The Structure of a 3-state Finite Transducer Representation for Prisoner's Dilemma

Jeffrey Tsang

***Abstract*—To facilitate systematic and automated analysis of game playing strategies, the fingerprint, a mathematical technique that generates a functional summary independent of representation, was developed. This study attempts to push the boundaries of full state space investigation, looking at 3-state finite transducers as a representation for playing Prisoner's Dilemma. There are a staggering 23,000 unique strategies in this space, which severely limits the choice of analysis methods. These strategies are fingerprinted and pairwise distances computed, then hierarchical clustering reduces them to a manageable size for further experiments with multidimensional scaling and the mutational connectivity network. Results indicate there are no obvious cutoff scales of structure; mutational distance is not correlated with fingerprint distance; and a level of similarity with past results on smaller state spaces.**

I. Introduction

A simple to understand model for simulating interactions is the mathematical game; the simplest a game could get is a simultaneous, symmetric two-move game, such as Prisoner's Dilemma. To preserve possibilities for complex strategies, we iterate the game, allowing response and counter-response to your opponent. One oft-used way of experimenting is with evolutionary game theory, generating an unlimited stream of arbitrarily convoluted strategies.

A series of papers [4], [3], [5] presented the concept of fingerprinting, which turns the strategies into normal mathematical functions recording the strategy's behaviour against a reference opponent, after which they become easier to handle. This has enabled previously inattainable studies in representational sensitivity [7], [8], evolutionary time and population size [9], the effect of noise [10], [11] among others.

Using the updated model presented in [1], which is a generalization of the one in [4], avoids some problems in the original specification, including discontinuities and some distinguishing power limitations. We can also define a distance on the space of fingerprints, quantifying the notion of "similar" strategies.

Due to the combinatorial explosion, investigations into the entirety of representations is difficult to perform; here we attempt to push the limits one step further by looking into a standard 3-state finite transducer representation. There are 93,312 nominal strategies; even after removing all duplicates there are 23,000 unique strategies in the space: this effectively prevents consideration of any cubic or worse analysis method. We will still look into the global structure imposed by fingerprint distance, as well as the mutational connections within the space.

Jeffrey Tsang is with the Department of Mathematics & Statistics, University of Guelph, Ontario, Canada. (email: jeffrey.tsang@ieee.org).

This study differs from the previous work done in [2] due to the change of metric involved: it is more natural to define the distance between two probability distributions via $\mathcal{L}_1$ (also well-known as the total variation distance [12]) rather than $\mathcal{L}_2$ distance, which would involve squares of probabilities — it is not even mentioned in [12].

For example, under $\mathcal{L}_2$ the distance between Always Cooperate and Always Defect (which is maximal) is non-constant, depending on parameter values; under $\mathcal{L}_1$ it is identically 1. Still, this causes major ripple effects, many of which come as consequences of breaking the nicer mathematical structure of $\mathcal{L}_2$.

The rest of the paper is organized as follows: the fingerprint is defined and useful properties given in Section II, the experiments are listed and described in Section III, the results and interpretation follow in Section IV, finally the discussion and conclusion are in Section V.

II. Background

As developed in [1] and exposited in [2], the fingerprint operator used in this study is based on the length-weighted probability of a move pair occuring, when the given agent plays against a parametrized k-state probabilistic finite state transducer. We will restrict our consideration to a 1-state machine, which can be parametrized as $(x, y, z) \in [0, 1]^3$, where x is the probability of cooperating on the initial move, y is the probability of cooperating in response to a cooperate, and z the probability of cooperating in response to a defect.

The operator takes as input a *specification* of a game playing agent P, which is a function ρ_P that gives the probability the agent plays as an input move history s (a string of moves) up to its length, given that its opponent plays as another input move history w as directed. That is, $\rho_P(s, w) = \Pr(\forall i\ P \text{ plays } s_i \text{ in turn } i \mid \forall j \text{ opponent plays } w_j \text{ in turn } j)$. Call the parametrized opponent $O_1(\vec{v})$ with $\vec{v} = (x, y, z)$, and define $\rho_{O_1(\vec{v})}$ similarly.

