Graph Isomorphism and Genotypical Houses

Ruth Conroy Dalton
r.conroy-dalton@ucl.ac.uk
University College London

Ciler Kirsan
c.kirsan@ucl.ac.uk
University College London

Abstract

This paper will introduce a new method, known as small graph matching, and demonstrate how it may be used to determine the genotype signature of a sample of buildings. First, the origins of the method and its relationship to other 'similarity' testing techniques will be discussed. Then the range of possible actions and transformations will be established through the creation of a set of rules. Next, in order to fully explain this method, a technique of normalizing the similarity measure is presented in order to permit the comparison of graphs of differing magnitude[1]. The last stage of this method is presented, this being the comparison of all possible graph-pairs within a given sample and the mean-distance calculated for all individual graphs. This results in the identification of a genotype signature[2]. Finally, this paper presents an empirical application of this method and shows how effective it is, not only for the identification of a building genotype, but also for assessing the homogeneity of a sample or sub-samples.

Introduction to small graph matching

It was demonstrated in earlier work by Conroy Dalton (Conroy 2001) that a technique drawn from mathematics/information theory termed string matching could be brought to bear upon an architectural problem (that of characterizing paths and samples of paths through an environment). Essentially, this technique comprised of comparing any pair of paths to determine their similarity. Recently, this method was applied successfully to analyze paths taken through a museum in a NSF[3]-funded project (Wineman et al. 2002; Peponis, Conroy Dalton, and Wineman 2003). Another graph theoretic method that was inspired by string matching is graph matching, a relatively new technique still unfamiliar to many mathematicians and graph theorists. Graph matching enables the comparison of any pair of small, labeled, directed graphs by determining how similar they are. It achieves this by calculating how many operations are needed to fully transform one graph

---

[1] In this paper, the *magnitude* of a graph is defined as the *order* of a graph (or the number of vertices) plus the *size* of a graph (the number of edges). *Magnitude*, $(M)G = (V)G+(E)G$. The authors of this paper have been unable to determine if such a term exists in graph theory and have, therefore, defined this term for the purposes of explaining this method.

[2] The distinction between a genotype and a *genotype signature* is a distinction drawn by Bafna, in our personal communication; since a genotype is a purely theoretic construct; it is not realizable, whereas a *genotype signature* is an evidence-based example deemed to most denotative of its theoretic genotype.

[3] National Science Foundation Informal Science Education Grant, number 9911829; the principal investigators were Peponis and Wineman.

into another. The kinds of transformations that will be discussed in this paper are the insertion or deletion of a vertex, edge or vertex-edge dyad[4] in a graph and vertex/edge/vertex-edge dyad substitution — for example, an edge that connects vertex A to vertex B, {A,B} may be substituted by an edge connecting vertex A to vertex C, {A,C}.

Partly due to the relative newness of this mathematical method and partly due to its complexity[5], there is some debate about the methods and algorithms that may be used in graph matching. For example, the act of a vertex or edge substitution can be held to be equivalent to a vertex or edge deletion followed by a vertex or edge insertion. The 'weight', 'cost' or 'penalty' for such an act could equally be argued to be one (it is only one operation) or two (it is two sequential operations). Equally, were a vertex to be deleted which is linked to the rest of the graph by an edge, is its associated edge also considered to have been *implicitly* deleted, since an edge can not exist independently? And what should the weight of this action be? Not only are the relative weights of such operations still to be established but also, due to the complexity of the problem, the algorithms used to perform graph matching vary considerably with many approximations and heuristics being used, such as genetic search algorithms and simulated annealing (Jiang, Munger and Bunke 1999; Jiang Munger and Bunke 2001; Bunke and Gunter 2001). However, given such fundamental disparities in application, it is, nevertheless, a technique that is beginning to be used successfully in many applied areas such as chemistry (for molecule comparison) and computer vision (for object recognition and shape analysis).

In this paper, it is proposed that graph matching may be used to compare graph representations of simple building types, for example house plans. House plans are a particularly challenging area for study, for as Hanson explains:

> *Real houses are a complex expression of the social and individual worlds of their occupants, in which social structure and convention seems inextricably bound up with the idiosyncratic, whimsical, arbitrary or even chaotic circumstances of people's everyday lives. That this is so nearly always poses problems in understanding and interpreting the hidden order in houses and homes.* (Hanson 1998, 269)

It is suggested that the methods of graph matching might be particularly suitable for addressing this problem. Furthermore, this paper will demonstrate how basic methods of graph matching may be extended to produce the calculation of a *median graph* in order to determine the genotype or genotype signature of a sample of small graphs.

