 1$;
- given the cluster memberships, vertices $i \in C_u$ and $j \in C_v$ are connected independently, with probability p_{ij} such that

$$\ln \frac{p_{ij}}{1 - p_{ij}} = \beta_{iv} + \beta_{ju}, \tag{2.4}$$

  for any $1 \leq u, v \leq k$ pair.

The parameters are collected in the vector $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_k)$ and the $n \times k$ matrix $\boldsymbol{B}$ of b_{iu}'s ($i \in C_u$, $u = 1, \ldots, k$). The likelihood function is the following mixture:

$$\sum_{1 \leq u, v \leq k} \pi_u \pi_v \prod_{i \in C_u, j \in C_v} p_{ij}^{a_{ij}} (1 - p_{ij})^{(1 - a_{ij})}.$$

Here $\boldsymbol{A} = (a_{ij})$ is the incomplete data specification as the cluster memberships are missing. Therefore, it is straightforward to use the EM algorithm:to complete our data matrix $\boldsymbol{A}$ with latent membership vectors $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_n$ of the vertices that are k-dimensional i.i.d. $Multy(1, \boldsymbol{\pi})$ (multinomially distributed) random vectors. Starting with initial parameter values $\boldsymbol{\pi}^{(0)}$, $\boldsymbol{B}^{(0)}$ and membership vectors $\boldsymbol{\mu}_1^{(0)}, \ldots, \boldsymbol{\mu}_n^{(0)}$, the t-th step of the iteration is the following ($t = 1, 2, \ldots$).

- **E**-step: we calculate the conditional expectation of each $\boldsymbol{\mu}_i$ conditioned on the model parameters and on the other cluster assignments obtained in the previous step, then using them, we relocate the vertices between the clusters.
- **M**-step: we estimate the parameters separately within the blocks. In the within-cluster scenario, we use the parameter estimation of α-β model. In the between-cluster scenario, we use the bipartite graph β-γ model.

By the general theory of the EM algorithm in exponential family, the algorithm converges to a local maximum of the likelihood function. We also proved the following.

Theorem 2 *Let the $n \times n$ symmetric matrix $\boldsymbol{S}$ contain the log-odds satisfying the model equation (2.4) as its entries. Then* $\operatorname{rank}\boldsymbol{S} \leq 2k$.

We applied the algorithm to two sets of randomly generated data, one with uniform parameters and the other with Gaussian distributed ones. Our algorithm converged to the estimated parameters with good fit of $MSE = 1.14634$ and 1.12556 respectively. Then we estimated the parameters of two real world data sets using the algorithm and showed how **LBM** is a better fine tuning of the spectral clustering than the stochastic block-model **SBM** is.

2.2 A random graph model with beta-distributed edge-weights

Let $\boldsymbol{W} = (w_{ij})$ be the $n \times n$ (usually not symmetric) edge-weight matrix of a random directed graph on n vertices: $w_{ii} = 0$ $(i = 1, \ldots, n)$ and $w_{ij} \in [0, 1]$ is the weight of the $i \to j$ edge $(i \neq j)$. We define the two $n \times n$ matrices $\boldsymbol{U} = \boldsymbol{U}(\boldsymbol{W})$ and $\boldsymbol{V} = \boldsymbol{V}(\boldsymbol{W})$ of general entries $u_{ij} = \ln w_{ij}$ and $v_{ij} = \ln(1 - w_{ij})$ respectively. Our model is the following: the $i \neq j$ weight obeys a beta-distribution with parameters $a_i > 0$ and $b_j > 0$. The parameters are collected in $\mathbf{a} = (a_1, \ldots a_n)$ and $\mathbf{b} = (b_1, \ldots b_n)$, or briefly, in $\boldsymbol{\theta} = (\mathbf{a}, \mathbf{b})$. Here a_i can be thought of as the potential of the vertex i to send messages out, and b_i is its resistance to receive messages in.