Denote by $\mathcal{F}_P$ the output of the operator on P; the (m_1, m_2)th component of the fingerprint function is defined as

$$\mathcal{F}_P(\vec{v})_{m_1 m_2} = \sum_{n=1}^{\infty} \mu(n) \sum_{\substack{(s,w) \text{ has length } n \\ s \text{ ends with } m_1 \\ w \text{ ends with } m_2}} \rho_{O_1(\vec{v})}(w, s)\rho_P(s, w)$$

the first sum is the two-way probability the players play (m_1,m_2) respectively on the nth move, weighting that by

978-1-4673-5311-3/13/$31.00 ©2013 IEEE

a given function $\mu(n)$. For special properties, we will use the family of geometric distributions: $\mu(n;\alpha) = (1-\alpha)\alpha^{n-1},\ \alpha \in [0,1)$. The limit as $\alpha \to 1^-$ exists and equals the expected value, but we regain all the problems such as discontinuity from using it.

To compute this for agents representable by finite state transducers, create the following Markov chain: the state space is $Q \times \{\mathrm{C},\mathrm{D}\}^2$, the states of the agent paired with the last moves of P then O_1. The transition matrix T has entries $(q_1, m_1, m_2) \to (q_2, m_3, m_4)$ equal to $\Pr(P$ transitions from q_1 to q_2 outputting m_3 seeing $m_2) \times \Pr(O_1$ outputs m_4 seeing $m_1)$.

The fingerprint function is then

$$\mathcal{F}_P(x,y,z;\alpha)_{m_1m_2} = (1-\alpha)\chi^T_{m_1m_2}\big(I-\alpha T(y,z)\big)^{-1}Q_0(x)$$

where $\chi_{m_1m_2}$ be the indicator vector whose entry is 1 if the state indexed has last move-pair (m_1, m_2), 0 otherwise, and $Q_0(x)$ is the initial state probability vector.

Define the distance between two fingerprints using the $\mathcal{L}_1$ or statistical distance:

$$\|\mathcal{F}_{P1} - \mathcal{F}_{P2}\| = \int_{[0,1]^3} \sum_m |(\mathcal{F}_{P1} - \mathcal{F}_{P2})(x,y,z)_m| \,\mathrm{d}x\mathrm{d}y\mathrm{d}z$$

III. Experimental Design

We consider a 3-state finite transducer representation for Prisoner's dilemma. This is a string, with an initial action, and for each of 3 states, a 4-tuple (transition on cooperate, action on cooperate, transition on defect, action on defect). All actions are cooperate/defect, and transitions point to any one of the 3 (labelled) states. The automaton always starts at the lexicographically first state.

There are thus $2^7 3^6 = 93,312$ representable strategies; using standard state-minimization algorithms to remove duplicates, there are 23,000 unique strategies in the space. This is comprised of 2,592 copies each of the 8 1-state strategies, 86–128 copies of the 288 2-state strategies, 2 copies each of the 3-state strategies (as state 2 and 3 can be interchanged).

Each of the 23,000 strategies was fingerprinted, computed using the matrix formula using the LAPACK linear algebra package into a 4-component function of y, z for $x = 0, 1$ at $\alpha = 0.8$, a value found in previous studies to have good separation properties [2].

Approximate pairwise distances are calculated with a composite third order product Gaussian cubature method (4 points at $(\pm 1/\sqrt{3}, \pm 1/\sqrt{3})$ for the region $[-1,1]^2$, see [13]) with a grid of 64×64 evenly spaced squares (16,384 evaluation points). These are summed in a binary divide-and-conquer fashion to decrease roundoff error.

Hierarchical clustering with the unweighted pair group with arithmetic mean method (UPGMA) is performed on this $23,000 \times 23,000$ distance matrix. The closest pair of clusters is repeatedly merged, and the distance between two clusters is defined to be the average over all possible pairs, one from each cluster.

As the tree does not exhibit any strong cutoff points, we pick a level of 980 clusters (by removing the 979 largest distance mergings), which is the last noticeable jump in distance below 1000, for further analysis.

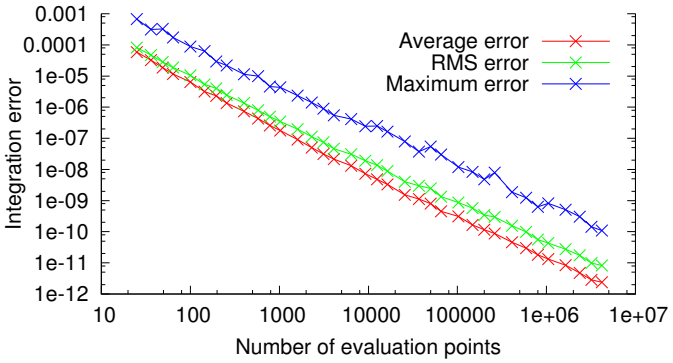

Fig. 1. Several measures of error in a set of sample calculations of fingerprint distance. Note the log-log scale.