---

[4] In this paper, we are using the term *vertex-edge dyad* to indicate a vertex of degree 1 plus its incident edge, with the two elements being considered a single entity. This is also known as a *pendant vertex* (plus incident edge) in graph theory and a *dangle vertex* (plus incident edge) in GIS.

[5] Graph matching is suspected to be neither a *P* nor an *NP complete* problem (although certainly it is *NP*). A problem is said to be Nondeterministically Polynomial (NP) if we can find a nondeterministic Turing machine that can solve the problem in a polynomial number of nondeterministic moves. That is to say, a problem is NP complete if its solution comes from a finite set of possibilities, and it takes polynomial time to verify the correctness of a candidate solution.

The concept of the genotype, particularly with respect to domestic building, has been established in space syntax literature (Hillier and Hanson 1984, 12-13, 143, 154; Hillier, Hanson and Graham 1987; Hillier 1996, 36, 83, 249-250, 429-431; Hanson 1998, 32, 269-270). In particular, Hanson provides the following, useful definition:

> *Function thus acquired a spatial expression which could also be assigned a numerical value. Where these numerical differences were in a consistent order across a sample of plans from a region, society or ethnic grouping then we could say that a cultural pattern existed, one which could be detected in the configuration itself rather than in the way in which it was interpreted by minds. We called this particular type of numerical consistency in spatial patterning a housing 'genotype'.* (Hanson 1998, 32)

The authors mentioned above, argued that the existence of a consistent inequality genotype across a sample of houses points to a cultural genotype and that the type of analysis to identify such a cultural genotype is both functional and spatial. Once this has been detected they argued it is possible to detect imprint of social relations and even of the human mind on buildings, (Hillier, Hanson and Peponis 1984, 61-72).

This definition of the genotype has most recently been elaborated by Bafna (Bafna, 2001), as *"The sociologically important structure of spatial configuration – it is the part of the design that the architect cannot tamper with, if the design is to function as a working house."* By this definition, a building genotype or rather a *genotype signature* would be the graph of a building that is most similar, on average, to all other graphs in a given sample[6]. It would then be possible to measure the variance between any single building phenotype and the genotypic graph by measuring the 'distance' (in graph matching terms) between them. Although the concepts of genotypes and phenotypes have been theorized in the context of building typologies, actual phenotypic graphs of a single building genotype may vary to such an extent, even at the level of the rank order of the integration values of their spaces, that in, practice, the genotypic signature of buildings seems very difficult to determine. As well as providing mathematical clarity and a reproducible method to the hitherto theoretic concept of genotype and phenotype, with respect to building types, this method also facilitates a discussion of the homogeneity of a sample of buildings through the calculation of the average distance of all graphs each from one another. This provides a powerful tool for the study of precedent building types. This paper will include an empirical section examining small sample of Greek and Turkish houses from Cyprus, analyzed comparatively using 'graph-matching' techniques in order to fully illustrate this method.

---

[6] This is only true, given a high degree of homogeneity of the sample; a more heterogeneous sample could be indicative of the presence of more than one genotype/genotype signature.

From String Matching to Small Graph Matching

How the ideas of string matching may be applied to small graphs

A string is simply a sequential (and hence one-dimensional) list of text-characters, for example, S-T-R-I-N-G. Any string will have various properties: it has a length (in this case, six characters), it has a beginning character (in this case, S), it has an end (in this case, G), if of sufficient length, it may contain a number of sub-strings (e.g. R-I-N-G or T-R-I), it may be reversed G-N-I-R-T-S and finally, it may be transformed into another string. For example, if the character I in the string S-T-R-I-N-G is substituted for the letter O, the string is transformed into the word S-T-R-O-N-G.

According to Stephen, in his report on string matching (Stephen 1992), the essential problem behind string matching may be generalized as being one of "*Find[ing] the patterns which occur more than once in a given sequence of symbols*." However, more generalized applications of string matching quickly lead to the family of methods known as *string distance metrics*, which are usually employed to determine the similarity between any candidate pair of strings. The degree of similarity is equivalent to the number of transformations required to fully transform one string into another. The act of transformation is often considered to be a *cost*, and the purpose of all string distance algorithms is to calculate the *least cost* way of transforming one string into another. On a string, these acts consist of character deletion, character insertion and character substitution. Once the range of permissible acts has been determined, the weight or cost for each act must be established and, it should be noted that the application of cost may not be uniform.