The likelihood function is

$$L_{\mathbf{a},\mathbf{b}}(\boldsymbol{W}) = \prod_{i \neq j} \frac{\Gamma(a_i + b_j)}{\Gamma(a_i)\Gamma(b_j)} w_{ij}^{a_i - 1} (1 - w_{ij})^{b_j - 1}.$$

After factorizing it, and by the Neyman–Fisher factorization theorem, the row-sums $R_1, \ldots, R_n$ of $\boldsymbol{U}$ and column-sums $C_1, \ldots, C_n$ of $\boldsymbol{V}$ are sufficient statistics for the parameters. Moreover, $\mathbf{t} = (\boldsymbol{R}, \mathbf{C}) = (R_1, \ldots, R_n, C_1, \ldots, C_n)$ is the canonical sufficient statistic, which is also minimal.

The system of likelihood equations is obtained by making the derivatives of $L_{\mathbf{a},\mathbf{b}}(\boldsymbol{W})$ with respect to the parameters equal to 0 then put it in the form $\boldsymbol{\theta} = f(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\mathbf{a}, \mathbf{b})$ which can be used as a fixed point iteration. Then we have:

$$\begin{aligned} a_i &= \psi^{-1}\left[\frac{1}{n-1} R_i + \frac{1}{n-1} \sum_{j \neq i} \psi(a_i + b_j)\right] =: g_i(\mathbf{a}, \mathbf{b}), \quad i = 1, \ldots, n \\ b_j &= \psi^{-1}\left[\frac{1}{n-1} C_j + \frac{1}{n-1} \sum_{i \neq j} \psi(a_i + b_j)\right] =: h_j(\mathbf{a}, \mathbf{b}), \quad j = 1, \ldots, n, \end{aligned} \tag{2.5}$$

where $\psi(x) = \frac{\partial \ln \Gamma(x)}{\partial x} = \frac{\Gamma'(x)}{\Gamma(x)}$ for $x > 0$ is the *digamma function*. Here g_i's and h_j's are the coordinate functions of $f = (g, h) : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$. Starting at $\boldsymbol{\theta}^{(0)}$, we use the successive approximation $\boldsymbol{\theta}^{(it)} := f(\boldsymbol{\theta}^{(it-1)}$ for $it = 1, 2, \ldots$, until convergence. Now the the statement of convergence of the above iteration to the theoretically guaranteed unique $\hat{\boldsymbol{\theta}}$ follows.

Theorem 3 *Let $\hat{\boldsymbol{\theta}} = (\hat{\mathbf{a}}, \hat{\mathbf{b}})$ be the unique solution of the ML equation. Then the above mapping $f = (g, h)$ is a contraction in some closed neighborhood K of $\hat{\boldsymbol{\theta}}$, and so, starting at any $\boldsymbol{\theta}^{(0)} \in K$, the fixed point of the iteration $\boldsymbol{\theta}^{(it)} = f(\boldsymbol{\theta}^{(it-1)})$ exists and is $\hat{\boldsymbol{\theta}}$.*

We introduce the following statements to ensure a good starting in the iteration.

Proposition 2 *Let*

$$M := \max\left\{ \max_{i \in \{1,\ldots,n\}} \left(-\frac{R_i}{n-1}\right), \max_{i \in \{1,\ldots,n\}} \left(-\frac{C_i}{n-1}\right) \right\} \tag{2.6}$$

and $\varepsilon > 0$ be the (only) solution of the equation $\psi(2x) - \psi(x) = M$. Then $(\hat{\mathbf{a}}, \hat{\mathbf{b}}) \geq \varepsilon \mathbf{1}$.

Proposition 3 *With the solution ε of $\psi(2x) - \psi(x) = M$ of (2.6), we have $f(\varepsilon \mathbf{1}) \geq \varepsilon \mathbf{1}$.*

It is also clear that we have the following.

Proposition 4 *If* $(\mathbf{a}, \mathbf{b}) \geq (\mathbf{x}, \mathbf{y}) > \mathbf{0}$*, then* $f(\mathbf{a}, \mathbf{b}) \geq f(\mathbf{x}, \mathbf{y})$.

