Combinatorial and Algebraic Approaches to Network Analysis*

M. Kraetzl[†], R. Laubenbacher[‡], M.E. Gaston[◇]

[†]Communications Division, DSTO, Salisbury, South Australia
miro.kraetzl@dsto.defence.gov.au
[‡]Dept. of Mathematical Sciences, New Mexico State University,
Las Cruces, NM, U.S.A.
reinhard@nmsu.edu
[◇]US DoD, Operations Research Modeling & Simulation Office
Ft. Meade, MD, U.S.A.
megasto@ncsc.mil

January 17, 2001

Abstract

In this paper novel approaches to change analysis of time series of dynamic networks are investigated. Combinatorial, algebraic, and topological techniques are proposed for measuring the distance between weighted graphs and digraphs. Various approaches to detection of abnormal changes in time series of graphs are explored. Using simplicial complexes as models, f-vectors, Q-analysis, A-theory, singular homology, and Betti numbers of Stanley-Reisner rings in measuring graph distance are considered.

1 Introduction

Abnormal change detection in a time series of dynamic communications or information networks is of great importance in network management and various other network analysis and control applications. Known techniques comprise various pattern recognition approaches,

*The authors are grateful to A. Jarrah and B. Stigler for their help with data analysis, and to G. Roberts for providing help with graphics. The second author thanks DSTO for its hospitality, while part of this work was completed.

spectral graph theory, string comparison, and mean/median graphs applications [7, 18]. In this paper we present several new graph measures based on combinatorics, algebraic and combinatorial topology, and abstract algebra. The key feature of our approach is to model the network as a simplicial complex, with an emphasis on detecting qualitative features of the network structure. There are several ways of doing this, depending on which graph theoretic features one wishes to focus on. Simplicial complexes are combinatorial versions of topological spaces and can be analyzed with combinatorial, topological, or algebraic methods. This feature makes them particularly versatile as models. We then use all three fields to associate numerical measures to a simplicial complex, which, in turn, provide measures for the underlying network. Applications of these measures to a real intranet network will be addressed in [12].

The contents of the paper are as follows. Section 2 introduces simplicial complexes, while in Section 3 several ways of constructing a simplicial complex from a digraph are described, including the neighborhood complex, the complete subgraph complex, containing as simplices all sub-cliques, and other complexes obtained by using any monotone graph property. Section 4 contains a description of several invariants of simplicial complexes, namely its A-homotopy groups, Q- and f-vectors, and the Betti numbers of the Stanley-Reisner ring associated to the complex. Sections 5 and 6 are devoted to the derivation of graph distance measures, using those invariants. Section 7 contains the definition of a difference and order complexes and a description of several possibilities of generalizing distance measures to time series of graphs. Section 8 introduces thresholding as a possible approach to implementation of graph distances in measuring distances of weighted graphs. Section 9 contains a detailed example of a time series of networks and their associated measures.

2 Simplicial Complexes

In this section we summarize the basic definitions, examples, and facts about simplicial complexes, in order to make the paper more self-contained.

Definition. A *simplicial complex* Δ on a finite set $V = \{v_1, \dots, v_n\}$ of *vertices* is a nonempty subset of the power set of V with the property that Δ is closed under the formation of subsets. That is, if $\sigma \in \Delta$ and $\tau \subset \sigma$, then $\tau \in \Delta$. The elements of Δ are called *simplices* or *faces*. The *dimension* of a simplex σ is equal to one less than the number of vertices defining it. The *dimension of* Δ is the maximum of the dimensions of all simplices in Δ.

Example. Let $V = \{1, 2, \ldots, 6\}$, and let Δ consist of the subsets $\{2, 3, 4\}, \{3, 4, 5, 6\}, \{1, 5, 6\}$ and their subsets. Then Δ can be represented geometrically as in Figure 1.

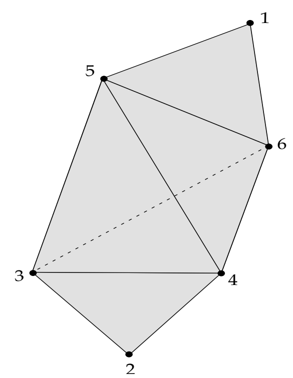

Figure 1: Example of a simplicial complex.

Its dimension is 3, as there is a 3-dimensional simplex, in addition to two 2-dimensional simplices attached to it.

This example shows that a simplicial complex, as defined here, is a purely combinatorial object, which in addition has a representation as a geometric object, and which can therefore be viewed as a topological space. It is this versatility of simplicial complexes that makes them extremely useful as models.

A convenient way to represent a simplicial complex, especially for computations, is via an incidence matrix, whose columns are labeled by its vertices, and the rows are labeled by the simplices. It is clearly sufficient to represent only those rows corresponding to maximal simplices (with respect to inclusion). For the complex in the above example an incidence matrix is

$$\begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 \end{pmatrix},$$

by using the column labels $1, 2, \ldots, 6$ from left to right.

