Structural Similarity in Graphs[*]

A Relaxation Approach for Role Assignment

Ulrik Brandes and Jürgen Lerner[**]

Department of Computer & Information Science, University of Konstanz

Abstract. Standard methods for role assignment partition the vertex set of a graph in such a way that vertices in the same class can be considered to have equivalent roles in the graph. Several classes of equivalence relations such as regular equivalence and equitable partitions have been proposed for role assignment, but they all suffer from the strictness of classifying vertices into being either equivalent or not. It is an open problem how to allow for varying degrees of similarity. Proposals include ad-hoc algorithmic approaches and optimization approaches which are computationally hard.

In this paper we introduce the concept of *structural similarity* by relaxation of equitable partitions, thus providing a theoretical foundation for similarity measures which enjoys desirable properties with respect to existence, structure, and tractability.

1 Introduction

Role assignment is an important tool in the analysis of, e.g., social networks and food webs. Given a graph one tries to assign roles to vertices by looking for partitions of the vertex set such that equivalent vertices can be considered to occupy the same kind of structural position [20].

Whether a vertex partition yields a "meaningful" role assignment is up to some notion of compatibility with the edges of the graph. In their seminal paper, Lorrain and White [18] proposed that vertices have the same role if they are *structurally* equivalent, i.e. have identical neighborhoods. This rather restrictive requirement has later been weakened in many ways (see [21, 12] for an overview).

Vertex partitions are well investigated in some areas of computer science or graph theory. E.g. *equitable* partitions in algebraic graph theory [15] serve to find *quotients* of graphs and to determine the spectrum of highly symmetric, huge graphs. *Stable* partitions (also *bisimulations* or *regular equivalences*) are used to determine equivalent states of finite state processes in the calculus of communicating systems [19]. Shortly, a vertex coloring is called *regular* (*equitable*) if equivalent vertices have the same (number of each) colors in their neighborhood.

However, practical applicability of these kinds of vertex partitions to social networks or other irregular graphs is severely limited. Firstly it is $\mathcal{NP}$-complete

---

[**] Corresponding author, `lerner@inf.uni-konstanz.de`

to decide whether a graph has a regular, or equitable, partition with a given quotient graph (see Theorem 7). Secondly, most graphs analyzed in the social sciences have small or even trivial automorphism groups, irregularly distributed vertex degrees, etc., so that they hardly ever admit a non-trivial equitable partition anyway, while tractable regular partitions are often trivial as well (e.g. the maximal regular equivalence of an undirected graph is simply the division into isolates and non-isolates). Finally, while it makes sense that small perturbations like adding or deleting single edges destroy or establish the equivalence of vertices in, e.g., finite state processes, this is counterintuitive in the case of social networks, which often contain measurement errors and are expected to display regular patterns only approximately. In the case of irregular graphs we expect that small perturbations cause vertices to be more or less *similar* but do not decide about equivalence or non-equivalence.

We summarize these remarks by observing that the real obstacle in applying role assignment methods to empirical grpahs is not an inappropriate formulation of compatibility of equivalence and graph structure, but the equivalence relation itself. In social networks and other graphs in which many pairs of vertices are somehow related, but not exactly equivalent, we need a notion of *similarity* of vertices, rather than equivalence.

Other attempts to introduce degrees of similarity have been unsatisfactory so far: Algorithmic approaches like REGE and CATREGE lack "a theoretical rationale for the measure of similarity produced" [7–p. 375], while optimization approaches (e.g. [3]) are computationally hard and suffer from the existence of many local optima.

We therefore propose the notion of *similarities* as relaxations of equivalence relations. From a compatibility requirement generalizing equitability, we obtain the subset of *structural* similarities which displays highly desirable mathematical properties. Our main results are that the set of all structural similarities can be described compactly by the eigenvalue decomposition of the adjacency matrix and that, in this framework, a graph R is a quotient of a graph G if and only if the characteristic polynomial of R divides the characteristic polynomial of G. In particular, this implies an efficient algorithm for the role assignment problem.

This paper is organized as follows. Basic notation is provided in Sect. 2. In Sect. 3 we derive similarities as relaxations of equivalence relations and introduce the condition for similarities to be compatible with a graph's structure. A characterization and some interesting properties of the set of edge-compatible similarities are given in Sect. 4. Computational issues are addressed in Sect. 5 and are illustrated on a well-studied dataset [14] in Sect. 6.

2   Preliminaries

2.1   Basic Notation

In this paper we consider *undirected* graphs, possibly with *multiedges*, *loops*, or edge *weights*. For background on graph theory see, e.g., [11].