Theorem 4 *With* ε *satisfying* $\psi(2\varepsilon) - \psi(\varepsilon) = M$ *of (2.6), and starting at* $\boldsymbol{\theta}^{(0)} = \varepsilon\mathbf{1}$*, the sequence* $\boldsymbol{\theta}^{(it)}$ *of the iteration* $\boldsymbol{\theta}^{(it)} = f(\boldsymbol{\theta}^{(it-1)})$ *for* $it \to \infty$ *converges at a geometric rate to the unique solution* $(\hat{\mathbf{a}}, \hat{\mathbf{b}})$ *of the ML equation.*

We applied the algorithm to randomly generated and migration data between 34 countries, see [5].

Chapter 3

Matrix and discrepancy view of generalized random and quasirandom graphs

Here we conjecture equivalence of generalized quasirandom properties based on former results in the literature and partly make up for the missing chains in the implications. We also give a short proof for the Expander Mixing Lemma for irregular graphs, and estimate the k-th largest singular value of normalized contingency table with its k-way discrepancy. We also extend the notion of multiway discrepancy to rectangular arrays, of which undirected or directed, unweighted or weighted graphs are special cases. The results are supported by computer simulations and processing migration data on the directed graph of which the spectral relaxation technique is illustrated.

The notion of multiway discrepancy

Let $\boldsymbol{A} = (a_{ij})$ be an $m \times n$ matrix with $a_{ij} \geq 0$. We assume that $\boldsymbol{A}\boldsymbol{A}^T$ (when $m \leq n$) or $\boldsymbol{A}^T\boldsymbol{A}$ (when $m > n$) is irreducible. Consequently, the row-sums $d_{row,i} = \sum_{j=1}^{n} a_{ij}$ and column-sums $d_{col,j} = \sum_{i=1}^{m} a_{ij}$ of $\boldsymbol{A}$ are strictly positive, and the diagonal matrices $\boldsymbol{D}_{row} = \mathrm{diag}(d_{row,1}, \ldots, d_{row,m})$ and $\boldsymbol{D}_{col} = \mathrm{diag}(d_{col,1}, \ldots, d_{col,n})$ are invertible. Without loss of generality, we mostly assume that $\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij} = 1$, since neither the discrepancy to be introduced, nor the normalized table

$$\boldsymbol{A}_D = \boldsymbol{D}_{row}^{-1/2}\boldsymbol{A}\boldsymbol{D}_{col}^{-1/2}, \tag{3.1}$$

are affected by the scaling of the entries of $\boldsymbol{A}$. It is well known that the singular values of $\boldsymbol{A}_D$ are in the [0,1] interval.

Definition 4 *The multiway discrepancy of the rectangular array $\boldsymbol{A}$ of nonnegative entries in the proper k-partition $R_1, \ldots, R_k$ of its rows and $C_1, \ldots, C_k$ of its columns is*

$$\mathrm{disc}(\boldsymbol{A}; R_1, \ldots, R_k, C_1, \ldots, C_k) = \max_{1 \leq u,v \leq k} \max_{X \subset R_u,\, Y \subset C_v} \mathrm{disc}(X, Y; R_u, C_v), \tag{3.2}$$

where

$$\begin{aligned} \mathrm{disc}(X, Y; R_u, C_v) &= \frac{|a(X,Y) - \rho(R_u, C_v)\mathrm{Vol}(X)\mathrm{Vol}(Y)|}{\sqrt{\mathrm{Vol}(X)\mathrm{Vol}(Y)}} \\ &= |\rho(X,Y) - \rho(R_u, C_v)|\sqrt{\mathrm{Vol}(X)\mathrm{Vol}(Y)}. \end{aligned} \tag{3.3}$$