3 Simplicial Complexes from Digraphs

In this section we associate several simplicial complexes to a directed graph (digraph). Any undirected graph should be considered directed by making each edge bidirectional. Later on we will consider digraphs with weighted edges. Let G be a digraph with vertices $v_1, \dots, v_n$. We construct several simplicial complexes from G. The first one is the so-called *neighborhood complex* $\mathfrak{N}(G)$ of G. Its vertices are the vertices of G. For each vertex v of G there is a simplex containing the vertex v, together with all vertices w corresponding to directed edges $v \to w$. By including all faces of those simplices, we obtain the neighborhood complex. Equivalently, $\mathfrak{N}(G)$ is generated by the simplices represented as the rows of the adjacency matrix of G, augmented by 1's in all diagonal entries.

The second complex associated to G which we want to consider has as simplices the complete subgraphs of G. Here we may consider G as undirected, if convenient. To be precise, the complex $\mathfrak{C}(G)$ has as vertices again the vertices of G. The maximal simplices are given by the collections of vertices that make up maximal (un)directed complete subgraphs, or *cliques* of G. Note that if a vertex of a complete subgraph is deleted then we again obtain a complete subgraph. For an example of both these constructions see the digraphs in Figure 6 and their associated complexes in Figure 7.

Note that, more generally, we could use any property of the graph G that is *monotone*, in the sense that it is preserved under deletion of vertices (or edges). For example, we could construct a simplicial complex whose simplices consist of all subgraphs of G without (directed or undirected) cycles. (This leads to graph matroids.) See [6] for details.

4 Invariants of Simplicial Complexes

We now describe a number of measures of simplicial complexes from several different points of view. Firstly, one can view a simplicial complex as a combinatorial model of a topological space and can then consider a variety of algebraic topological measures, such as homotopy and homology groups. Secondly, one can view the complex simply as a combinatorial object and consider several numerical invariants associated to it, such as its dimension, or its so-called f-vector, a more subtle invariant which counts the number of simplices in different dimensions. Finally, we can consider an algebraic model of the complex, its so-called Stanley-Reisner ring, the quotient of a polynomial ring on variables corresponding to the vertices of the complex, divided by the ideal generated by the non-faces of the complex. Each point of view provides very different measures of the complex (and, by extension, of the digraph that produced the complex).

First we describe a family of invariants that arise from Q-analysis and its extensions. Q-analysis, developed by R. Atkin in [2, 3], has been applied to various aspects of connectivity characterizations in social science [1, 4, 14], and transport networks [15]. In recent work by X. Kramer, and R. Laubenbacher, an extension of Q-analysis has been developed [16, 17] and, in work with M. Coombs and A. Taha, applied to detection and evaluation of IO in $\mathrm{C}^4\mathrm{I}$ networks [11]. Possible applications to time series analysis of influence structures in decision networks are outlined in [10]. A detailed version of this extension of Q-analysis, termed A-theory, can be found in [5]. A-theory generates a family of groups associated to a simplicial complex, similar to the homotopy groups of a topological space.

To be precise, let Δ be a simplicial complex of dimension d, $0 \leq q \leq d$ an integer, and σ a maximal simplex of Δ of dimension greater than or equal to q. There are groups

$$A_n^q(\Delta, \sigma), \quad 1 \leq n,$$

which are analogous to the homotopy groups of a topological space. The problem with classical homotopy of spaces applied to simplicial complexes is the insensitivity to the combinatorial structure of the complex. For instance, a hollow tetrahedron and a triangulated 2-sphere made up of a large number of 2-dimensional simplices (triangles) both look like a 2-sphere from the point of view of homotopy. In situations where the combinatorial information is important to retain, this is less than useful. Nonetheless, from the point of view of network analysis, homotopy-like measures are of great interest, because they can be used to detect essential network features that may be present under a variety of graph configurations. One can view A-theory as a version of homotopy which is sensitive to the combinatorics of the configuration, but nonetheless gives qualitative information about the network topology.

One disadvantage of A-theory, as of homotopy theory, is that it is hard to compute. At present the only existing (and implemented) algorithm computes the abelianization of the first group A_1^q. (In analogy to classical topology, the A_1-groups are in general nonabelian.) It has been successfully applied to the detection of interesting features in decision networks (see [9]). In Section 6 we will explain how to obtain measures of graph distance based on A-theory.

While invariants from classical topology applied to simplicial complexes have the drawback of not being very sensitive to the combinatorics, as explained above, they can still lead to useful measures. In the next section we will describe a graph measure based on the singular homology groups

$$H_n(\Delta, k), \; n \geq 0, k \text{ a field},$$

of the complex viewed as a topological space. For a definition see, e.g., [6].