We also need some definitions from linear algebra (see, e.g., [1]). In particular we need the notions of *Euclidean* vectorspaces, i.e. real vectorspaces $\mathcal{V}$ supplied with an *inner product* or *dot product* $\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \to \mathbb{R}$, and *eigendecomposition* of *symmetric* linear mappings in such spaces.

2.2   Graph Spaces

As in [4], we associate with a graph $G = (V, E)$ the *vertex space* $\mathcal{V} := \mathcal{V}(G) := \{f : V \to \mathbb{R}\}$, that is the vectorspace of all real-valued functions on the vertex set. $\mathcal{V}$ has *standard basis* $(f_v)_{v \in V}$, where $f_v(v) := 1$ and $f_v(u) := 0$ if $u \neq v$. An element f_v of the standard basis can be associated with the vertex v. Further, we define $\alpha(G) : \mathcal{V} \to \mathcal{V}$ to be the endomorphism determined by the action of the adjacency matrix $A(G)$ on the standard basis. This immediately generalizes to weighted graphs. Finally, we get an inner product $\langle \cdot, \cdot \rangle$ on $\mathcal{V}$ which is uniquely defined by $\langle f_v, f_v \rangle := 1$ and $\langle f_u, f_v \rangle := 0$, if $u \neq v$, for $u, v \in V$.

Definition 1. *Given a graph G, the pair $\mathcal{G}(G) = (\mathcal{V}(G), \alpha(G))$ is called the* graph space *of G. For two vectors $u, v \in \mathcal{V}$ the* weight *of the pair (u, v) is defined to be $w(u, v) := \langle u, \alpha(v) \rangle$.*

Conversely, suppose we are given a pair $\mathcal{G} = (\mathcal{V}, \alpha)$, where $\mathcal{V}$ is a Euclidean vectorspace and $\alpha : \mathcal{V} \to \mathcal{V}$ a symmetric endomorphism. Then $\mathcal{G}$ induces a complete, weighted graph $G = (V, E)$ by choosing an orthonormal basis of $\mathcal{V}$ as the vertex set V, and defining the weight of an edge $e = \{u, v\}$ by $w(e) := \langle u, \alpha(v) \rangle$. Thus a graph space might be seen as a collection of graphs. We stress that different bases will lead to different, possibly non-isomorphic, graphs.

Two graph spaces are identical if and only if the underlying vectorspaces and endomorphisms are the same. But, as for graphs, it is reasonable to identify graph spaces which differ only by a renaming of the vertex space. If $G = (V, E)$ and $H = (W, F)$ are graphs then an *isomorphism* from G to H is a bijection $\phi : V \to W$ such that for all $u, v \in V$, $A(G)_{uv} = A(H)_{\phi(u)\phi(v)}$, i.e., the adjacency matrices are, after appropriate ordering of vertices, identical. In the following definition, we relax the condition that isomorphisms have to be bijections between discrete sets to the condition that they have to be vectorspace-isomorphisms.

Definition 2. *Let $\mathcal{G} = (\mathcal{V}, \alpha)$ and $\mathcal{H} = (\mathcal{W}, \beta)$ be two graph spaces. An* isomorphism *from $\mathcal{G}$ to $\mathcal{H}$ is an orthogonal linear isomorphism $\varphi : \mathcal{V} \to \mathcal{W}$, such that for all $u, v \in \mathcal{V}$*

$$\langle u, \alpha(v) \rangle = \langle \varphi(u), \beta(\varphi(v)) \rangle \quad .$$

This is true if and only if $\alpha = \varphi^{\mathrm{T}} \beta \varphi$. In this case $\mathcal{G}$ and $\mathcal{H}$ are called isomorphic graph spaces.

Graph isomorphisms are special, more restrictive types of graph space isomorphisms.

3   Structural Similarities

Relaxing (boolean) equivalence, we introduce a (continuous) notion of similarity. It is derived from projections that relax the surjective mapping associated with an equivalence relation (i.e. which maps elements to their equivalence classes). We show that for each projection there is a unique associated similarity and vice versa. A natural generalization of equitability then ensures that similarities respect the structure of a graph.

Let us start by considering the discrete case. For an equivalence relation $\sim$ on the vertex set V of a graph $G = (V, E)$, let $W := V/\sim$ be its set of equivalence classes. The associated surjective mapping $\phi : V \to W$, that maps vertices to their equivalence class, defines a binary $|W| \times |V|$ matrix P where for $v \in V$, $w \in W$, $P_{wv} = 1$ iff $\phi(v) = w$. Such a matrix P satisfies the equation $PP^T = \mathrm{id}_W$, for boolean matrix multiplication.

Relaxations of such class mappings allow for the entries p_{wv}, instead of 1's ("v is in class w") and 0's ("v is not in class w"), real numbers ("v's degree of membership to w is $p_{wv} \in \mathbb{R}$").