Now, in order to compare the similarity of graphs rather than strings, we must simply exchange the elements upon which the actions are performed. Strings only contain one kind of element, the character, and a string is one-dimensional. Graphs consist of two kinds of element: vertices (or nodes) and edges (or arcs). In addition, a graph may be two-dimensional (a planar graph) or even three-dimensional (a non-planar graph). However, given the apparent disparity between string-elements and graph-elements, the range of actions is surprisingly the same; borrowing directly from string-matching, we may apply acts of deletion, insertion or substitution to these graph elements. This is the essence of string matching as applied to small graphs. Although this seems quite straightforward in principle, there are a few caveats; given the differences between graphs and strings a few rules and conditions need to be clarified in order to render the method transferable. These are set down in the next section.

Conditional use of labeled graphs

The primary requirement for the application of this method is the use of labeled graphs. The use of labels in these small graphs permits the problem of small graph matching to be computable. Without labels, there could potentially exist many possible solutions. Therefore, one view of the act of graph labeling is that it is about reducing the solution search space. Computational arguments aside, the use of labeled graphs does seem crucial from an architectural standpoint. For example, it would be possible to have two buildings which were identical from the point of view of their graphs but fundamentally different from a functional perspective. In this case, rooms or spaces which were highly integrated in one building might be quite segregated in the other and vice versa. We would clearly feel that these buildings were essentially different to each other. This view is substantiated and elaborated by Hanson in her discussion of building types:

> *If we now consider labels, or more precisely the relation of the various labels within the space configurations [of four courtyard buildings]…, we can begin to detect certain regularities in terms of the relations between syntactic positions within the complex and way in which labels are assigned to spaces. For example, the space labeled A is always as shallow as any other in the complex, whereas space B is always as deep as it is possible to go from the outside. Space D is always on a ring, except where there are no rings in which case there is no space D. Space E is always on the shortest path from A to B. Finally the position of C is randomized. Since this is the only space which is so, this may well be considered significant.* (Hanson 1998, 31)

So, not only are labels necessary in order for the graph matching method to be computable but they are fundamental to the very concept of the building genotype. This has been established clearly in the earlier work by Hillier and Hanson, The Social Logic of Space, where authors argued that labels are more significant in gamma[7] than in alpha and that, '*…a genotype in gamma can be identified in terms of associations between labels of spaces and differentiations in how those spaces relate to the complex as a whole, in terms of the syntactic dimensions'*. (Hillier and Hanson 1984, 154)

The only other consideration, with respect to labeled graphs, which needs to be clarified in this paper is when instances of multiple labels (i.e. vertices sharing the same label) are present in a graph, such as is common in domestic architecture (multiple bedrooms, for example), this need present no problem. In order to accommodate such cases, a maxim of *comprehensive interchangeability* may be applied. This is to say, any two vertices or edges bearing the same labels may be considered to be fully interchangeable. However, in a few extreme cases, this may necessitate some internal rearrangement in order to complete the transformations, See the next section on exceptions to the rules for an example of this.

Rules governing the range of possible actions/transformations

The range of possible actions and transformations is best illustrated through the definition of a set of rules. These may be defined as follows:

---

[7] Gamma and Alpha analyses are the types of syntactical analyses provided for building interiors and settlements respectively which had been introduced in The Social Logic of Space.

Deletions and Insertions (InDels)

1. Any edge may be deleted, *on condition* that it does not "orphan[8]" a pre-existing vertex. (I.e. an edge that is incident upon a vertex of degree 1, also known as a *pendant* vertex).
2. Any edge may be inserted *on condition* that it must connect exactly two[9] *pre-existing* vertices.
3. It is not possible to delete a vertex on its own; the vertex-edge dyad (the *pendant* vertex and its incident edge) must be deleted as a single unit.
4. It is not possible to insert a vertex on its own; the vertex-edge dyad (the *pendant* vertex and its incident edge) must be inserted as a single unit.

Without these rules, it could be possible to create *disconnected* graphs as intermediate steps in the graph matching process. This method is based on the premise that all actions should result in viable-graphs (that is to say, viable from the point of view of the real-world building that the graph is representing).

Substitutions

5. An edge may be substituted *on condition* that there is a common vertex shared by the two edges being substituted. In other words, if the two edges are *adjacent*. This could be interpreted as edge 'reassignment'.
6. A vertex may be substituted; since we are using labeled graphs, a vertex substitution simply equates to an act of re-labeling a vertex. A single substitution is written $A \rightarrow B$. A two-way substitution, or *swap*, is written $A \leftrightarrow B$.
7. A vertex-edge dyad may not be substituted by another vertex-edge dyad. The reason for this is that there is no common structure to restrict the process (as with edge substitution). The act of vertex-edge dyad substitution may be achieved by combining edge substitution followed by vertex substitution (or vice versa). This is clearly two actions and not one single action.