Metric multidimensional scaling is used to embed these clusters into the Euclidean plane. This works by minimizing the *stress* loss function

$$\sum_{i,j} w_{i,j}(\delta_{i,j} - d_{i,j})^2$$

where $w_{i,j}$ is the product of the cluster sizes of i, j, $\delta_{i,j}$ is the true distance of clusters i, j, and $d_{i,j}$ is the distance between the points on the plane representing clusters i, j.

The stress majorization SMACOF algorithm [14] is used for this purpose, with the best fit chosen from multiple runs starting at uniformly random initial points.

We can define a simple mutation operator that takes any one position in the string representation and changes it to any other value. This induces a mutational *connectivity network* on the space that can be investigated, being an important property of the representation. We take all 93,312 strategies and find each of their 19 neighbours 1 mutation away, then display the propensity of a cluster to mutate into each other cluster.

IV. Results

The distance matrix required 2 CPU-years to compute. There is no mathematically proven error bound usable, so an experimental approach was adopted.

A. Integration error

Random pairs of 3-state automata were generated, and their fingerprint distance was approximated using the order 3 product Gaussian quadrature with various sizes of grids.

The exact integral is not analytically computable, and so the same method with 16,777,216 evaluation points was used as the ground truth. Statistics of the error of the method, over 13,337 independent trials, were calculated and plotted in Figure 1.

The results give strong experimental evidence that the error of integration is bounded well below 10^{-6} (at 16,384 evaluation points). As the smallest nonzero distance is about 0.00179, integration error is not an issue.

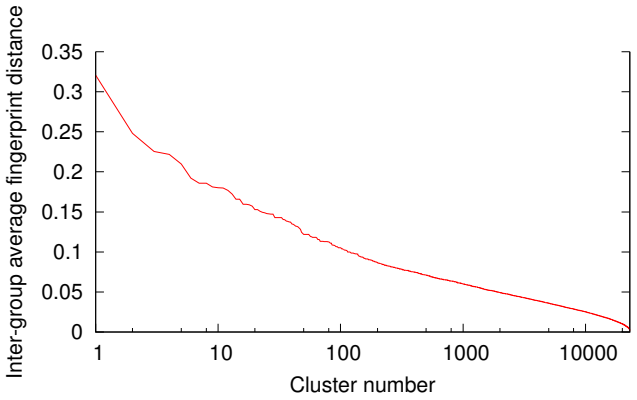

Fig. 2. The distance between clusters n and $n+1$ when combined in the UPGMA hierarchical clustering tree. Note the logarithmic scale.

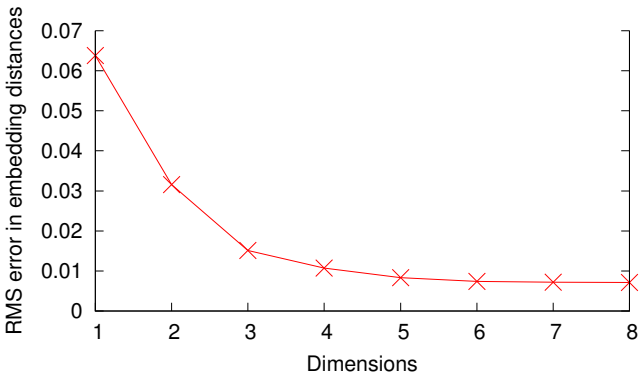

Fig. 3. The best-fit weighted root mean square error in embedding the pairwise distance matrix for the 980 clusters into $\mathbb{R}^n$, for various n.

B. Colouring

Assigning a colour to each strategy allows extra dimensions of information in a plot and perhaps we can find correlations of position to colour. We propose here a schema:

Ignore the outputs for now and consider only the transition graph of the automaton, which has 3 states and 7 edges total (2 each from each state and the initializing edge). Run the fingerprint calculation $(1-\alpha)(I-\alpha T)^{-1}q_0$ assigning each edge except the initializer a probability of 0.5, with the same $\alpha = 0.8$; q_0 is the first state. This gives a probability distribution on the 3 states; the initial 'state' has weight $1-\alpha = 0.2$.