Definition 3. *Let $\mathcal{V}$ and $\mathcal{W}$ be two euclidian vectorspaces. A surjective linear homomorphism $\pi : \mathcal{V} \to \mathcal{W}$ is called a* projection, *if $\pi\pi^T = \mathrm{id}_{\mathcal{W}}$.*

These generalized mappings to classes suggest the following generalization of quotients of graphs to quotients of graph spaces. Considering the discrete case, let P be the (binary) characteristic matrix of a surjective mapping of the vertices of G onto the set W of equivalences classes. Then, following [15–Sect. 9.6], we obtain a (directed, weighted) graph $H = (W, F)$ (called the *quotient* of G modulo $\sim$) that is defined by its adjacency matrix $A(H) := (PP^T)^{-1}PAP^T$, where A is the adjacency matrix of G. This definition is motivated by the fact that for two classes, say $w, w' \in W$, $A(H)_{ww'}$ is the average number of edges between vertices in w and vertices in w'. The translation to projections is straight-forward:

Definition 4. *For a graph space $\mathcal{G} = (\mathcal{V}, \alpha)$ and a projection $\pi : \mathcal{V} \to \mathcal{W}$ to a euclidian vectorspace $\mathcal{W}$, π and α induce an endomorphism $\beta : \mathcal{W} \to \mathcal{W}$ by*

$$\beta := \pi\alpha\pi^T \ .$$

The resulting graph space $\mathcal{G}/\pi = (\mathcal{W}, \beta)$ is called the quotient *of $\mathcal{G}$ modulo π.*

Note that, since α is symmetric, β is symmetric.

We are now ready to define similarities as relaxations of equivalence relations, such that similarities and projections are associated with each other just like equivalence and class membership relations. Reconsidering the discrete case, let $\phi : V \to W$ be a surjective mapping and P be the (binary) characteristic matrix of ϕ. The equivalence relation induced by ϕ has characteristic matrix $S = P^T P$, since two vertices u and v are equivalent, iff the corresponding columns of P have the 1 in the same row, iff $s_{uv} = 1$.

If we relax P to a projection π, then $\sigma := \pi^T \pi$ is *symmetric*, i.e. $\sigma^T = \sigma$, and *idempotent*, i.e. $\sigma^2 = \sigma\sigma = \sigma$ and these properties serve to define our relaxation of equivalence relations.

Definition 5. *For a euclidian vectorspace $\mathcal{V}$ an endomorphism $\sigma : \mathcal{V} \to \mathcal{V}$ is called a* similarity *if it is symmetric, i.e. $\sigma^{\mathrm{T}} = \sigma$, and idempotent, i.e. $\sigma^2 = \sigma$.*

An equivalence relation $\sim$ induces a similarity represented by the matrix $S_\sim$, called here the *normalized matrix* of $\sim$, which is given by

$$(S_\sim)_{uv} := \begin{cases} 0 & \text{if } u \nsim v \\ 1/c & \text{if } u \sim v \text{ and } c \text{ is the size of the equivalence class of } v \end{cases} .$$

For example, the partition $\{\{1, 2, 3\}, \{4, 5\}\}$ of $\{1, \ldots, 5\}$ has normalized matrix

$$S = \begin{bmatrix} \left[\frac{1}{3}\right]_{3\times3} & [0]_{3\times2} \\ [0]_{2\times3} & \left[\frac{1}{2}\right]_{2\times2} \end{bmatrix} .$$

Note that, when multiplying normalized matrices, idempotency corresponds to transitivity. For a mapping σ, idempotency means that σ restricted to its *image* $\operatorname{im}\sigma$ is the identity.

Like equivalence relations and their associated surjective mappings, similarities and projections are just two points of view of the same concept: Let $\mathcal{V}$ and $\mathcal{W}$ be two euclidian vectorspaces and $\pi : \mathcal{V} \to \mathcal{W}$ a projection. Then, $\sigma_\pi := \pi^{\mathrm{T}}\pi$ is symmetric and idempotent, i.e., a similarity. Conversely, let $\mathcal{V}$ be a euclidian vectorspace and $\sigma : \mathcal{V} \to \mathcal{V}$ a similarity. Set $\mathcal{W} := \operatorname{im}\sigma$. Then, $\pi_\sigma = (v \mapsto \sigma(v)) : \mathcal{V} \to \mathcal{W}$ is a projection.

The next theorem states that $\sigma \mapsto \pi_\sigma$ and $\pi \mapsto \sigma_\pi$ are mutually inverse (up to isomorphism).

