ALGORITHMIC CLASSIFICATION OF DRAINAGE NETWORKS ON MARS AND ITS RELATION TO MARTIAN GEOLOGICAL UNITS.

T. F. Stepinski[1], R. Vilalta[2], M. Achari[2], P. J. McGovern[1], [1]*Lunar and Planetary Institute, Houston TX 77058-1113, USA, (tom@lpi.usra.edu, mcgovern@lpi.usra.edu)*, [2]*Dept. of Computer Science, University of Houston, 4800 Calhoun Rd. Houston TX, 77204-3010, USA, (vilalta@cs.uh.edu, amkchari@cs.uh.edu).*

Abstract. We present an algorithmic classification of drainage networks on Mars. Altogether, 368 drainage networks were computationally extracted using MOLA topographical data from various martian locations covering 16 major geological units. The classification is quantitative and objective, as it is based on a numerical description of drainage networks. Applying a clustering algorithm to our dataset, we have found that the networks can be best divided into 9 clusters. These clusters are separated from each other, so they can be used for classification purposes. Our partition does not correlate with an existing division into geological units. A morphological interpretation for this emergent classification is still been developed. One particular cluster was studied in detail, and we have determined that it groups networks overlaying landscapes dominated by topographical basins.

Introduction. The morphology of martian landscapes is of great interest because it helps to identify physical processes responsible for the presently observable topography. Traditionally, the descriptive morphology has been used to study and categorize different types of martian landscapes. This resulted in dividing the martian surface into a number of geological units, terrains with common morphological features (1). Our long-term goal is to devise an algorithm capable of classifying martian landscapes objectively and quantitatively, solely on the basis of topographical data provided by the Mars Orbiter Laser Altimeter (MOLA). Such an algorithmic classification may cut across existing divisions.

In this paper we present the first step toward our goal, an algorithmic classification of martian drainage networks. Using MOLA topographical data, a drainage network can be computationally extracted (2) from any terrain, including terrains that never experienced any real flow. We prefer to work with drainage networks because they are much simpler than their underlaying landscapes. We assume that there is enough correspondence between a landscape and its drainage network that the classification of networks is tantamount to the classification of landscapes.

A morphology of a drainage network can be encapsulated (3) in a so-called "network descriptor", a list of four numbers $A = (\tau, \gamma, \beta, \rho)$. Briefly, τ, γ, and β are parameters that characterize distributions of contributing areas, lengths of main streams, and dissipated energy, respectively. The parameter ρ measures the spatial uniformity of drainage. The network descriptor offers an abstract but very compact characterization of a drainage network. Our methodology is to classify drainage networks based on the values of their network descriptors (two networks are similar if the values of their descriptors are similar). Our dataset consist of values of A derived from martian drainage networks. We applied a clustering algorithm over our dataset that produces as output a set of clusters $\{c_1, c_2, \cdots, c_k\}$, where each cluster c_i covers certain number of network descriptors (networks) that are similar to each other and distinct from the rest. The set of clusters is mutually exclusive and exhaustive, it constitutes a good basis for classification if the resulting clusters are well separated from each other.

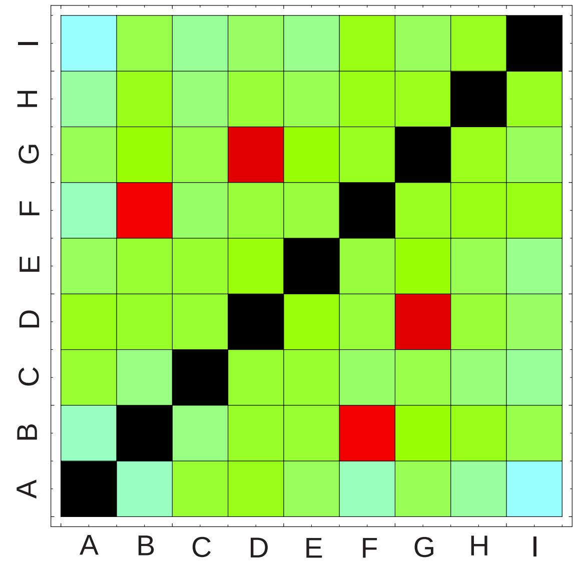

Figure 1: The matrix of normalized distances between clusters A to I. The matrix is symmetric. Black color corresponds to $d = 0$, red colors correspond to $d < 1$, green colors correspond to $d > 1$. Lighter green, turning into blue corresponds to larger value of d (bigger separation).

Data and Methods We have extracted 386 drainage networks from martian locations with a wide range of latitudes and elevations that represent all three major epochs and 16 geological units: Npl1, Npl2, Npld, Nple, Nplr, Nh1, Had, Hh3, HNu, Hpl3, Hr, Hvk, Ael1, Aoa, Apk, Aps (1). There are 152 (39%) networks extracted from Noachian surfaces, 145 (38%) from Hesperian, and 89 (23%) from Amazonian.

The probabilistic clustering algorithm used in our experiments (4) groups records into clusters by modelling each cluster through a probability density function. Each record in dataset then has a probability of class membership and is assigned to the cluster with highest posterior probability. Unless known a priori, the optimal number of clusters is normally not known and one must employ some form of cross-validation to estimate this value. The implementation we used (5) varies the number of clusters automatically until it reaches an optimum value.