If we view the complex Δ as a combinatorial object there are several numerical invariants attached to it that lead to potentially interesting graph measures. First of all, there is of course the dimension of the complex. More subtly, we can consider its f-vector. It is an integer vector with $\dim(\Delta) + 1$ entries, with the i-th one being equal to the number of i-dimensional simplices in Δ. The f-vector is a much studied invariant which plays an important role in network reliability [8].

Another integer vector associated to Δ is provided by Q-analysis. Fix an integer q, with $0 \leq q \leq \dim(\Delta)$. Call two simplices σ and τ *q-connected* if there is a sequence of simplices

$$\sigma, \sigma_1, \dots, \sigma_n, \tau$$

in Δ such that any two consecutive ones share a q-dimensional face. It is straightforward to see that q-connectivity is an equivalence relation. We associate to Δ its *Q-vector*, an integer vector of the same length as the f-vector, whose i-th entry is equal to the number of i-connectivity classes. It too can be used as a graph measure, as we will see later.

Finally, we consider an invariant of Δ that comes from abstract algebra. Let $x_1, \dots, x_n$ be the vertices of Δ, and let k be a field, for instance the field $\mathbb{R}$ of real numbers. We consider the polynomial ring $k[x_1, \dots, x_n] = k[\mathbf{x}]$, containing all polynomials in the variables $x_1, \dots, x_n$, with addition and multiplication of polynomials as the ring operations. Note that each simplex $\{x_{i_1}, \dots, x_{i_r}\}$ of Δ corresponds to a unique (square-free) monomial in $k[\mathbf{x}]$, namely $x_{i_1} \cdots x_{i_r}$. Now let $I \subset k[\mathbf{x}]$ be the ideal generated by all square-free monomials that correspond to non-faces in Δ, that is, collections of vertices that do not represent simplices in Δ. The *Stanley-Reisner ring of* Δ is the quotient ring

$$R_\Delta = k[\mathbf{x}]/I.$$

It should be viewed as an algebraic model of Δ, and many combinatorial properties of Δ are reincarnated as algebraic properties of R_Δ. One can associate to R_Δ its *Betti numbers*, a sequence of non-negative integers that provide a subtle measure of the relationships among the monomials that generate the ideal I, hence among the non-faces of the complex Δ. They are defined using homological algebra. It is known that each module over $k[\mathbf{x}]$ has a minimal free resolution, in particular so does R_Δ. That is, there exists an exact sequence

$$0 \longrightarrow k[\mathbf{x}]^{b_r} \longrightarrow k[\mathbf{x}]^{b_{r-1}} \longrightarrow \cdots \longrightarrow k[\mathbf{x}]^{b_1} \longrightarrow k[\mathbf{x}]^1 \longrightarrow R_\Delta \longrightarrow 0,$$

where $k[\mathbf{x}]^{b_i}$ stands for the Cartesian product

$$k[\mathbf{x}] \times \cdots \times k[\mathbf{x}]$$

with b_i factors, a free $k[\mathbf{x}]$-module of rank b_i. The image of each module homomorphism in the sequence is equal to the kernel of the subsequent homomorphism. The sequence is of minimal length, and the b_i are minimal. Such a minimal free resolution is unique up to isomorphism. The numbers $\{b_0 = 1, b_1, \ldots, b_r\}$ are the *Betti numbers* of R_Δ. It is also known that r is at most equal to n, the number of variables, that is, the number of vertices of Δ. For details see [19]. In the next section we will develop a graph measure based on the Betti numbers.

5 Graph Measures from Invariants of Simplicial Complexes

In this section we derive graph measures as well as measures of graph distance from the list of invariants of simplicial complexes that we described in the previous section. The graph measures will all consist of vectors with non-negative integer entries. Let Δ be a simplicial complex of dimension d, which we may think of as obtained from a digraph as outlined in Section 3.

5.1 The f-vector and the Q-vector

In the previous section we have defined the f-vector $(f_0, \ldots, f_d)$ and the Q-vector $(Q_0, \ldots, Q_d)$ of Δ. They are straightforward to compute, and several implementations are available. If so desired, one can convert both vector-valued measures into numerical measures by taking the (possibly weighted) sum of the entries.

5.2 A-theory

As mentioned earlier, the A-groups $A_n^q(\Delta, \sigma)$ of Δ based at a simplex $\sigma \in \Delta$ are at present only computable for $n = 1$. Even in this case, it is generally impossible to compute the group itself, and one has to be satisfied with its abelianization. Thus, we get a sequence of finitely generated abelian groups

$$A_1^0(\Delta, \sigma)^{\mathrm{ab}}, \ldots, A_1^d(\Delta, \sigma)^{\mathrm{ab}},$$

for each choice of base simplex σ. For the following it is important that we elaborate on the role of this base simplex. It is shown in [5] that, for a given q, the group $A_1^q(\Delta, \sigma)$ does not change (up to isomorphism), if we replace σ by a simplex in the same q-connected component. It can, however, change dramatically if we change q-connected components. Thus, A-theory can be viewed as a measure of a complex that refines the calculation of the q-connected components. Thus, we obtain one A_1^q-group for each q-connected component of Δ.