Theorem 1. *Let $(\mathcal{V}, \alpha)$ be a graph space and $\sigma : \mathcal{V} \to \mathcal{V}$ a similarity. Then $\sigma = \sigma_{\pi_\sigma}$. Conversely, let $(\mathcal{V}, \alpha)$ be a graph space, $\mathcal{W}$ a euclidian vectorspace, $\pi : \mathcal{V} \to \mathcal{W}$ a projection, $\mathcal{U} := \operatorname{im}\sigma_\pi$ and $\pi' := \pi_{\sigma_\pi}$. Then, with $\beta := \pi\alpha\pi^{\mathrm{T}}$ and $\beta_\mathcal{U} := \pi'\alpha\pi'^{\mathrm{T}}$, $(\mathcal{W}, \beta)$ and $(\mathcal{U}, \beta_\mathcal{U})$ are isomorphic graph spaces and π and π' are the same (up to this isomorphism).*

Proof. For proving the second part, note that $\varphi := \pi_{|\mathcal{U}} = \pi\pi'^{\mathrm{T}}$ is an isomorphism of the graph spaces $(\mathcal{W}, \beta)$ and $(\mathcal{U}, \beta_\mathcal{U})$. The proof therefore follows from the definitions and properties of the mappings in question. $\square$

As a consequence of Theorem 1 we conclude that for each similarity there is a unique associated projection and vice versa.

A similarity σ defines for a pair of vectors $u, v \in \mathcal{V}$ a real number $s(u, v)$, which is a measure for the similarity of u and v, by:

$$s(u, v) := \langle \pi(u), \pi(v) \rangle = \langle u, \pi^{\mathrm{T}}\pi(v) \rangle = \langle u, \sigma(v) \rangle ,$$

where $\pi := \pi_\sigma$. If u and v are basis vectors (vertices) then their similarity is the uv'th entry of the matrix of σ.

Until now we have introduced similarities as relaxations of arbitrary equivalence relations. The following condition is a natural generalization of equitability and ensures that roles are based on the graph's structure.

Definition 6. *Let $(\mathcal{V}, \alpha)$ be a graph space, $\mathcal{W}$ a euclidian vectorspace and π : $\mathcal{V} \to \mathcal{W}$ a projection. Then, π is called* structural *if, with $\beta := \pi\alpha\pi^{\mathrm{T}}$, the diagram*

$$\begin{array}{ccc} \mathcal{V} & \xrightarrow{\ \alpha\ } & \mathcal{V} \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ \mathcal{W} & \xrightarrow{\ \beta\ } & \mathcal{W} \end{array}$$

commutes, i.e., $\pi\alpha = \beta\pi$. The associated similarity σ_π is called structural *as well.*

4 Properties of Structural Similarities

We present several equivalent characterizations of structural similarity and show that equitable partitions form a special, restrictive case. Moreover, we generalize the partial order on the set of equivalence relations to a partial order on the set of similarities and show that the set of structural similarities forms a sublattice of the lattice of all similarities.

4.1 Characterization

Structural similarities can be characterized in several ways. A subspace $\mathcal{U} \subset \mathcal{V}$ is called α-*invariant* if $\alpha(\mathcal{U}) \subset \mathcal{U}$.

Theorem 2. *Let $(\mathcal{V}, \alpha)$ be a graph space, $\mathcal{W}$ a euclidian vectorspace and π : $\mathcal{V} \to \mathcal{W}$ a projection. Let $\sigma := \sigma_\pi$ be the associated similarity. Then the following assertions are equivalent:*

    1. π is structural;   *3. $\ker\sigma$ (or $\ker\pi$) is α-invariant;*

    2. $\sigma\alpha = \alpha\sigma$;   *4. $\operatorname{im}\sigma$ is α-invariant.*

Proof. Let $\beta := \pi\alpha\pi^{\mathrm{T}}$. *2 implies 1*: Multiplying $\sigma\alpha = \alpha\sigma$ with π from the left, we get $\pi\pi^{\mathrm{T}}\pi\alpha = \pi\alpha\pi^{\mathrm{T}}\pi$, which implies $\pi\alpha = \beta\pi$. *1 implies 2*: Given π with $\pi\alpha = \beta\pi$ we get

$$\sigma\alpha = \pi^{\mathrm{T}}\pi\alpha = \pi^{\mathrm{T}}\beta\pi = \pi^{\mathrm{T}}\pi\alpha\pi^{\mathrm{T}}\pi = \sigma\alpha\sigma \ .$$