Combining transformations

8. In cases where a lesser number of actions (or greater number of actions at a lower cumulative cost) can be combined to produce the equivalent final transformation in comparison to a differing (but greater) set of actions (producing the same result), then the weight must always be dictated by the lower number.

Exceptions: zero weighting of substitutions

---

[8] Another way of saying this is that an edge may be deleted on condition that it does not cause a degree of vertex 0 (or an *isolate*) to be created.

[9] Less than two edges would contradict the essential definition of an edge, more would constitute a *hyperedge*, which would have no real-world meaning in this example.

There are some extreme cases where actions, which would otherwise be assigned a weight of one or two, are given a weight of zero. These occur only in certain cases, where graphs contain multiple vertices which share the same label, and where it may be necessary to 'reassign' an edge (or vertex-edge dyad) from one instance of a label to another. In such cases, seemingly strange actions such as {A,C} → {A,C} may arise and, if so, should be given a weight of *zero* not one (as per the rules above). This may occasionally occur in particular cases of edge or vertex-edge dyad substitutions.

Summary

These rules can be summarized in Table 1, where permitted actions are indicated by a tick and excluded actions shown with a cross.

		Graph Elements		
		Vertex	Edge	Vertex-Edge Dyad
Actions	**Del**etion	✗	✓	✓
	Insertion	✗	✓	✓
	Substitution	✓	✓	✗

Table 1 Range of permissible transformations

It is clear from the eight rules (and one exception) written above that order and conditions are important in making any graph transformations. These will be discussed in the next section. Equally, it is not enough to establish what actions may be performed on a graph, it is also necessary to determine what weight or value will be assigned to each transformation. This is best done through a process of considering in detail, what each of these possible actions could mean, at the level of the physical building, rather than at the abstraction of the graph.

Normalizing the graph 'distance' measure, why and how?

The concept of graph comparison normalization

The concept of graph normalization arises from Conroy Dalton's work on string matching (Conroy 2001). In her review of string-matching techniques, she realized that many techniques were restricted in their application to the comparison of strings of the same length, for example, Hamming distance. In her subsequent work, she allowed for differences in lengths of strings by normalizing the number of transformations necessary to fully transform one string into another by dividing this number by the average length of the two strings being compared.

Without graph-matching normalization a comparison of graphs of different ‘sizes’ would equally be impossible. In string matching, the elements being transformed were text-characters and the total length of the string was used for the normalization process. However, in graph matching both vertices and edges may be transformed. Therefore, any comparison between two small graphs should be normalized by the average number of vertices plus the average number of edges per graph. The number of vertices in a graph is known as a graph’s *order* while the number of edges is known as a graph’s *size*. We propose that the sum of a graph’s *order* and *size* be referred, in this paper, as a graph’s *magnitude* and that the mean magnitude of any pair of graphs be used to normalize the transformation measure.

Given two graphs, G_1 and G_2, the magnitude, M of G_1 and G_2 may be defined as follows,

Equation 1

$$M(G_1) = |V(G_1)| + |E(G_1)|$$
$$M(G_2) = |V(G_2)| + |E(G_2)|$$

&

Equation 2

$d(G_1, G_2) =$ the *distance metric* between two graphs, G_1 and G_2.

Obviously, $d(G_1,G_2)$ is equivalent to the number of transformation required to fully transform one graph into another.

Then the *normalized distance*, $\hat{d}$, between G_1 and G_2 can be written as,

Equation 3

$$\hat{d}(G_1, G_2) = \left(\frac{2d(G_1, G_2)}{M(G_1) + M(G_2)} \right)$$

This is a distance metric, between any two graphs, which normalizes the distance through the average magnitude of the two graphs. This normalization has been tested with a variety of examples and can be applied to all cases, even the case of comparing a null graph (a graph with no vertices or edges) into a non-null graph, should this be relevant.