Here $a(X,Y) = \sum_{i \in X}\sum_{j \in Y} a_{ij}$ is the cut between $X \subset R_a$ and $Y \subset C_b$, $\mathrm{Vol}(X) = \sum_{i \in X} d_{row,i}$ is the volume of the row-subset X, $\mathrm{Vol}(Y) = \sum_{j \in Y} d_{col,j}$ is the volume of

the column-subset Y, whereas $\rho(X,Y) = \frac{a(X,Y)}{\mathrm{Vol}(X)\mathrm{Vol}(Y)}$ denotes the density between X and Y. The minimum k-way discrepancy of $\boldsymbol{A}$ is

$$\mathrm{disc}_k(\boldsymbol{A}) = \min_{\substack{R_1,\ldots,R_k \\ C_1,\ldots,C_k}} \mathrm{disc}(\boldsymbol{A}; R_1, \ldots, R_k, C_1, \ldots, C_k).$$

This definition extends naturally to quadratic and symmetric matrices.

3.1 Generalized random and quasirandom graphs

Definition 5 *Let n be a natural number and $k \leq n$ be a positive integer. The graph $G_n(\boldsymbol{P}, \mathcal{P}_k)$ is a generalized random graph with probability matrix $\boldsymbol{P}$ and proper k-partition $\mathcal{P}_k = (V_1, \ldots, V_k)$ of the vertices if it satisfies the following. The vertex set is V, $|V| = n$; the $k \times k$ symmetric matrix $\boldsymbol{P}$ is such that its entries satisfy $0 \leq p_{ij} \leq 1$ $(1 \leq i \leq j \leq k)$. Then vertices of V_i and V_j are connected independently, with probability p_{ij}, $1 \leq i \leq j \leq k$.*

In [1, 2, 3] the following properties of a generalized random graph was proved where we will introduce the proof of the fourth one.

Proposition 5 *Let $G_n(\boldsymbol{P}, \mathcal{P}_k)$ be a generalized random graph on n vertices with vertex-classes $\mathcal{P}_k = (C_1, \ldots, C_k)$ of sizes $n_1, \ldots n_k$ and $k \times k$ symmetric probability matrix $\boldsymbol{P}$. Let k be a fixed positive integer and $n \to \infty$ in such a way that $\frac{n_u}{n} \geq c$ $(u = 1, \ldots, k)$ with some constant $0 < c \leq \frac{1}{k}$ (called balancing condition). Then the following hold almost surely for the adjacency matrix $\boldsymbol{A}_n$ and the normalized modularity matrix $\boldsymbol{M}_{D,n}$ of $G_n(\boldsymbol{P}, \mathcal{P}_k)$.*

1. *$\boldsymbol{A}_n$ has k so-called structural eigenvalues that are $\Theta(n)$, while the remaining eigenvalues are $O(\sqrt{n})$ in absolute value. Further, the k-variance $S^2_{k,n}$ of the k-dimensional vertex representatives, based on the eigenvectors corresponding to the structural eigenvalues of $\boldsymbol{A}_n$ (see (1.2)), is $O(\frac{1}{n})$.*

2. *There exists a positive constant $0 < \delta < 1$ independent of n (it only depends on k) such that $\boldsymbol{M}_{D,n}$ has exactly $k-1$ structural eigenvalues of absolute value greater than δ, while all the other eigenvalues are less than $n^{-\tau}$ in absolute value, for every $0 < \tau < \frac{1}{2}$. Further, the weighted k-variance $\tilde{S}^2_{k,n}$ of the $(k-1)$-dimensional vertex representatives, based on the transformed eigenvectors corresponding to the structural eigenvalues of $\boldsymbol{M}_{D,n}$ (see (1.3)), is $O(n^{-\tau})$.*

3. *There is a constant $0 < \theta < 1$ (independent of n) such that $\mathrm{disc}_1(G_n(\boldsymbol{P}, \mathcal{P}_k)) > \theta$, $\ldots$, $\mathrm{disc}_{k-1}(G_n(\boldsymbol{P}, \mathcal{P}_k)) > \theta$, and the k-way discrepancy $\mathrm{disc}_k(G_n(\boldsymbol{P}, \mathcal{P}_k); C_1, \ldots, C_k)$ is $O(n^{-\tau})$.*

4. *(a) For every $1 \leq u \leq v \leq k$ and $i \in C_u$:*

$$\sum_{j \in C_v} a_{ij} = p_{uv} n_v + o(n).$$

(b) For every $1 \leq u \leq v \leq k$ and $i, j \in C_u$:

$$\sum_{t \in C_v} a_{it} a_{jt} = p^2_{uv} n_v + o(n).$$

Now we will discuss similar properties of the generalized quasirandom graphs, which are the deterministic counterparts of the generalized random graphs and are spectrally equivalent to them.