Taking the transpose of this equation and using symmetry of α and σ we obtain $\alpha\sigma = \sigma\alpha\sigma$, hence $\sigma\alpha = \alpha\sigma$. *3 holds iff 4 holds* since, for the symmetric endomorphism σ, $\operatorname{im}\sigma$ is α-invariant iff $\ker\sigma$ is α-invariant. *2 implies 3*: Assume that $\alpha(\ker\sigma) \not\subset \ker\sigma$. Then there is a $v \in \mathcal{V}$ such that $\sigma(v) = 0$ and $\sigma\alpha(v) \neq 0$. Then, $\sigma\alpha(v) \neq 0 = \alpha\sigma(v)$, whence $\sigma\alpha \neq \alpha\sigma$. *3 implies 2*: Let $v \in \mathcal{V} = \ker\sigma \oplus \operatorname{im}\sigma$. We have $v = u_1 + u_2$ with $u_1 \in \ker\sigma$ and $u_2 \in \operatorname{im}\sigma$. Then, since $\alpha(u_1) \in \ker\sigma$ and $\alpha(u_2) \in \operatorname{im}\sigma$ (by *3 iff 4*), we have that $\sigma\alpha(v) = \alpha(u_2) = \alpha\sigma(v)$, whence $\sigma\alpha = \alpha\sigma$. $\qquad\square$

As a corollary, Theorem 2 yields a compact description of structural similarities by the eigenvalue decomposition of the adjacency matrix.

Corollary 1. *Let $(\mathcal{V}, \alpha)$ be a graph space. A similarity $\sigma : \mathcal{V} \to \mathcal{V}$ is structural, iff $\ker \sigma$ (or $\operatorname{im} \sigma$) is generated by eigenvectors of α.*

Proof. By Theorem 2, σ is structural iff $\ker \sigma$ (or $\operatorname{im} \sigma$) is α-invariant. If $\ker \sigma$ ($\operatorname{im} \sigma$) is generated by eigenvectors, then $\ker \sigma$ ($\operatorname{im} \sigma$) is obviously invariant. Conversely, if $\mathcal{U} := \ker \sigma$ ($\mathcal{U} := \operatorname{im} \sigma$) is α-invariant, then α restricted to $\mathcal{U}$ is a symmetric endomorphism on $\mathcal{U}$ and, thus, $\mathcal{U}$ admits a basis of eigenvectors. ☐

If follows from Theorem 2 that structural similarities are indeed a relaxation of equitable partitions.

Theorem 3. *Let $G = (V, E)$ be a graph and S the normalized matrix of an equivalence relation $\sim$ on V. Then, S defines a structural similarity for the graph space $\mathcal{G}(G)$ if and only if $\sim$ is equitable for G.*

Proof. Is a consequence of [15–Lemma 9.3.2] and the equivalence of 1. and 4. in Theorem 2. ☐

4.2   Lattice Structure

We recall the notion of a lattice (see, e.g., [16]). A *lattice* L is a partially ordered set L such that for each $a, b \in L$ there is a *supremum* (least upper bound) and an *infimum* (greatest lower bound) of a and b. A lattice L is called *complete* if suprema and infima exist for each subset $L' \subset L$.

In the next lemma we establish a connection between similarities and subspaces, which we use to define a partial order on the set of similarities.

Lemma 1. *Let $\mathcal{U} \subset \mathcal{V}$ be a subspace of a Euclidean vectorspace $\mathcal{V}$. Then there is a unique similarity $\sigma : \mathcal{V} \to \mathcal{V}$ such that $\mathcal{U} = \ker \sigma$.*

Proof. Suppose σ is such a similarity. First we observe that, since σ has to be symmetric, $\operatorname{im} \sigma$ is the orthogonal complement of $\ker \sigma = \mathcal{U}$, hence $\operatorname{im} \sigma$ is determined by $\mathcal{U}$. Thus, each $v \in \mathcal{V}$ admits a unique representation $v = v_1 + v_2$, where $v_1 \in \operatorname{im} \sigma$ and $v_2 \in \mathcal{U}$. Further, since σ is idempotent, it is the identity on $\operatorname{im} \sigma$. Thus

$$\sigma(v) = \sigma(v_1 + v_2) = v_1, \tag{1}$$

which shows the uniqueness. On the other hand, (1) can be used to define a similarity σ, such that $\ker \sigma = \mathcal{U}$, which shows the existence. ☐

By the above lemma, we get a bijection between the set of subspaces $\mathbb{U}(\mathcal{V}) := \{\mathcal{U} \subset \mathcal{V}\}$ and the set of similarities $\mathbb{S}(\mathcal{V}) := \{\sigma : \mathcal{V} \to \mathcal{V}; \sigma^{\mathrm{T}} = \sigma \text{ and } \sigma^2 = \sigma\}$. Via this bijection $\mathbb{S}(\mathcal{V})$ can be supplied with a partial order $\leq$ by

$$\sigma \leq \tau \iff \ker \sigma \subset \ker \tau, \text{ for } \sigma, \tau \in \mathbb{S}(\mathcal{V}) \ .$$

If $\sim_1$ and $\sim_2$ are two equivalence relations on the vertex set V of a graph $G = (V, E)$ such that $\sim_1$ is finer than $\sim_2$, then the corresponding normalized matrices define similarities σ_1 and σ_2 satisfying $\sigma_1 \leq \sigma_2$. Thus, the embedding of

equivalence relations on V into the set of similarities $\mathbb{S}(\mathcal{V}(G))$ is order-preserving and the above-defined partial order is a generalization of the order on the set of equivalence relations.