Results. Applying the clustering algorithm directly over

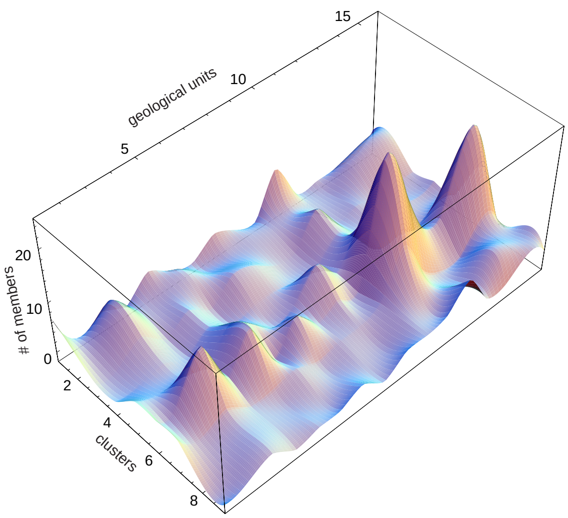

Figure 2: The match between clusters and geological units. The clusters, A to I (1 to 9) are on the x axis, the geological units, 1 to 16, in the order given earlier in the text, are on the y axis. The z axis corresponds to the the value of each entry in matrix M. Distribution of peaks along a diagonal of $x - y$ plane would indicate a good correlation between clusters and geological units.

our dataset of networks resulted in dividing the 386 networks into 9 clusters, which we would label by the first 9 letters of the alphabet (A to I). The number of networks in these clusters is as follow: 29, 37, 23, 16, 49, 129, 40, 28, 35, respectively. To assess the degree of separation between clusters, we constructed a matrix of normalized distances, d, between the clusters, $d > 1$ indicates strong separation, $d < 1$ indicates only partial separation. Strong separation of clusters indicates that our dataset has been divided into distinct groups that could be used for network classification. Figure 1 illustrates normalized distances between our clusters. In overwhelming number of cases the separation of clusters is strong, The only cases where $d < 1$ is between clusters B and F ($d = 0.95$) and D and G ($d = 0.88$). Thus, from a formal point of view we succeeded in classifying network descriptors of martian drainage networks.

To assess the degree of match between our nine clusters and 16 martian geological units we constructed a matrix M where each row corresponds to a cluster and each column to a geological unit. Each entry $M[i, j]$ in M indicates the number of networks covered by a given cluster and a given geological unit. Under a perfect match we could always align rows and columns in such a way that the diagonal of M shows high values and all other entries very low values. Figure 2 illustrates matrix M; columns and rows are ordered to try to maximize the values along the diagonal. As shown in Fig. 2, the degree of match between our clusters and geological units is very small. Thus, our clusters do not correspond to martian geological units.

In order to see what particular morphological feature (or features) determines a given cluster we need to interpret values of a network descriptor in morphological terms. This work is in progress and it would determine whether our clustering translates into a useful classification of actual networks, and, ultimately, into a useful classification of martian landscapes.

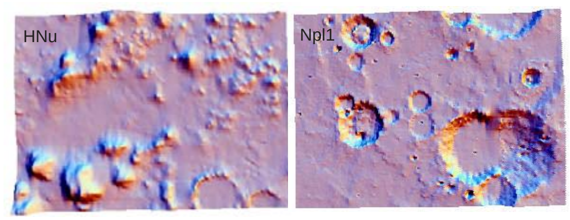

Figure 3: Two examples of martian terrains from different geological units (as labelled), whose drainage networks have been assigned to cluster A. Both terrains feature large, enclosed depressions.

For example, cluster A is characterized by low values of τ and ρ indicating a relatively non-hierarchical organization and irregular structure of networks belonging to this cluster. A visual investigation of all cluster A networks showed that these networks overlay terrains that are dominated by topographical basins, large, enclosed depressions in a landscape. These are mostly, but not exclusively, large craters. Our drainage network extraction algorithm considers such features as "lakes" and drains them in a regular (non-hierarchical fashion), hence low values of τ and ρ. This example illustrates why our clusters do not correspond to geological units, a basin can occur in any geological unit (see Fig. 3 for two examples), it's a feature not related to the present classification. In addition, the analysis of cluster A suggests that our entire classification could be biased toward hydrological aspects of landscape morphology. A similar analysis for other eight clusters is in progress.

References

(1).Scott, D.H., and Carr, M.H. 1977, Geological Map of Mars, *U.S.G.S. Misc. Geol. Inv.*, Map I-1093

(2). Tarboton, D.G., Bras, R.L., and Rodriguez-Iturbe, I. 1989, Technical Report no. 326, Ralf M. Parsons Lab., Massachusetts Institute of Technology, Cambridge, MA.

(3). Stepinski, T.F., Marinova, M.M., McGovern, P.J., and Clifford, S.M. 2002, Geophysical Research Letters, 29, No 8, 10.1029/2002GL014666.

(4). Witten, I.H. and Frank, E. 2000, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Academic Press, London U.K.

(5). Cheeseman, P. and Sturz 1995, Bayesian Classification (AutoClass) Theory and Practice. Advances in Knowledge Discovery and Data Mining. MIT Press