The four 1-state strategies are coloured as follows: ALLC (always cooperate) is green, ALLD (always defect) is red, TFT (tit-for-tat) is blue, and PSY (reverse tit-for-tat) is black. Consider only the actions taken while at each state, and give each state the colour of its corresponding 1-state strategy.

Now weight the states according to the above probability distribution, with the initial move assigned as ALLC/ALLD. This average is the colour of the automaton. Because the sum of all components is 1, being a probability vector, assigning black (equivalently no contribution) to PSY allows the nominally 4-dimensional surface to be fit conveniently in 3-dimensional RGB colour space.

For example, consider the strategy represented by the string 0102001112021. The transition graph in matrix form is

$$T = \begin{bmatrix} 0 & 0.5 & 0 \\ 0.5 & 0.5 & 0 \\ 0.5 & 0 & 1 \end{bmatrix}$$

with T_{ij} the probability of transitioning from state j to state i. Then $(1-\alpha)(I-\alpha T)^{-1}q_0$ is $[3/11, 2/11, 6/11]$. Thus the colour of this strategy is $(1-\alpha)\cdot\text{Green} + \alpha(3/11\cdot\text{Green} + 2/11\cdot\text{Red} + 6/11\cdot\text{Blue})$.

The colour assigned to a cluster is a simple average of the colours of each constituent automaton.

C. Hierarchical clustering

The distance between the clusters that are combined at each step (the minimum distance between any two clusters at that stage) is plotted in Figure 2. It may be a consequence of the large number of points, but the distances do not exhibit any clear jumps that would indicate distinct levels of structure.

The distances also loosely fit to a logarithmic curve, which suggests that the cloud of points are evenly dispersed throughout the space. Cluster size after the merging at each step (omitted, see Figure 4) says otherwise. That distribution is very spread out, meaning many clusters are much larger than others, even at the same cutoff distance — this shows the points are not that evenly dispersed.

D. Multidimensional scaling

We chose a cutoff at 980 clusters mostly arbitrarily as there are no clear levels of structure in the data. The UPGMA algorithm is re-run until 980 clusters remain, then the reduced distance matrix (now 980×980) is used for multidimensional scaling.

Princical component analysis is not applicable because the data has non-Euclidean geometry; indeed attempts to do so (omitted) create a matrix, required to be positive semidefinite, with almost half its eigenvalues negative. Further attempts to ignore this and simply take the components corresponding to the largest positive eigenvalues have the strange property that approximation error *increases* beyond the first few.

To investigate the inherent dimensionality of the data under MDS, the algorithm was re-run embedding the points into different dimensions (the target space is a parameter choice). The best-fit error versus dimension is shown in Figure 3. The general idea is that if adding an extra dimension lowers the error by an insignificant amount, then it is superfluous and can be safely not done.

The results indicate that 3, possibly 4 dimensions suffice to account for most of the data. The error is almost constant beyond 6 dimensions; the residual error comes from the fact that the data is non-Euclidean, and hence cannot be exactly embedded in Euclidean space of any dimension at all.

A scatter plot in 3 dimensions is too difficult to handle, and so we choose to use the 2-dimensional fit, knowing that there is a dimension in the data left undisplayed. The root mean square error is 0.03158, with the maximal distance being 1.

These points, now explicitly in Euclidean space, are rotated to the princical components.

The clusters are plotted in Figure 4, with the colouring scheme described in Section IV-B. The first major implication from the plot is that the colouring scheme is clearly reflected in the position of the points: there is a strong correlation with the colour green with the positive y-direction, red with negative y, blue with positive x and black with negative x.

This proves that the scheme works, as well as that the 8 1-state strategies are important — ALLC and ALLD are at the extreme edges of the plot, while TFT and PSY are close to the boundary (only a few tiny clusters are further out in the x direction). These are the much larger clusters, with sizes ranging from 2,592 (two of them are purely TFT and purely PSY) to 2,922 (ALLC and other nearby strategies).

Notice the symmetry of the configuration: the axis of symmetry involves replacing an automaton with another that plays its reverse move at every step. From the plot it seems this induces a *rotational* symmetry in the points. There is still an evocative semblance to reflective symmetry across both the coordinate axes.