Towards a concept of the genotype graph or an all-pairs comparison

The mean, normalized distance measure

The final stage in using this method of calculating a normalized distance metric between any two graphs is to take the radical step of calculating the distance metric between all possible pairs of graphs. If we have a small sample of graphs, say G_1 to G_n, with the total number of graphs being n, then the number of comparisons that can be made are $n(n-1)/2$ (since every graph will be identical to itself and graph comparisons are reflexive). If every graph is compared to every other graph in the sample of n, and the mean, normalized distance found between all pairs of graphs, then each graph will have a single value, its mean, normalized distance, $\bar{\hat{d}}$, which represents how similar it is, on average, to the whole of the rest of the sample. This value can be expressed mathematically as,

Equation 4

$$\bar{\hat{d}} \equiv \sum_{i,j=1}^{n} \left(\frac{2\hat{d}_{i,j}}{n(n-1)} \right)$$

What information does an all-pairs comparison provide?

How may the mean, normalized distance measure of a graph be interpreted? First, let us rename it; from this point onward, in this paper, the mean, normalized distance of a graph shall be referred to as its *distinction*. If the *distinction* of a graph is low, then this means that, on average, fewer acts of transformation are required to turn it into any other graph in the sample. Note that this is a one-to-many relationship, no longer a one-to-one. Equally if the *distinction* of a graph is towards the lower end of the range of values for a given sample, this means that it is approaching the *genotype signature*. The *genotype signature* would be the graph with the lowest distinction value of a sample. This precisely matches Bafna’s assertion that the genotype signature should be an evidence-based entity since the genotype itself is a purely theoretic construct. The graph with the highest *distinction* value of a sample would the graph which differs most greatly from all others, that is too say, it is the graph (or in this case, the building) which is most idiosyncratic.

The *homogeneity* of a sample can be defined as the standard deviation[10] of *distinction* values, see Equation 5; the smaller the range of values the more homogenous the sample and the higher the values the more heterogeneous the sample.

Equation 5

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

---

[10] In many texts this is referred to as *standard error* as it represents the standard deviation of a sample of means not a sample of individual values.

Where $x_1, x_2... x_n$, are the set of *distinction* values for a sample of buildings with a mean value, $\bar{x}$ and where n is the total number of buildings in the sample.

Now that all the parts of the small graph matching method have been defined, the next section of this paper presents an example application of this method to a sample of houses.

Application of the method to a small sample of graphs: Greek and Turkish Cypriot houses

Morphological studies have shown that one of the most conspicuous ways in which cultures differ from one another is through different forms of spatial organization, (Hillier et al. 1987; Hillier and Hanson 1984; Hanson 1998). Numerous studies which have been conducted recently by using space syntax theory and its analytical tools suggest that it is likely to expect culture and ethnicity to manifest itself through one or two dominant syntactic 'genotypes' at any time and with variation associated with the passage of time. In this respect we would expect the *homogeneity* to be higher across the ethnic groups and lower within the group.

Description of the house sample

In order to illustrate these theoretical and methodological ideas a small sample of Greek and Turkish traditional Cypriot houses have been analyzed comparatively to explore their cultural differences. The sample of house plans analyzed date from the period before 1974, which marks the division of Cyprus into two ethnically diverse areas and the resultant exchange of populations and of their houses. The sample has been constructed in a way so as to allow cross-ethnic comparisons. It comprises of six houses, three of which have been selected from a village, which used to be Greek before 1974 and three from a Turkish one; both villages are located in the Mesarion region of the Island. The spatial layouts have been reconstructed from the present situation of the houses through first hand observations and fieldwork[11].

The houses are all 'courtyard houses', essentially driven by a peasantry-based agricultural economy and way of life. Besides their main domestic functions, they serve as a 'workshop' or 'laboratory' due to the abundance of work-related functions (Christodoulou 1959). Regardless of their Greek or Turkish origins, both groups' houses are formed by similar spatial elements as; 'courtyards', multi-functional living spaces referred as 'rooms' and 'main rooms', multi-functional transitional spaces as 'central hallways' and 'loggias', secondary spaces as 'kitchens' and storage facilities for animals and goods[12]. They follow a simple rectangular arrangement developed around an inner courtyard and are one or two storey-high.

---

[11] Observations and fieldwork were all conducted by Ciler Kirsan between April 2000 and August 2000 as part of her PhD studies.

[12] The labeled spaces in the plans and graphs are represented as follows: C = courtyard; c = animal courtyard; L = loggia; R = room; M = main room; K = kitchen; A = animal shed; E = exterior; m = upper main room; S = semi-closed central space/hallway; H = closed

Figure 1 shows the six plan layouts and their corresponding justified access graphs. The houses have been named after their ethnic identities and named one through to three, with G representing Greek houses and T the Turkish ones.