Now observe that any finitely generated abelian group can be written as a direct sum of a torsion group and a free abelian group, which is isomorphic to $\mathbb{Z}^r$ for some nonnegative integer r. The integer r is called its *free rank*. Applying this to the abelianized A_1-groups we can associate a numerical measure to each group, which in turn leads to a vector-valued measure of Δ based on the A-groups.

Define the A-vector

$$(A^0, A^1, \dots, A^d)$$

of Δ as follows. Let A^q be the sum of the free ranks of all the groups $A_1^q(\Delta, \sigma)$, where σ varies over all the q-connected components of Δ.

5.3 Singular Homology

Viewing Δ as a topological space rather than a combinatorial object one can compute its singular homology with coefficients in a field k, such as $\mathbb{Q}$ or $\mathbb{R}$. Heuristically speaking, the singular homology groups of a topological space measure the existence of "holes" of various dimensions in the space. They are defined and computed algebraically, via a sequence of k-vector spaces and linear transformations. For details see, e.g., [20]. We obtain a sequence of k-vector spaces

$$H_0(\Delta, k),\ H_1(\Delta, k), \dots,\ H_d(\Delta, k).$$

(It is known that these groups vanish in all dimensions above the dimension d of the complex.) Assembling their vector space dimensions H_i into the *H-vector*, we obtain another vector-valued measure

$$(H_0, \dots, H_d)$$

associated to Δ.

To illustrate the significance of the H-vector, consider the following example. Let Δ be the complex given by a hollow tetrahedron, with a one-cycle graph attached to it at one of its vertices. Then the H-vector of Δ is

$$H_\Delta = (1, 1, 1).$$

Here, the first 1 indicates that the complex is connected, that is, has a single connected component, whereas the second 1 measures the attached graph, which topologically looks like a one-dimensional sphere. Finally, the third 1 measures the 2-dimensional hole inside the hollow tetrahedron.

5.4 The Betti Numbers

Finally, we derive another vector-valued measure from the Betti numbers of the Stanley-Reisner ring attached to Δ, by assembling the Betti numbers $b_0 = 1, \ldots, b_r$ into the B-vector

$$B_\Delta = (b_0, b_1, \ldots, b_r).$$

In order to give the reader a feeling for the B-vector we compute it for a simple example. Let Δ be the complex on the vertex set $\{1, \ldots, 4\}$, with simplices $\{1,2,3\}$, $\{3,4\}$, and their faces. The missing simplices in Δ are $\{1,4\}$ and $\{2,4\}$. Hence the ideal $I_\Delta \subset k[x_1, \ldots, x_4]$ is generated by the monomials x_1x_4 and x_2x_4. We construct a minimal free resolution for Δ as

$$0 \longrightarrow k[\mathbf{x}] \longrightarrow k[\mathbf{x}]^2 \longrightarrow k[\mathbf{x}] \longrightarrow R_\Delta \longrightarrow 0.$$

Hence

$$B_\Delta = (1,2,1).$$

To see this, observe that the kernel of the projection $k[\mathbf{x}] \longrightarrow R_\Delta$ is equal to I_Δ, which is generated minimally by x_1x_4 and x_2x_4, that is, no fewer than two generators will generate I_Δ. Consider the composition of the two $k[\mathbf{x}]$-module homomorphisms

$$k[\mathbf{x}]^2 \longrightarrow I_\Delta \longrightarrow k[\mathbf{x}],$$

where the first map sends the first canonical basis element $(1,0)$ of $k[\mathbf{x}]^2$ to x_1x_4, and the second to x_2x_4. The second map is simply the inclusion. The composition is a map $\phi : k[\mathbf{x}]^2 \longrightarrow k[\mathbf{x}]$ whose image is precisely I_Δ, the kernel of the map $k[\mathbf{x}] \longrightarrow R_\Delta$.

Now observe that the element $(x_2, -x_1)$ lies in the kernel of ϕ since $x_2(x_1x_4) - x_1(x_2x_4) = 0$. In fact, one can show that it generates the kernel. We now add the map

$$k[\mathbf{x}] \longrightarrow k[\mathbf{x}]^2$$

to the sequence which sends 1 to the element $(x_2, -x_1) \in k[\mathbf{x}]^2$. This map is one-to-one, and we have constructed a free resolution. It is not hard to show that this is indeed the minimal one.