E. Princical components

We hypothesize here, by analogy with the results in [1], [2], that the 2 principal components in the embedded clusters correspond to *cooperativity* (propensity to favour one move over the other), and *coordination* (how correlated your move is to your opponent's).

This is bourne out by the striking gradient of colours in Figure 4, which suggests a quantitative test: we can measure the Pearson correlation coefficient between the colouring of the clusters and their position.

We use as a measure of cooperativity the linear predictor Green − Red in the colouring scheme (bounded between -1 and 1), and for coordination the predictor Blue − Black.

The Pearson correlation (bounded in [-1,1]) between Green − Red and the y-coordinate of the points is 0.99857, between Blue − Black and the x-coordinate 0.93469. We consider the hypothesis proven beyond a reasonable doubt.

F. Cluster-wise mutation

To make the data easier to display, we reduce the number of clusters to 139 the same way (removing the 138 largest distance mergings done by UPGMA). Each automaton in the space has 19 neighbours according to our simple string-based mutation.

We compute for each cluster i the probability distribution of which cluster j it moves to in 1 mutation, by averaging over all automata in the cluster. This can be displayed as a heatmap in Figure 5. The ordering of the clusters is as the hierarchical clustering tree: if two clusters would be merged in the (removed) UPGMA, they are adjacent. The colour bar displays each cluster's colour and can be compared to the colours of the points in Figure 4.

One feature that leaps out of that map is the special properties of the 1-state automata: not only are they the largest clusters, they also have the highest probability of null mutations, as evidenced by the brightest colour of those squares. This is not surprising, as one way of having a 1-state automaton is by completely not using a state or two — all mutations that affect unreachable states are inherently null.

Several larger-scale squares (groups of adjacent clusters) can be discerned by their relative reachability under mutation — a lot of mutations do not move you far from the diagonal, hence to a nearby cluster.

A distinct band at mid-range can be seen: that is a mutation that crosses the axis dividing the ALLC/ALLD halves of the space. Note that the outer diagonals of the entire heatmap are empty: there is no way to go from ALLC to ALLD in 1 move.

V. Discussion and Conclusion

Hierarchical clustering shows that this space of strategies is mostly dispersed, but in an uneven manner. With multidimensional scaling, 980 clusters of strategies were embedded into the plane and show a high degree of symmetry across both principal component axes.

A simple colouring scheme based on the fraction of time a strategy plays (for one move) as each of the 1-state strategies (against a RANDOM opponent) is dramatically successful in explaining the spatial distribution of the clusters. The smooth colour gradient in the scatter plot of Figure 4 is evidence of that.

Furthermore, the simple linear predictor, fraction of time as ALLC minus fraction of time as ALLD, has an almost perfect correlation with the y-coordinate of the clusters, meaning not only is this dominant direction seen to be *cooperativity*, it can be precisely *quantified* as well.

The analogous linear predictor (fraction as TFT minus fraction as PSY) has an extremely high correlation with the x-coordinate. Thus the second principal coordinate of the strategy space is seen to be *coordination*. That the correlation is less perfect may be due to the use of 2 instead of 3 or 4 dimensions, as identified by the error-vs-dimension plot Figure 3.

Repeating these experiments with the data embedded into 4 dimensions can verify these results, as well as possibly find the interpretation of the third and fourth coordinates.

From comparison to [2], it is interesting that the dominant two components (at $\alpha = 0.8$) still retain their role, even as the underlying metric is modified, and the representation changed. This serves as validation that the specific choice of metric did not severely affect the data. Also in [2], the third coordinate was found to be *bias*, or favouring one move over another. It would be interesting to see if that hypothesis holds for this data.

Another direction for future work is replicating the analysis for different choices of α, which was not done here due to computational constraints.