T1 G1

T2 G2

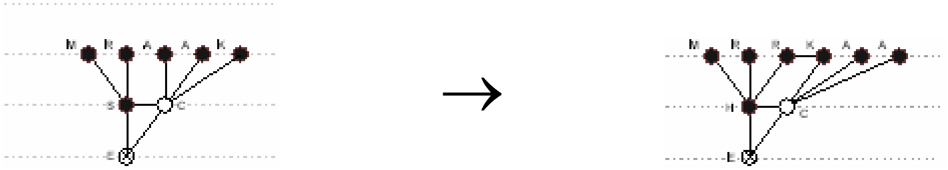
T3 G3

Figure 1 House plans

---

central space/hallway; P = produce store; T = straw store; h = upper closed central space/hallway; s = upper semi-closed central space/hallway; r = upper room.

Simple worked example: house G2 to house G3



G2 to G3

Figure 2 Graphs G2 and G3

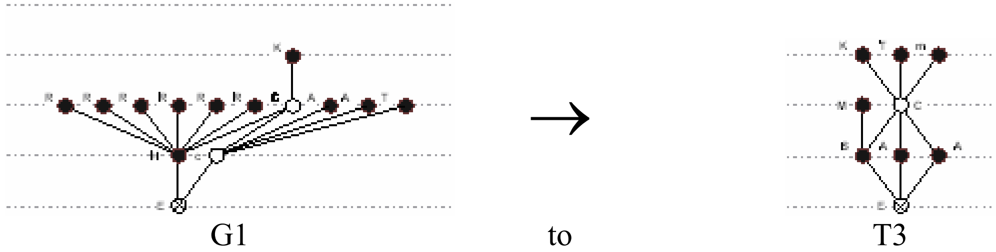
Transformation	*Weight*
S → H	1
{} → R{R,K)	1
{} → {R,H}	1
Total	**3**

Table 2 Sequence of transformations (house G2 to house G3)

In this example, we are comparing two of the Greek houses, G2 to G3 (see Figure 2). In order to transform G2 into G3, we need to perform the following operations: The vertex that is labeled S in house G2 must be relabeled as vertex H. A new pendant vertex, R, and its adjacent edge, {R,K} must be inserted into G2. Finally, a new edge, {R,H} must be inserted. All of these transformations carry the weight of one and therefore the total cost of transforming house G2 into G3 (or vice versa) is three.

The *size* of G2 is 8, the *order* is 8 and therefore the *magnitude* of G2 is 16.
The *magnitude* of G3 is 19. The mean magnitude is, therefore, $(16+19)/2$, which is 17.5.
The mean normalized distance, or *distinction*, between G3 and G2 is 3/17.5 = 0.17.

Complex worked example: house G1 to house T3



G1 to T3

Figure 3 Graphs G1 and T3

Transformation	*Weight*
({A,c} → {A,C})*2	1*2
H → S	1
R → M	1
{} → {E,A}	1
{T,c} → {T,C}	1
(R{R,S} → {})*5	1*5
c → m	1
{E,m} → {E,A}	1
Total	**13**

Table 3 Sequence of transformations (house G1 to house T3)

In the example illustrated in Figure 3, we are attempting to transform a Greek house, G1, into a Turkish house, T3. This is a particularly complex set of transformations, which must be performed in strict order. We have four instances of edge substitution, three cases of vertex substitution, one edge insertion and five example of vertex-edge dyad deletion (the multiple instances of vertices R and their adjacent edges). These transformations are listed, using graph notation, in Table 3. It can be seen from this table, that thirteen separate acts of transformation are required to turn house G1 into house T3. This is far more than the previous example, which required only three.

The *size* of G1 is 14, the *order* is 14 and therefore the *magnitude* of G1 is 28.
The *magnitude* of T3 is 19. The mean magnitude is, therefore, $(28+19)/2$, which is 23.5.
The mean normalized distance, or *distinction,* between G1 and T3 is 13/23.5 = 0.55.

If we calculate the numbers of acts of transformation required to transform each of the six houses (G1, G2, G3, T1, T2 & T3) into every other house in the sample, these values are shown in the upper-left section of each cell in Table 4. Note that the values of the leading diagonal are all zero, as a house needs no acts of transformation to be turned into itself.