In summary, we have described several ways of generating an integer vector associated to a simplicial complex, namely the f-, Q-,A-, H-, and B-vectors. Combined with any encoding of a directed graph into a simplicial complex this provides many ways of obtaining vector-valued graph measures. Two of these encodings were described earlier, the neighborhood complex and the complete subgraph complex. In the next section we describe how to obtain graph distance measures from these.

6 Graph Distance Measures

Let G and G' be two digraphs. Assume that we have chosen a method of associating simplicial complexes $\Delta(G)$ and $\Delta(G')$ to them, and that we have chosen a vector-valued measure M on simplicial complexes. Let $M(G) = (m_0, \ldots, m_r)$ and $M(G') = (m'_0, \ldots, m'_s)$. Adding zero entries on the right, if necessary, we may assume that $r = s$, that is, the vectors have the same length.

Definition. Define the *M-distance* of G and G' to be

$$d_M(G, G') = \sum_{i=0}^{r} |m_i - m'_i|^2.$$

That is, $d_M(G, G')$ is the Euclidean distance of the M-vectors of G and G'.

Proposition. M-distance is symmetric and satisfies the triangle inequality.

Proof. This proposition follows immediately from the corresponding properties of Euclidean distance.

It is worth observing that the definition of M-distance can be generalized, if necessary, by assigning weights to the different dimensions of the vector. That is, given a sequence W of weights

$$w_0, w_1, \ldots, w_n, \ldots$$

consisting of real numbers, we can define the *weighted M-distance* of G and G' to be

$$d_M^W(G, G') = \sum_{i=0}^{r} w_i |m_i - m'_i|^2.$$

What prevents M-distance from being an actual metric on the space of digraphs and digraph morphisms is that non-isomorphic graphs may have M-distance zero for any of the measures and encodings defined. This simply reflects the fact that these measures detect qualitative differences in networks, for which graph isomorphism is too strong a condition. It remains an open problem to identify a suitable notion of digraph equivalence which has the property that two graphs are equivalent precisely if their M-distance is zero.

In summary, we have developed the following procedure to generate digraph metrics:

$$\boxed{\text{digraph} \longrightarrow \text{simplicial complex} \longrightarrow M\text{-vector}}$$

The M-distance of two digraphs is then measured as the distance, possibly weighted, of their M-vectors.

7 Measures for Time Series of Graphs

In this section we consider a time series of digraphs

$$G_1, \dots, G_n,$$

and develop some distance measures on it. With any graph distance measure $d(-,-)$ one can derive from this time series a sequence of successive distance measurements:

$$d(G_1, G_2), d(G_2, G_3), \dots, d(G_{n-1}, G_n);$$

$$|d(G_1, G_2) - d(G_2, G_3)|, \dots, |d(G_{n-2}, G_{n-1}) - d(G_{n-1}, G_n)|;$$
$$\dots$$

With graph measures via simplicial complexes several more sophisticated approaches become possible. Most importantly, "differences" can be taken at the simplicial complex level. Let Δ and Δ' be two complexes, represented by incidence matrices M and M'. We will assume that both complexes have the same vertex set. This is a reasonable assumption, since the digraphs in the time series frequently have the same set of nodes. We construct the *difference complex* $[\Delta, \Delta']$ as follows. Its incidence matrix is given by all rows from M, that do not represent a simplex which is a subsimplex of a simplex from a row of M', and vice versa.

To illustrate this construction consider the following example. Let Δ_1 and Δ_2 be given by the matrices M_1 and M_2, respectively:

$$M_1 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix}.$$

Then the complex $[\Delta_1, \Delta_2]$ is given by the incidence matrix

$$\begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix}.$$

We now apply this construction to distance measures on time series of graphs. Let $\Delta_1, \dots, \Delta_n$ be the simplicial complexes associated to the time series $G_1, \dots, G_n$ by some method. We form the series of consecutive differences

$$\Delta_{12} = [\Delta_1, \Delta_2], \Delta_{23} = [\Delta_2, \Delta_3], \dots, \Delta_{n-1,n} = [\Delta_{n-1}, \Delta_n],$$

$$\Delta_{123} = [\Delta_{12}, \Delta_{23}], \dots,$$
$$\dots$$
$$\Delta_{123\dots n}.$$

One can now apply the various vector valued graph measures to these difference complexes. It is interesting to compare the difference between the M-distance of two successive complexes and the M-measure of the difference complex.

A somewhat different approach one can take to a time series of digraphs is to view it not as a sequence of independent digraphs, but rather as a transformation of a digraph in time. If one adopts this point of view, then one is led to the notion of composition of digraphs in a time series. To explain this concept it is most convenient to represent a digraph in a different way.

Instead of the usual representation of a digraph G with n vertices we use the following diagram. We arrange two rows of n vertices, one above the other, and labeled as in G. Then we connect a top vertex v with a bottom vertex u by an edge if and only if there is a directed edge in G from v to u. The resulting structure can be interpreted as the Hasse diagram of a partially ordered set (poset). This representation is suggestive of interpreting a directed graph as a flow in time among nodes. This should not be confused with the flow in a network, as indicated by weights attached to the edges. Consequently, we want to refer to this partially ordered set as a *flow*. A flow can of course be represented by a binary adjacency matrix.