The mutational heatmap confirms several aspects of the hierarchical clustering, as several larger-scale clusters tend to retain members under mutation. The 1-state strategies are distinguished by their highest propensity for null mutations,

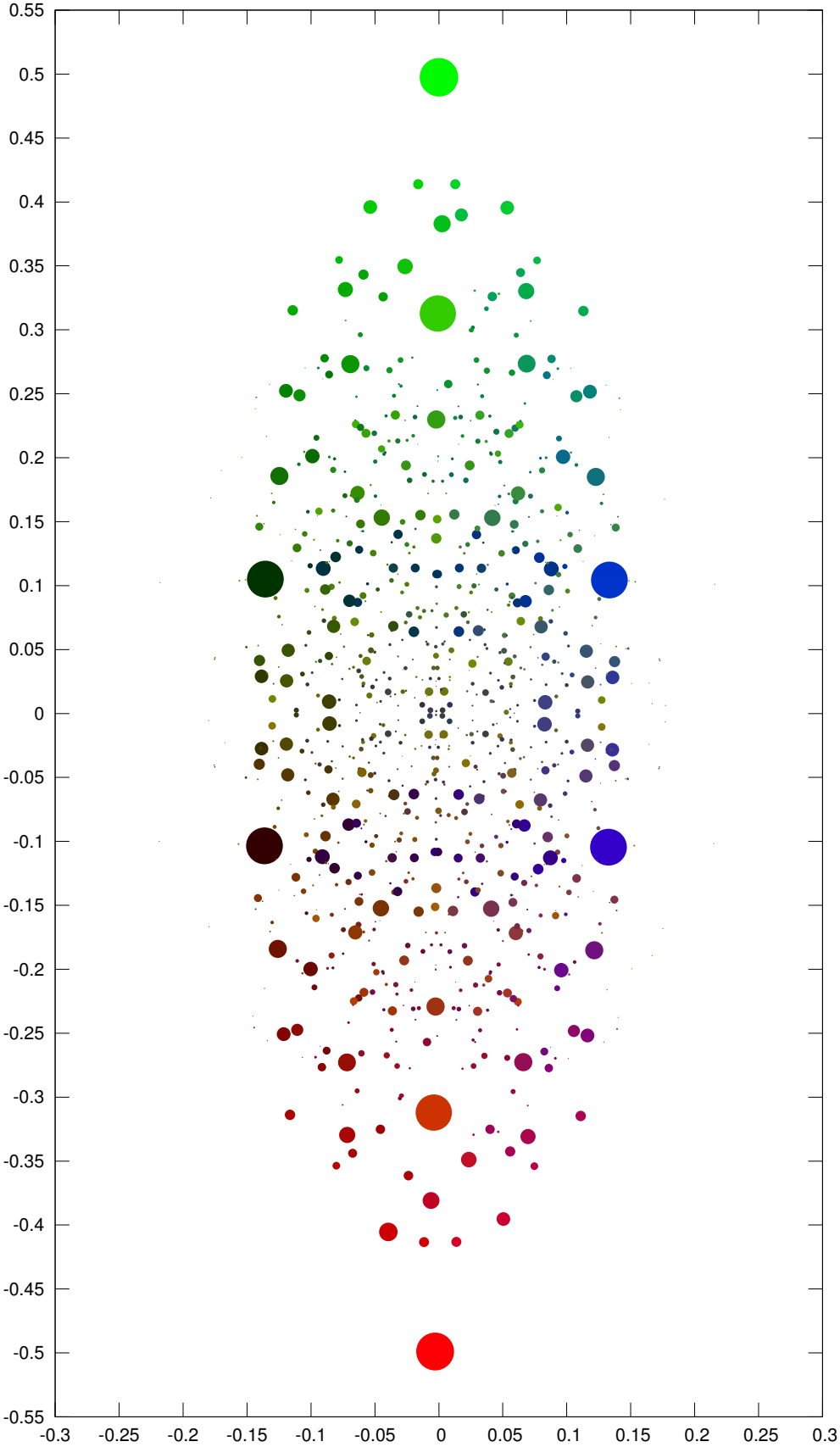

Fig. 4. Scatter plot of all 93,312 strategies, reduced to 980 clusters with UPGMA hierarchical clustering, projected to 2D with metric MDS. Point size (area) is directly proportional to cluster size; for colouring see section IV-B; axes are rotated to principal components, positive orientation is arbitrary.