$\mathbb{U}(\mathcal{V})$ is a complete lattice, where the infimum is given by intersection and the supremum by the sum of subspaces. Thus, $\mathbb{S}(\mathcal{V})$ is a complete lattice, too. We show that the set $\mathbb{S}_\alpha(\mathcal{V})$ of structural similarities forms a sublattice of $\mathbb{S}(\mathcal{V})$:

Theorem 4. *Let $(\mathcal{V}, \alpha)$ be a graph space. Then, the set $\mathbb{S}_\alpha(\mathcal{V})$ of structural similarities is a complete sublattice of $\mathbb{S}(\mathcal{V})$.*

Proof. By Theorem 2 it suffices to show that the intersection of α-invariant subspaces is α-invariant and that the sum of α-invariant subspaces is α-invariant. Proving this is straightforward. □

The above theorem implies (see Theorem 5) that, given a similarity σ, that is not necessarily structural, there is always a smallest structural similarity above σ and a biggest structural similarity below σ. Note, that this doesn't hold, neither for regular equivalence relations [9], nor for equitable partitions, which can easily be verified.

Definition 7. *Let $(\mathcal{V}, \alpha)$ be a graph space, and $\sigma : \mathcal{V} \to \mathcal{V}$ be a similarity on $\mathcal{V}$ (not necessarily structural). Then, a similarity $\tau : \mathcal{V} \to \mathcal{V}$ is called a* structural hull *of σ if τ is structural, $\sigma \leq \tau$, and for all $\tau' \in \mathbb{S}_\alpha(\mathcal{V})$, $\sigma \leq \tau'$ implies $\tau \leq \tau'$. A similarity $\tau : \mathcal{V} \to \mathcal{V}$ is called a* structural interior *of σ if τ is structural, $\tau \leq \sigma$, and for all $\tau' \in \mathbb{S}_\alpha(\mathcal{V})$, $\tau' \leq \sigma$ implies $\tau' \leq \tau$. If they exist, structural hulls and interiors are unique.*

Theorem 5. *Let $(\mathcal{V}, \alpha)$ be a graph space. Then, in $\mathbb{S}(\mathcal{V})$ there exist structural hulls and interiors.*

Proof. We utilize the fact that $L_1 := \mathbb{S}_\alpha(\mathcal{V})$ is a complete sublattice of the complete lattice $L := \mathbb{S}(\mathcal{V})$. Let $\sigma \in L$ then $\inf\{\tau \in L_1 \,;\, \sigma \leq \tau\}$ is the structural hull and $\sup\{\tau \in L_1 \,;\, \tau \leq \sigma\}$ is the structural interior of σ. □

5   Determining Structural Similarities

Corollary 1 yields a general procedure for determining structural projections: Select a subset of eigenvalues and associated eigenvectors of a graph and project its graph space onto the subspace generated by these eigenvectors. For example, the equitable partition of the graph in Fig. 1 corresponds to a structural similartity that is the projection onto the sum of eigenspaces corresponding to eigenvalues 3, 1 and -2.

If no eigenvalue has multiplicity greater than one (or we demand that eigenspaces must not be divided into subspaces), then Corollary 1 implies that, for fixed k, all structural projections onto k-dimensional image spaces can be listed efficiently. As a consequence of [13], the corresponding task for discrete partitions, i.e. to list all, say, regular equivalences with exactly k classes, is $\mathcal{NP}$-hard.
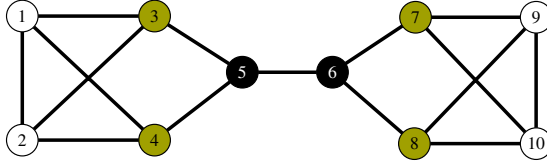
Fig. 1. Example from [6]. The vertex coloring represents an equitable partition of the drawn graph; the same partition can be found by a structural projection onto eigenvalues 3, 1, and -2

The following theorem characterizes role assignability with structural similarities.