It is clear that from a flow we can reconstruct the digraph that gave rise to it. Thus, there is a one-to-one correspondence between digraphs and flows. The new represenation has several advantages, aside from being more suggestive. It allows for several forms of *composition* of flows. For instance, we can simply "stack up" two flows, one on top of the other to obtain a flow of length two. More generally, we can stack up any finite number of flows. Given a time series of digraphs of length n, we obtain a "stacked" flow of height n.

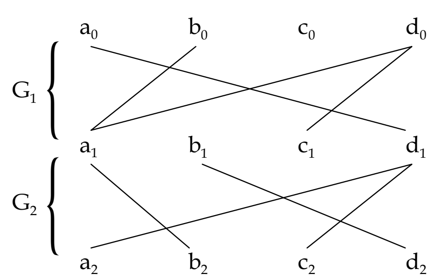

Figure 2: Stacked flow.

Associated to such a stacked flow is another simplicial complex that is potentially of great interest in network analysis. We can interpret a stacked flow S as a poset, whose elements are the nodes at each level, and whose partial order relation is given by the edges between levels. Then we associate to S its *order complex.* It has as vertices the elements of S, and the simplices are given by all totally ordered chains of S. This construction has been studied extensively. For details see, e.g., [6]. We illustrate it with an example. Consider the stack in Figure 2, constructed from the two digraphs in Figure 3.

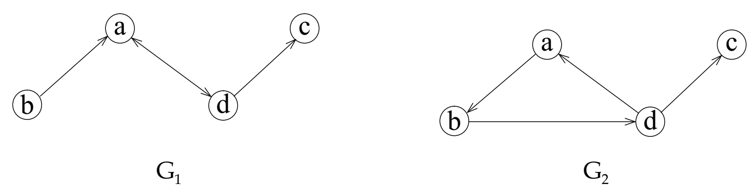

Figure 3: Time series of digraphs.

Its order complex is given in Figure 4.

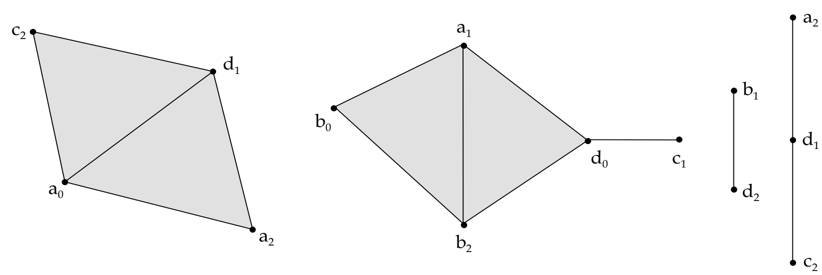

Figure 4: Order complex of stacked flow.

One can now apply various graph measures to the order complex. Stacked flows provide a combinatorial approach to the analysis of the flow of interactions in decision networks [10].

Alternatively, we can compose two flows to obtain another flow by multiplying the corresponding incidence matrices. This amounts to composing two flows by composing all individual components that make up the flow. Composing flows is conceptually easier than composing two directed graphs to obtain a third one. For the stack in Figure 2, the composed flow is given in Figure 5.

Composition of flows or equivalently, digraphs, provides us with yet another measure attached to a time series of digraphs. It will not be a difference measure, but rather a measure of aggregation. Nonetheless,

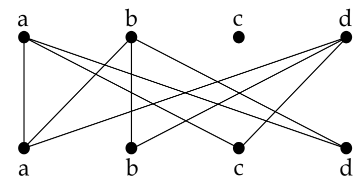

Figure 5: Composed flow.

it might provide a useful global measure of network transformation over time.

8 Measures for Weighted Digraphs

In order to apply our measures to digraphs with weights attached to the edges, we represent a weighted digraph as a sequence of unweighted digraphs as follows. Let G be a weighted digraph with vertices $1, \ldots, n$, and weight $w_{ij} \in \mathbb{R}_{\geq 0}$ attached to directed edge (i, j). Choose a *threshold unit* $c > 0$ in $\mathbb{R}$. If w is the largest weight that occurs, then we generate a sequence of unweighted digraphs by taking G_1 to be the subgraph of G whose edges correspond to those edges of G whose weight v is such that

$$0 \leq v \leq c.$$

Next, define G_2 to have edges corresponding to those edges with weight between c and $2c$. The last graph G_r has edges with weight v such that

$$rc \leq v \leq w.$$

In this way we associate to a weighted digraph G a sequence

$$(G_0, \ldots, G_r)$$

of unweighted digraphs. We can now apply the measures for unweighted digraphs developed earlier to each of the graphs in the sequence. In order to obtain a single vector-valued graph measure, one can aggregate the measures for the sequence of graphs by summing, possibly using a weighted sum, if appropriate. In this way one can directly apply M-distance as a graph distance measure to weighted digraphs. That is, if $W = \{w_0, w_1, \ldots\}$ is a sequence of weights, we define

$$M^W(G) = M^W(G_0, \ldots, G_r) = \sum_{i=0}^{r} w_i M(G_i).$$

9 An Example

In this section we carry out the calculation of all the measures defined so far on the time series of the two digraphs given in Figure 6, in order to illustrate the various concepts we have introduced, their similarities and differences.