House	G1	G2	G3	T1	T2	T3	$\bar{\hat{d}}$
G1	0 28.0 **0.00**	11 22.0 **0.50**	10 23.5 **0.43**	15 22.5 **0.62**	14 20.5 **0.68**	13 23.5 **0.55**	0.45
G2	11 22.0 **0.50**	0 16.0 **0.00**	3 17.5 **0.17**	3 16.5 **0.18**	5 14.5 **0.34**	5 17.5 **0.29**	0.26
G3	10 23.5 **0.43**	3 17.5 **0.17**	0 19.0 **0.00**	5 18.0 **0.28**	7 16.0 **0.44**	7 19.0 **0.37**	0.28
T1	15 22.5 **0.62**	3 16.5 **0.18**	5 18.0 **0.28**	0 17.0 **0.00**	4 15.0 **0.27**	6 18.0 **0.33**	0.31
T2	14 20.5 **0.68**	5 14.5 **0.34**	7 16.0 **0.44**	4 15.0 **0.27**	0 13.0 **0.00**	8 16.0 **0.50**	0.40
T3	13 23.5 **0.55**	5 17.5 **0.29**	7 19.0 **0.37**	6 18.0 **0.33**	8 16.0 **0.50**	0 19.0 **0.00**	0.35

Table 4 Combined transformation and normalization matrix

Given the method, as set out in the previous section of the paper, it now becomes necessary to normalize these values using the mean magnitudes of the graph-pairs (Equation 3); these are noted in the upper-right portion of the cells in Table 4. By dividing the number of transformations by the mean-magnitude (upper-left value by upper-right value), the normalized, distance value for each pair of houses is calculated. These values are indicated in bold in Table 4. Again, the values of the leading diagonal remain zero. Finally, in order to calculate the *distinction* value for each house, the mean, normalized distance must be calculated for each graph, see Equation 4. For each row, the values are summed and divided by (*n-1*); this produces the *distinction* values in the final column of the table.

It can be seen from these *distinction* values, that the house with the lowest *distinction*, namely the *genotype signature* of this sample, is G2 (see Figure 1 for the plans and graphs). The *distinction* value of 0.26 means that, on average, 0.26 acts of transformation per graph element (vertex or edge) are needed to turn this house into any of the other five houses. Another way of looking at this is that, of every three graph elements, two will remain unchanged after any act of transformation required to turn this house into any of the others. Equally, the house with the greatest *distinction* is G1. Its *distinction* value of 0.45 is almost double that of G2. For every single act needed to transform G2, two such acts would be needed to transform G1. Our interpretation of this result is that G1 is the most idiosyncratic house in this sample. If you refer to the plans and graphs in Figure 1, this does appear to be intuitively correct. The house, G1, is quite different from the others; it has a double courtyard, not a single courtyard, is slightly deeper than average, with an extremely high proportion of pendant vertices.

The *homogeneity* of the whole sample is 0.07. The *homogeneity* of the Greek houses is 0.03 and the *homogeneity* of the Turkish houses is 0.06 enabling us to conclude that the Greek sample is more homogenous than the Turkish sample. The Greek *genotype signature* is house G3 (albeit from a sample of only three houses) and the Turkish *genotype signature* is house T1 (again, only from a sample of three). The houses with the highest *distinction* values (i.e. the most idiosyncratic) for each ethnic sample are G1 and

T3 respectively. Having considered the properties of each individual sample, let us now examine the differences between the two samples. If, from Table 4, we consider the mean normalized values for each quadrant (representing, the Greek to Greek pairs, the Greek to Turkish pairs and the Turkish to Turkish pairs) and average the mean, normalized distances for each quadrant[13] we can produce the following matrix.

	Greek	Turkish
Greek	0.25	0.44
Turkish	0.44	0.24

Table 5 Greek and Turkish comparisons

This means, that on average, it takes 0.25 acts of transformation, per graph element to transform *any* Greek house into *any other*. Equally, 0.24 transformations are needed to transform any Turkish house into any other Turkish house. However, almost double the number of transformations is needed to transform a Greek house into a Turkish house (0.44). This is an extremely interesting result and fully supports any hypothesis that there exists an identifiable difference between the two ethnic samples.

Discussion and Conclusion

The method known as small graph matching seems to be applicable to a variety of situations and flexible enough to allow for instances of multiple vertex labeling and even null-graph transformation. It results in a single, quantifiable measure, termed *distinction*, which represents the overall similarity of a single house (graph) to a sample of houses. This method was applied to a small sample of six houses, with apparent success. In a small, informal study undertaken, the *distinction* measure of the six houses seemed to match people's subjective intuition. That is to say, the majority of people, when asked to select the most and least distinctive house in the sample, correctly identified those with the smallest and greatest value of distinction. Furthermore, the matrix of transformations between Greek and Turkish houses, which indicates that there is a measurable difference between the two ethnic types of house, is an extremely interesting finding and tentative vindication of the utility of the method. Hanson discusses the growth of new technologies and computer techniques and how, through these, there is the potential for giving *"clearer descriptions of the invariance that exists in large numbers of house plans, search for spatial 'genotypes' and dimension the extent of individual differences within the set."* (Hanson 1998, 270). The authors of this paper feel that the technique of small graph matching fulfils these criteria and provides a highly useful augmentation to existing methods of analysis. However, a cautionary note should be added; the sample upon which this method was first tried was extremely small and can not be deemed statistically significant.