Let us start with the $k = 1$ case. Chung, Graham and Wilson [7] used the term *quasirandom* for simple graphs that satisfy any of some equivalent properties, where these properties are closely related to the properties of expander graphs, including the 'large' spectral gap. Chung and Graham [8] proved that 'small' discrepancy is caused by a 'large' *spectral gap*, which is $1 - \|\boldsymbol{M}_D\|$. This relation is summarized in the following proposition that is a straightforward generalization of the *Expander Mixing Lemma for irregular graphs.*

Proposition 6

$$\mathrm{disc}(G) = \mathrm{disc}_1(G) \le \|\boldsymbol{M}_D\| = |\mu_1|,$$

where $\|\boldsymbol{M}_D\|$ is the spectral norm of the normalized modularity matrix of G.

Though, with different notation, even a stronger version of this proposition is proved in [8], we gave another short proof.

The sequence (G_n) of edge- and possibly vertex-weighted graphs is said to be convergent if the sequence $t(F, G_n)$ of homomorphism densities converges for any simple graph F as $n \to \infty$. The limit object that is a symmetric, bounded, measurable function $W : [0,1]\times[0,1] \to \mathbb{R}$, called *graphon* (see [6]). The stepfunction graphon W_G is assigned to the weighted graph G in the following way: the sides of the unit square are divided into intervals $I_1, \dots, I_n$ of lengths of the relative vertex-weights, and over the rectangle $I_i \times I_j$ the stepfunction takes on the value that is the edge-weight between vertices i and j. The convergence of (G_n) is also equivalent that the stepfunction graphon W_G converges to the limiting graphon in the so-called cut-metric. In terms of the graph convergence, in Section 4 of [3] the following was proved.

Proposition 7 *Consider the generalized random graph sequence $G_n(\boldsymbol{P}, \mathcal{P}_k)$ with $\mathcal{P}_k = (C_1, \dots, C_k)$, $|C_u| = n_u$ $(u = 1, \dots, k)$. Let $n \to \infty$ in such a way that $\frac{n_u}{n} \to r_u$ with some $r_1, \dots, r_k > 0$, $\sum_{u=1}^k r_u = 1$. Then $G_n(\boldsymbol{P}, \mathcal{P}_k) \to W_H$ as $n \to \infty$, where H is a vertex- and edge-weighted graph on k vertices with vertex-weights $r_1, \dots, r_k$, the edge-weights are the entries of $\boldsymbol{P}$, and W_H is the step-function graphon corresponding to H.*

In [10] the following definition of a generalized quasirandom graph sequence was given.

Definition 6 *Given a model graph H on k vertices with vertex-weights $r_1, \dots, r_k$ and edge-weights $p_{uv} = p_{vu}$, $1 \le u \le v \le k$ (entries of $\boldsymbol{P}$), (G_n) is H-quasirandom if $G_n \to W_H$ as $n \to \infty$ in terms of the homomorphism densities.*

In order to construct a generalized quasirandom graph with given k, $\boldsymbol{P}$, and vertex-weights of the model graph H, consider k clusters $C_1, \dots, C_k$ of the vertices of sizes $n_1, \dots, n_k$ such that $\frac{n_u}{n} = r_u$ $(u = 1, \dots, k)$. Let us choose the independent irrational numbers α_{uv} $(1 \le u \le v \le k)$. Then the subgraph on the vertex-set C_u is constructed as follows: $i, j \in C_u$, $i < j$ are connected if and only if