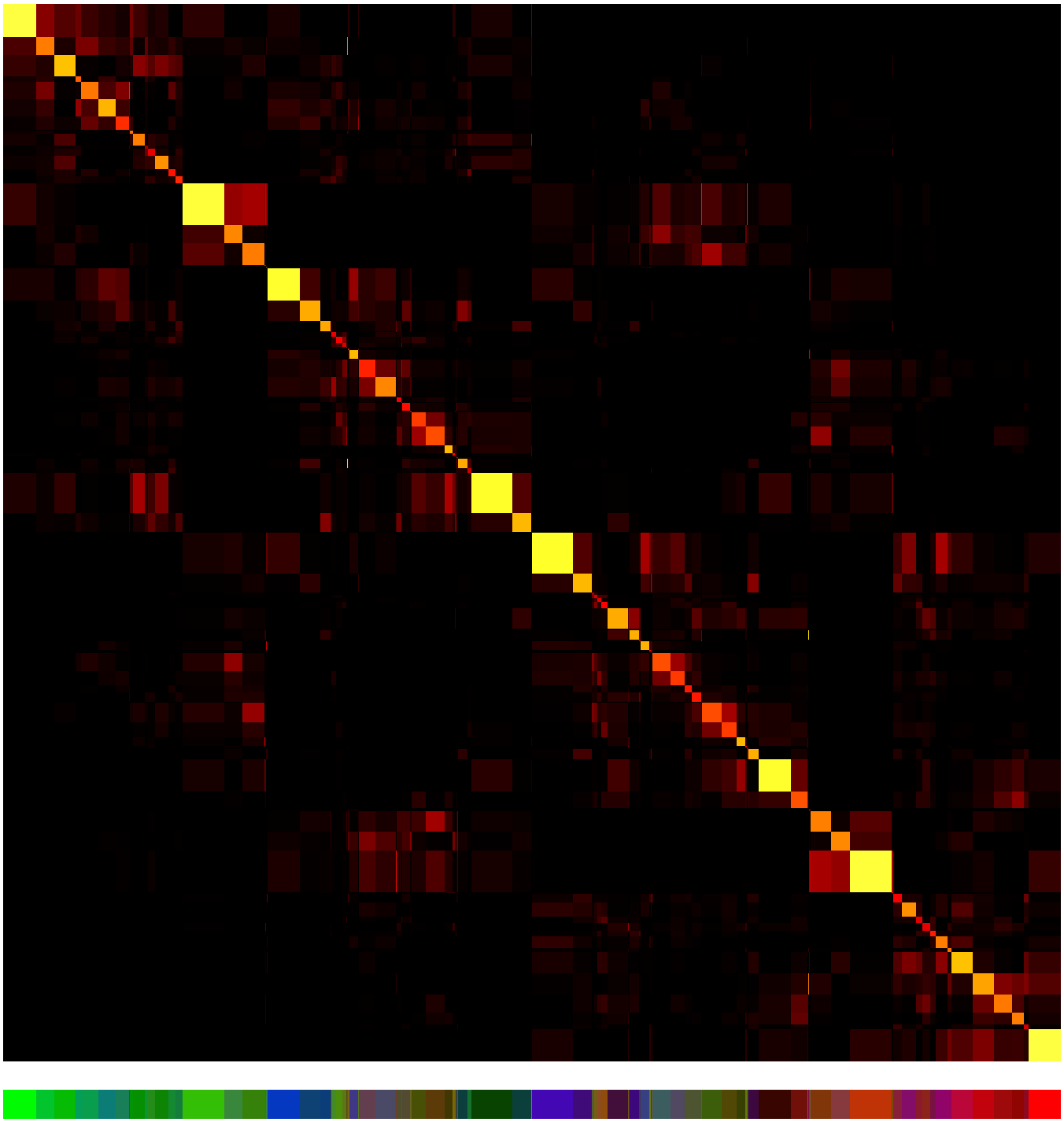

Fig. 5. Single mutation connectivity heat-map: cluster width is proportional to size and clusters are listed in the hierarchical tree order. Rectangle (i, j) has as value the (cluster average) probability of transitioning from cluster i to cluster j in one mutation: the colourmap is piecewise linear, black at 0, red at 0.25, yellow at 0.5 and white at 1.0. The colour bar on the bottom displays each cluster's colour as in section IV-B.

as well as their large cluster size. This may explain their relative frequency of occurrence in evolutionary experiments.

In conclusion, the structure of a space of 3-state automata under the fingerprint metric was investigated using a variety of tools. It was found to be dispersed in an irregular manner, have at most 4 or 5 major dimensions, of which the top 2 are cooperativity and coordination. A simple quantification of these attributes was found and enjoys unqualified success.