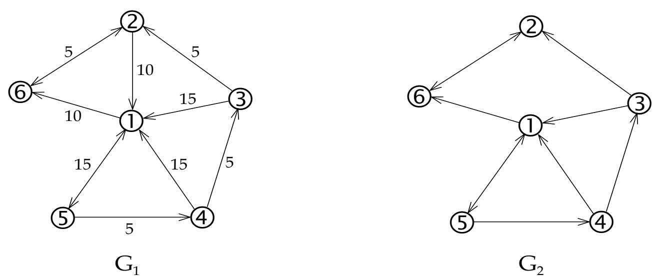

Figure 6: Time series of digraphs.

First we disregard the weights on the first digraph. Their neighborhood complexes and complete subgraph complexes are given in Figure 7.

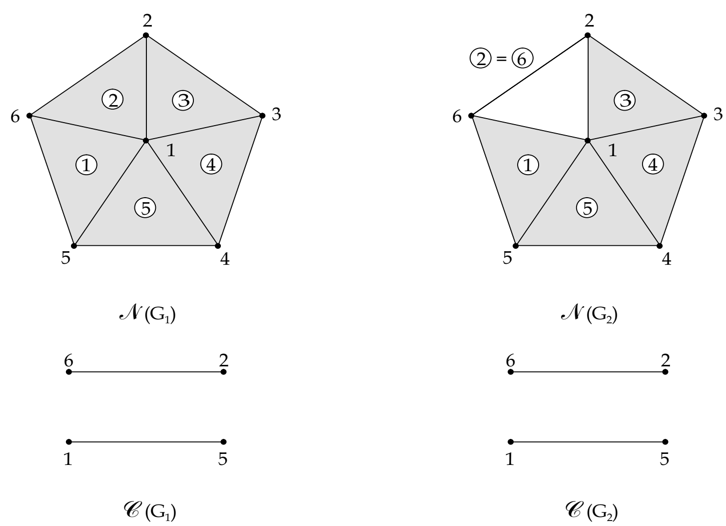

Figure 7: Neighborhood and complete subgraph complexes.

We compute the vector-valued measures for $\mathfrak{N}(G_i)$ as follows:

$$\begin{array}{ll} f_1 = (6,10,5), & f_2 = (6,10,4); \\ Q_1 = (1,1,5), & Q_2 = (1,2,4); \\ A_1 = (0,1,0), & A_2 = (0,0,0); \\ H_1 = (1,0,0), & H_2 = (1,1,0); \\ B_1 = (1,5,5,1), & B_2 = (1,6,8,4,1). \end{array}$$

The computations of the f- and Q-vector are straightforward. To compute the A-vector, observe that the 1-loop given by the sequence of 2-simplices

$$(1) \to (2) \to (3) \to (4) \to (5) \to (1)$$

generates a nontrivial element of $A_1^1(\mathfrak{N}(G_1),(1))$. Now, as a topological space, $\mathfrak{N}(G_1)$ is a disk, hence contractible to a point. On the other hand, $\mathfrak{N}(G_2)$ is homotopic to a circle, so that its first homology group is isomorphic to $\mathbb{Z}$. Finally, the B-vector was computed using the computer algebra system *Macaulay2* [13].

The complexes $\mathfrak{C}(G_i)$ are substantially less interesting, and the computation of the vector-valued invariants is left to the reader as an exercise.

We now compute the following graph distance measures:

$d_f(G_1,G_2) = 1,\ d_Q(G_1,G_2) = 2,\ d_A(G_1,G_2) = 1,\ d_H(G_1,G_2) = 1,$
$d_B(G_1,G_2) = 20.$

Finally, consider G_1 as a weighted graph, as in Figure 6. Using a threshold unit of 5, we obtain the sequence of digraphs given in Figure 8. The first graph reflects edges with weights $0 \le w \le 5$, the second one weights $0 \le w \le 10$, and, finally, the third one which contains all edges, reflects all weights.