Theorem 6. *Let $\mathcal{G} = (\mathcal{V}, \alpha)$ and $\mathcal{R} = (\mathcal{W}, \beta)$ be two graph spaces. Then there is a structural projection $\pi : \mathcal{V} \to W$ such that $\beta = \pi\alpha\pi^{\mathrm{T}}$ if and only if the characteristic polynomial of β divides the characteristic polynomial of α.*

Proof. The *only if* part can be proved along the same lines as [15–Theorem 9.3.3]. *If*: Let $\{(\lambda_1, \nu_1), \ldots, (\lambda_r, \nu_r)\}$ be the set of different eigenvalues λ_i of β with multiplicity ν_i. For each $1 \leq i \leq r$ choose ν_i linearly independent eigenvectors $v_{i1}, \ldots, v_{i\nu_i}$ of α with eigenvalue λ_i. Let $\mathcal{U} \subset \mathcal{V}$ be the subspace, generated by $(v_{11}, \ldots, v_{1\nu_1}, v_{21}, \ldots, v_{r\nu_r})$. Then, the projection $\pi_1 : \mathcal{V} \to \mathcal{U}$ is structural and $\beta_1 := \pi_1\alpha\pi_1^{\mathrm{T}} = \alpha_{|\mathcal{U}}$ has exactly the eigenvalues, with multiplicities, of β. Finally define $\varphi : \mathcal{U} \to \mathcal{W}$ to be the isomorphism which maps $(v_{11}, \ldots, v_{r\nu_r})$ onto an orthonormal eigenbasis of $(\mathcal{W}, \beta)$ in such a way that the image of v_{ij} has eigenvalue λ_i. The assertion follows with $\pi := \varphi\pi_1$.   □

The above theorem implies that the role assignment problem for structural similarities is easier than its discrete counterparts.

Theorem 7. *It is $\mathcal{NP}$-complete to decide for two graphs G and R whether there is*

1. *a regular role assignment for G with given role graph R,*
2. *an equitable partition for G with quotient R.*

On the other hand, it can be tested efficiently whether there exists a structural similarity for the graph space of G with quotient $\mathcal{G}(R)$.

Proof. 1. is proved in [13]. 2. holds since the $\mathcal{NP}$-complete decision problem, whether a 3-regular graph has a perfect code [17], can be formulated as the problem whether this graph has an equitable partition with given quotient. The last statement follows immediately from Theorem 6.   □

6   Example: Southern Women Data

Theorem 6 is especially useful to determine structural projections. We illustrate this on a well-studied network representing 18 women who are connected

by weighted edges signifying the number of co-appearance at 14 selected social events ("Southern Women" data set [10]). A meta-analysis of this data is presented in [14].

It is a striking observation that a number of commonly used techniques can be seen as determining structural projections onto specific role graphs. The simple examples below mostly serve to illustrate that our method also gives a uniform interpretation to established, but seemingly unrelated methods of network analysis. It should be obvious that there are more sophisticated uses of Theorem 6 with more complex role graphs.

$R_c :$ $R_{c_1 c_2} :$ $R_{cp} :$

Fig. 2. Role graphs for centrality (*left*), 2-way partitioning (*middle*) and core/periphery structure (*right*)

Eigenvector Centrality. A widely-used approach to determine the structural importance of vertices in a graph is eigenvector centrality [5], in which vertices are valued by the entries of the principal eigenvector of the adjacency matrix.

Vertices are therefore central if they have central neighbors, so that centrality can be viewed as the degree of membership in the only role present in role graph R_c shown in Fig. 2.

From the proof of Theorem 6 we know that a similarity σ is structural with $G/\sigma = R_c$, if and only if σ is the projection onto a 1-dimensional space generated by an eigenvector associated with eigenvalue $c > 0$ of the adjacency matrix of G. Therefore, eigenvector centrality is precisely a structural projection onto R_c, i.e. a 1-dimensional role assignment, and the eigenvalue determines the weight of the loop.

Spectral k-Clustering. Eigenvectors are also frequently used to partition a graph into dense clusters. Using the spectrum of the adjacency matrix has been advocated, e.g., in [2]. Membership in a cluster can again be seen as a role, and partitioning into two roles thus corresponds to projecting onto role graph $R_{c_1 c_2}$ shown in Fig. 2.

Again, Theorem 6 implies that a similarity σ is structural with $G/\sigma = R_{c_1 c_2}$, if and only if σ is the projection onto a 2-dimensional space generated by eigenvectors corresponding to eigenvalues $c_1, c_2 > 0$. The result using the first two eigenvecctors is shown in Fig. 3(*left*). This figure shows the usefulness of real-valued degrees of membership: Vertices 8 and 9 are rather between the two clusters, which is consistent with the fact that commonly used methods disagree largely about the cluster these vertices belong to [14]. Vertices 16, 17, and 18 are assigned to the second cluster, but only with low degrees of membership. Again, this is a precise result consistent with [14], where many methods did not assign these vertices at all.
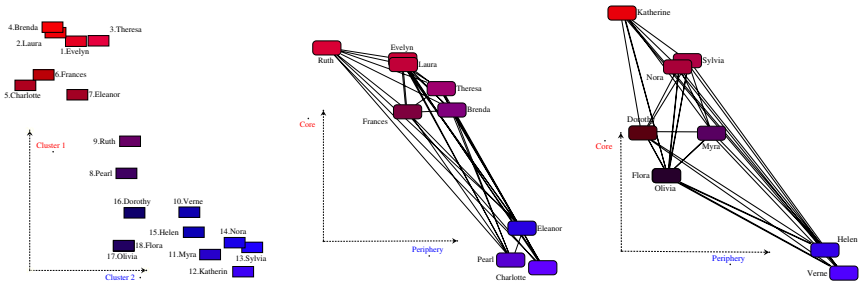
Fig. 3. projections for 2-clustering (*left*) and core/periphery structure in resulting clusters (*middle* and *right*)