Next steps for exploring more strategy spaces, and possibly confirming that the dominiant coordinates in general strategy spaces is invariant, include a grid of probabilistic automata. Leaving behind the restriction that the strategies be deterministic offers a good chance for encountering different "coordinates".

The other way to avoid the crushing weight of combinatorial explosion is to use a sampling technique. As usual evolutionary experiments would use representations of no less than 8 states, an attempt to catalogue that space would prove more directly applicable.

A further, direct test of the similarity of representations is to combine them into one set for analysis. Then using the distance matrix across the representations allows quantification of how different they are, especially with regards to their principal coordinates. The use of multiple representations simultaneously for comparison has been pioneered in [6], and a theoretical counterpart would be valuable.

ACKNOWLEDGMENTS

The author would like to thank Rajesh Pereira and Daniel Ashlock, University of Guelph. This work is supported by a Natural Sciences and Engineering Research Council of Canada Postgraduate Scholarship, and was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET:www.sharcnet.ca) and Compute/Calcul Canada.

REFERENCES

[1] J. Tsang, "The Parametrized Probabilistic Finite State Transducer Probe Game Player Fingerprint Model", *IEEE Transactions on Computational Intelligence and AI in Games* 2(3), 2010, pp.208–224.

[2] J. Tsang, "The Structure of a Depth-3 Lookup Table Representation for Prisoner's Dilemma", in *Proceedings of the 2010 IEEE Conference on Computational Intelligence in Games*, 2010, pp.54–61.

[3] D. Ashlock and E-Y. Kim, "Techniques for analysis of evolved prisoner's dilemma strategies with fingerprints", in *Proceedings of the 2005 Congress on Evolutionary Computation*, 2005, pp.2613–2620.

[4] D. Ashlock, E-Y. Kim and W.K. von Roeschlaub, "Fingerprints: enabling visualization and automatic analysis of strategies for two player games", in *Proceedings of the 2004 Congress on Evolutionary Computation*, 2004, pp.381–387.

[5] D. Ashlock and E-Y. Kim, "Fingerprinting: Visualization and Automatic Analysis of Prisoner's Dilemma Strategies", *IEEE Transactions on Evolutionary Computation* 12(5), 2008, pp.647–659.

[6] H. Ishibuchi, H. Ohyanagi and Y. Nojima, "Evolution of Strategies With Different Representation Schemes in a Spatial Iterated Prisoner's Dilemma Game", *IEEE Transactions on Computational Intelligence and AI in Games* 3(1), 2011, pp.67–82.

[7] D. Ashlock, E-Y. Kim and N. Leahy, "Understanding representational sensitivity in the iterated prisoner's dilemma with fingerprints", in *IEEE Transactions on Systems, Man and Cybernetics C* 36(4), 2006, pp.464–475.

[8] D. Ashlock and E-Y. Kim, "The impact of cellular representation on finite state agents for prisoner's dilemma", in *Proceedings of the 2005 Genetic and Evolutionary Computing Conference*, 2005, pp.59–66.

[9] W. Ashlock and D. Ashlock, "Changes in Prisoner's Dilemma Strategies Over Evolutionary Time With Different Population Sizes", in *Proceedings of the 2006 Congress on Evolutionary Computation*, 2006, pp.297–304.

[10] D. Ashlock, E-Y. Kim and W. Ashlock, "Fingerprint Analysis of the Noisy Prisoner's Dilemma Using a Finite State Representation", *IEEE Transactions on Computational Intelligence and AI in Games* 1(2), 2009, pp.157–167.

[11] D. Ashlock, E-Y. Kim and W. Ashlock, "Fingerprint Analysis of the Noisy Prisoner's Dilemma", in *Proceedings of the 2007 Congress on Evolutionary Computation*, 2007, pp.4073–4080.

[12] A.L. Gibbs and F.E. Su, "On choosing and bounding probability metrics", *International statistical review*, 70(3), 2002, pp.419–435.

[13] A.H. Stroud, *Approximate Calculation of Multiple Integrals*, Englewood Cliffs, N.J.: Prentice-Hall, 1971.

[14] J. de Leeuw, "Applications of Convex Analysis to Multidimensional Scaling", in *Recent Developments in Statistics*, Eds. J.R. Barra et. al, Amsterdam: North-Holland, 1977, pp.133–145.