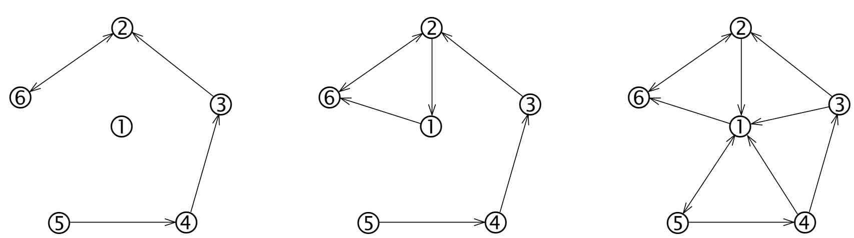

Figure 8: Representation of weighted digraph.

As before, one can now apply the above graph measures to this sequence of unweighted graphs.

10 Conclusion

In this paper a number of novel approaches to change detection of time series of communications networks are described. They comprise combinatorial and algebraic invariants viewed as distance (similarity/disimilarity) measures between weighted and unweighted graphs and digraphs. Various ways of assignment of simplicial complexes to graphs (and time series of digraphs) are explored. Their combinatorial, topological, and algebraic properties are used to characterize the time series behavior of the networks.

While a full investigation of these measures in experiments is still pending, it is clear that our approach opens up several rich mathematical areas as a source of graph distance measures and measures on time series. These measures focus primarily on the qualitative properties of the graphs rather than on their combinatorial differences. We are optimistic that our approach will lead to a fruitful new area of network analysis.

References

[1] R. Atkin, From Cohomology in Physics to q-connectivity in Social Science, *Int. J. Man-Machine Studies* **4**, 139–167, 1972.

[2] R. Atkin, An Algebra for Patterns on a Complex, I, *Int. J. Man-Machine Studies* **6**, 285–307, 1974.

[3] R. Atkin, An Algebra for Patterns on a Complex, II, *Int. J. Man-Machine Studies* **8**, 483–498, 1976.

[4] R. Atkin, *Combinatorial Connectivities in Social Systems*, Birkhäuser Verlag, Boston, 1977.

[5] H. Barcelo, X. Kramer, R. Laubenbacher, and C. Weaver, Foundations of a Connectivity Theory for Simplicial Complexes, *Adv. in Appl. Math.*, to appear.

[6] A. Björner, Topological Methods, in R. Graham, M. Grötschel, and L. Lovász (eds.), *Handbook of Combinatorics, II*, 1819–1872, Elsevier Science B. V., 1995.

[7] H. Bunke and A. Kandel, Mean and Maximum Common Subgraph of Two Graphs, in preparation, 2000.

[8] C. J. Colbourn, *The Combinatorics of Network Reliability*, Oxford University Press, Oxford, 1987.

[9] M. Coombs, O. Colon-Reyes, J. Lakey, R. Laubenbacher, and B. Stigler, The Combinatorial Detection of Local Control Mechanisms in Decision Networks, in preparation, 2000.

[10] M. Coombs, A. Jarrah, and R. Laubenbacher, A Foundation of Combinatorial Time Series Analysis, preprint, 2000.

[11] M. Coombs, R. Laubenbacher, and A. Taha, An Application of q-analysis as a Tool for Detecting and Evaluating the Presence of Information Warfare, Technical Report NMMP-97-02, Physical Science Laboratory, New Mexico State University, 1997.

[12] M.E. Gaston, M. Kraetzl, R. Laubenbacher, and B. Stigler, Anomalous Change Detection in Computer Networks Using Combinatorial and Algebraic Measures, in preparation, 2001.

[13] D. Grayson and M. Stillman, *Macaulay2*, available at `www.math.uiuc.edu/Macaulay2`.

[14] J. Johnson, *The Mathematics of Complex Systems* (ed.), Oxford University Press, Oxford, 1991.

[15] J. Johnson, The q-analysis of Road Intersections, *Int. J. Man-Machine Studies* **8**, 531–548, 1976.

[16] X. H. Kramer and R. Laubenbacher, Combinatorial Homotopy of Simplicial Complexes and Complex Information Networks, in C. Cox and B. Sturmfels (eds.), *Applications of Computational Algebraic Geometry*, Proc. Symposia in Appl. Math., Amer. Math. Soc., Providence, RI, 1997.

[17] R. Laubenbacher, X. H. Kramer, and M. Coombs, Algebraic Invariants of Simplicial Complexes and Analysis of Complex Information Networks, in M. Coombs and M. Sulcoski (eds.), *Control Mechanisms for Complex Systems*, to appear.

[18] P. J. Shoubridge, M. Kraetzl, W. D. Wallis, and H. Bunke, Detection of Abnormal Change in a Time Series of Graphs, submitted to Journal of Interconnection Networks, 2000.

[19] R. Stanley, *Combinatorics and Commutative Algebra*, 2nd ed., Birkhäuser Verlag, Boston, 1996.

[20] J. W. Vick, *Homology Theory. An Introduction to Algebraic Topology*, 2nd ed., Springer-Verlag, New York, 1994.