Core/Periphery Structure. Borgatti and Everett [8] discuss intuitive ideas and formal definitions for a frequently encountered phenomenon in social networks, namely their division into a cohesive *core* of densely knit actors and a loosely connected *periphery*.

The degree of membership to both core and periphery of a vertex can be determined by a structural projection onto role graph R_{cp} of Fig. 2, although the spectrum of the role graph is not as trivial as those of the other two. However, the eigenvalues λ_1, λ_2 of R_{cp} satisfy $c = \lambda_1 + \lambda_2$ and $-p^2 = \lambda_1 \cdot \lambda_2$. Since it is reasonable to demand $c, p \in \mathbb{R}_{\geq 0}$, these equations imply w.l.o.g. $\lambda_1 \leq 0 \leq \lambda_2$ and $|\lambda_1| \leq |\lambda_2|$. Standard results in spectral graph theory imply that for loopless connected graphs the smallest and the largest eigenvalue always satisfy these inequalities. Thus, by Theorem 6, the projection σ onto the 2-dimensional space generated by the corresponding eigenvectors of G is structural with $G/\sigma = R_{cp}$. The resulting role assignment for the two clusters suggested by [14] of the data set is in Fig. 3(*middle* and *right*). Observe that the more a vertex is in the core, the more it is connected to core and to peripheral vertices, whereas vertices that have high peripheral values are connected mostly to the core.

References

1. Artin, M.: Algebra. Prentice Hall (1991)
2. Barnes, E.R.: An algorithm for partitioning the nodes of a graph. SIAM Journal on Algebraic and Discrete Methods **3** (1982) 541–550
3. Batagelj, V., Doreian, P., Ferligoj, A.: An optimizational approach to regular equivalence. Social Networks **14** (1992) 121–135
4. Bollobás, B.: Modern Graph Theory. Springer (1998)
5. Bonacich, P.: Factoring and weighting approaches to status scores and clique identification. Journal of Mathematical Sociology **2** (1972) 113–120
6. Borgatti, S.P., Everett, M.G.: The class of all regular equivalences: Algebraic structure and computation. Social Networks **11** (1989) 65–88
7. Borgatti, S.P., Everett, M.G.: Two algorithms for computing regular equivalence. Social Networks **15** (1993) 361–376
8. Borgatti, S.P., Everett, M.G.: Models of core/periphery structures. Social Networks **21** (1999) 375–395

9. Boyd, J.P., Everett, M.G.: Relations, residuals, regular interiors, and relative regular equivalence. Social Networks **21** (1999) 147–165
10. Davis, A., Gardner, B., Gardner, M.: Deep south. The University of Chicago Press (1941)
11. Diestel, R.: Graph Theory. Springer-Verlag, New York (2000)
12. Everett, M.G., Borgatti, S.P.: Regular equivalence: General theory. Journal of Mathematical Sociology **19** (1994) 29–52
13. Fiala, J., Paulusma, D.: The computational complexity of the role assignment problem. In: Proceedings of the ICALP 2003., Springer–Verlag (2003) 817–828
14. Freeman, L.C.: Finding social groups: A meta-analysis of the southern women data. In Breiger, R., Carley, K.M., Pattison, P., eds.: Dynamic Social Network Modeling and Analysis. The National Academies Press (2003)
15. Godsil, C., Royle, G.: Algebraic Graph Theory. Springer (2001)
16. Grätzer, G.: General Lattice Theory. Birkhäuser Verlag (1998)
17. Kratochvíl, J.: Perfect Codes in General Graphs. Academia Praha (1991)
18. Lorrain, F., White, H.C.: Structural equivalence of individuals in social networks. Journal of Mathematical Sociology **1** (1971) 49–80
19. Milner, R.: A Calculus of Communicating Systems. Lecture Notes in Computer Science, 92. Springer Verlag, Berlin (1980)
20. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press (1994)
21. White, D.R., Reitz, K.P.: Graph and semigroup homomorphisms on networks of relations. Social Networks **5** (1983) 193–234