---

[13] In order to avoid *'mean of means'* errors, these values are calculated by summing all of the transformation values in a quadrant and dividing them by all of the mean magnitudes in the quadrant (rather than simply averaging the values).

The major disadvantage of the method is that it is extremely time-consuming and laborious to conduct by hand. Equally, due to the complexity of the transformations, there is a high potential for error and all the work presented in this paper needed to be 'double-checked' at least four or five times. The ideal solution would be to automate the method computationally. Yet, as mentioned in the beginning of the paper, this in itself is not straightforward as it is an *NP/NP complete* problem. However, it may be possible to approximate this method and some research on this has been conducted in other fields. So, given the encouraging results presented in this paper, it is concluded that an attempt to approximate this method should form the next stage of any further research.

Acknowledgements

To Dr. Sonit Bafna for his invaluable contribution of the concept and definition of a genotype signature and to Alasdair Turner for his assistance with the mathematical equations expressed in this paper.

Bibliography

Bafna, S. (1999, 29 March-2 April, 1999) *The Morphology of Early Modernist Residential Plans: Geometry and Genotypical Trends in Mies van der Rohe's Designs.* Paper presented at the Space Syntax - II International Symposium, Brasilia.

Bafna, S. (2001, 7-11 May, 2001). *Geometrical Intuitions of Genotypes.* Paper presented at the Space Syntax - III International Symposium, Atlanta.

Bafna, S. (2001). *A Morphology of Intentions: An Interpretation of Mies Van Der Rohe's Residential Designs.* Unpublished Doctoral, Georgia Institute of Technology, Atlanta.

Bunke, H., & Gunter, S. (2001). Weighted Mean of a Pair of Graphs. *Computing, 67*(3), 209-224.

Conroy, R. A. (2001). *Spatial Navigation in Immersive Virtual Environments.* Unpublished Doctoral, University College London, London.

Christodoulou, D., 1959, *The evolution of the rural land use pattern in Cyprus*, Geographical Publications Limited, Bude, Cornwall, England.

Hamming, R. W. (1980). *Coding and Information Theory*: Prentice Hall, Englewood Cliffs, NJ.

Hanson, J., 1998, *Decoding Homes and Houses*, Cambridge University Press, UK. 61-72

Hillier, B., 1996, *Space is the Machine*, Cambridge University Press, UK.

Hillier B., Hanson J., Peponis J., 1984, *What Do We Mean by Building Function* in *'Design for Building Utilisation'*, Powell J. & Cooper I. (ed.). London.

Hillier, B., Hanson, J., 1984, *The social Logic of Space*, Cambridge University Press, UK.

Hillier, B., Hanson, J., Graham, H., 1987, "Ideas are in Things: an Application of the Space Syntax Method to Discovering House Genotypes", *Environment and Planning: Planning and Design*, 14, pp. 363-385.

Jiang, X., Munger, A., & Bunke, H. (1999). *Computing the Generalized Median of a Set of Graphs.* Paper presented at the Second IAPR Workshop Graph-Based Representations.

Jiang, X., Munger, A., & Bunke, H. (2001). On Median Graphs: Properties, Algorithms, and Applications. *IEEE Trans. Pattern Analysis and Machine Intelligence, 23*(10), 1144-1151.

Peponis, J., Conroy Dalton, R., & Wineman, J. (2003, 17-19 June, 2003). *Path, Theme and Narrative in Open Plan Exhibition Settings.* Paper presented at the Space Syntax - IV International Symposium, London.

Stephen, G. A. (1992). *String Search* (Technical Report No. TR-92-gas-01). Bangor: University College of North Wales.

Wachowski, A., & Wachowski, L. (Writer), & A. Wachowski & L. Wachowski (Director) (1999). The Matrix [Film]: Warner Bros.

Wineman, J., Peponis, J., Conroy Dalton, R., Dalton, N. S. C., Flaningam, T., & Wilson, A. A. (2002). *Exhibition Layout and Visitor Movement in Science Museums*: National Science Foundation.