$$\{(i-j)^2\alpha_{uu}\} < p_{uu} \quad (u = 1, \dots, k),$$

where $\{.\}$ denotes the fractional part of a real number. The bipartite subgraph between C_u and C_v is constructed as follows: $i \in C_u$ and $j \in C_v$ are connected if and only if

$$\{(i-j)^2\alpha_{uv}\} < p_{uv} \quad (1 \le u < v \le k).$$

Analytical number theoretical considerations (see, e.g., [5, 14]) guarantee that, for any $1 \leq u \leq v \leq k$, the above fractional parts are well-distributed symmetrically in $[0,1]^2$ if $n \to \infty$ and $\frac{n_u}{n} \to r_u$ $(u = 1, \ldots, k)$.

Conjecture 1 *Consider the sequence of graphs G_n with vertex-set V_n, adjacency matrix $\boldsymbol{A}_n$, and normalized modularity matrix $\boldsymbol{M}_{D,n}$. Let k be a fixed positive integer and $|V_n| = n \to \infty$. Then the following properties are equivalent:*

PI. *(a) $\boldsymbol{A}_n$ has k structural eigenvalues that are $\Theta(n)$ in absolute value, while the remaining eigenvalues are $o(n)$.*

(b) The k-variance $S^2_{k,n}$ of the k-dimensional vertex representatives, based on the eigenvectors corresponding to the structural eigenvalues of $\boldsymbol{A}_n$, is $o(1)$. The k-partition $\mathcal{P}_k = (C_1, \ldots, C_k)$ minimizing this k-variance is such that $\frac{n_u}{n} \geq c$ $(u = 1, \ldots, k)$ holds with some constant c, where $n_u = |C_u|$.

PII. *(a) There exists a constant $0 < \delta < 1$ (independent of n, it only depends on k) such that $\boldsymbol{M}_{D,n}$ has $k-1$ structural eigenvalues that are greater than δ in absolute value, while the remaining eigenvalues are $o(1)$.*

(b) The weighted k-variance $\tilde{S}^2_{k,n}$ of the $(k-1)$-dimensional vertex representatives, based on the transformed eigenvectors corresponding to the structural eigenvalues of $\boldsymbol{M}_{D,n}$, is $o(1)$. The k-partition $\mathcal{P}_k = (C_1, \ldots, C_k)$ minimizing the above weighted k-variance is such that $\frac{n_u}{n} \geq c$ $(u = 1, \ldots, k)$ holds with some constant c, where $n_u = |C_u|$.

PIII. *There are vertex-classes $\mathcal{P}_k = (C_1, \ldots, C_k)$ of sizes $n_1, \ldots n_k$, satisfying $\frac{n_u}{n} \geq c$ $(u = 1, \ldots, k)$ and a constant $0 < \theta < 1$ (independent of n) such that $\mathrm{disc}_1(G_n), \ldots, \mathrm{disc}_{k-1}(G_n) > \theta$, and $\mathrm{disc}_k(G_n; C_1, \ldots, C_k) = o(1)$.*

PIV. *There are vertex-classes $\mathcal{P}_k = (C_1, \ldots, C_k)$ of sizes $n_1, \ldots n_k$, satisfying $\frac{n_u}{n} \geq c$ $(u = 1, \ldots, k)$ and a $k \times k$ symmetric probability matrix $\boldsymbol{P} = (p_{uv})$, such that every vertex of C_u has asymptotically $n_v p_{uv}$ neighbors in C_v for any $1 \leq u \leq v \leq k$ pair. Further, for the codegrees (number of common neighbors) the following holds: every two different vertices $i, j \in C_u$ have asymptotically $p^2_{uv} n_v$ common neighbors in C_v for any $1 \leq u \leq v \leq k$ pair. More exactly, for every $1 \leq u \leq v \leq k$ and $i, j \in C_u$:*

$$\sum_{t \in C_v} a_{it} = p_{uv} n_v + o(n);$$

and for every $1 \leq u \leq v \leq k$ and $i, j \in C_u$:

$$\sum_{t \in C_v} a_{it} a_{jt} = p^2_{uv} n_v + o(n).$$

The implications of some slightly modified versions have been proved by Bolla M. As for the **PIII→PII** implication, here we state more generally, for rectangular arrays, the next theorem. For the proof see [2].

Theorem 5 *Let $\boldsymbol{A}$ be an $m \times n$ rectangular array of nonnegative entries, and $s_1, \ldots, s_r$ be the positive singular values of the normalized table $\boldsymbol{A}_D$ introduced in (3.1), where $r = \mathrm{rank}(\boldsymbol{A})$. Then*

$$s_k = O(\sqrt{\log m \log n})\, \mathrm{disc}_k(\boldsymbol{A})$$

for any positive integer $k < r$, where $\mathrm{disc}_k(\boldsymbol{A})$ is defined in Definition 4.

Publications of Ahmed Elbanna

[1] Bolla, M., Elbanna, A., Estimating Parameters of a Probabilistic Heterogeneous Block Model via the EM Algorithm. Journal of Probability and Statistics, 1, 1-14 (2015).

[2] Bolla, M., Elbanna, A., Matrix and discrepancy view of generalized random and quasirandom graphs. Special Matrices, 1, 31-45 (2015).

[3] Bolla, M., Elbanna, A., Extending the Rash model to a multiclass parametric network model. In: Proceedings of the 1st International Conference on Future RFID Technologies, 5-7 November, 2014, Eger, Eszterházy Károly College (ed. G. Kruspér), pp. 15-22 (2014).

[4] M. Bolla, Elbanna, Multiple interaction strategies in networks related to graph spectra and dominant sets, accepted in Libertas Mathematica (2017).

[5] M. Bolla, Elbanna, A., Mala, J., Estimating parameters of a directed weighted graph model with beta-distributed edge-weights, submitted to Journal of Mathematical Sciences (2017).

Other references

[1] Bolla M., Recognizing linear structure in noisy matrices, Linear Algebra and Its Applications, 402, 228-244 (2005).

[2] Bolla M., Noisy random graphs and their Laplacians, Discrete Mathematics, 308, 4221-4230 (2008).

[3] Bolla,M., *Spectral Clustering and Biclustering. Learning Large Graphs and Contingency Tables*. Wiley (2013).

[4] Bolla M., Modularity spectra, eigen-subspaces and structure of weighted graphs, European Journal of Combinatorics, 35, 105-116 (2014).

[5] Bollobás B., Random Graphs, 2nd edition (Cambridge Univ. Press, Cambridge, 2001).

[6] Borgs C., J. T. Chayes, L. Lovász, V. T. Sós, K. Vesztergombi, Convergent graph sequences I: Subgraph Frequencies, metric properties, and testing, Advances in Math. 219, 1801 (2008).

[7] Chung, F., R. Graham, R. K. Wilson, Quasi-random graphs, Combinatorica 9, 345 (1989).

[8] Chung, F., R. Graham, Quasi-random graphs with given degree sequences, Random Struct. Algorithms 12, 1 (2008).

[9] Csiszár,V., Hussami,P., Komlós,J., Móri,T.F., Rejtő,L. and Tusnády,G., When the degree sequence is a sufficient statistic, *Acta Math. Hung.* **134**, 45-53 (2011).

[10] Lovász, L., T. Sós, V., Generalized quasirandom graphs, J. Comb. Theory B 98, 146 (2008).

[11] Motzkin, T. S., Straus, E. G., Maxima for graphs and a new proof of a theorem of Turán, *Canad. J. Math.* **17**, 533–540 (1965).

[12] Pavan, M., Pelillo, M., Dominant sets and hierarchical clustering. In Proc. of the 9th IEEE International Conference on Computer Vision (ICCV 2003) 2-Volume0Set, 362-369 (2003).

[13] Pavan, M., Pelillo, M., Dominant sets and pairwise clustering, *IEEE Trans. Pattern Anal. Machine Intell.* **29**(1) (2007), 167–172.

[14] Pinch, R. G. E., Math. Proc. Cambridge Phil. Soc. 99, 19 (1986).