GROUPING CHEMCAM TARGETS BY VISUAL CHARACTERISTICS IMPROVED BY AUTOMATIC PARTITIONING. A. Essunfeld[1] (aessunfeld@lanl.gov), J. M. Comellas[1,2], P. J. Gasda[1], D. Oyen[1], N. Lanza[1], O. Gasnault[3], D. Delapp[1], R. Wiens[1], S. Clegg[1], C. C. Bedford[5,6], E. Dehouck[7], B. Clark[8], R. Anderson[4], [1]Los Alamos National Laboratory, [2]UH Mānoa, [3]IRAP, Toulouse, France, [4]USGS, [5]LPI, USRA, [6]NASA JSC, [7]Univ. Lyon, [8]SSI.

Introduction: NASA's *Curiosity* rover has been traversing Gale crater, a lacustrine region chosen as the landing site of the rover due to its potential for past habitability, since landing there in 2012 [1]. *Curiosity* spent the first ~760 martian solar days (sols) of the mission in the Bradbury formation, an ancient fluvio-lacustrine system [1, 2].

In the nine years since landing on Mars, *Curiosity* has observed a wide variety of rock types, and several classification methods have been developed with the aim of sorting these rocks into process-oriented facies [e.g., 3-5]. But accurate classification of rocks can be challenging when information is limited to images and chemical composition, meaning process-oriented classifications risk introducing bias. [6] addressed this issue by developing a classification system based only on simple visual attributes. This system involved three phases: first, (1) the manual process of reviewing each target's RMI and encoding its visual attributes as a 17-digit binary number; then (2) an initial algorithmic grouping of the targets; and finally (3) a manual review, which refined the algorithm-generated groupings [6]. The second phase generated ten graph components with varying connectivity, and the relatively weakly connected components seemed to correlate with worse target image association [6]. In this work, we attempt to automate the third phase by automatically partitioning the components with weak connectivity.

Methods: Curiosity's ChemCam instrument uses Laser-Induced Breakdown Spectroscopy (LIBS) to obtain chemical information about rock targets [7, 8]. With each LIBS analysis, high-resolution images of the target are taken with the Remote Micro Imager (RMI) [7, 8]. In this work, we use the same dataset studied by [6, 9]. This dataset includes the visual attribute documentation described in [6]: a 17-digit binary number for each target, encoding its visual attributes.

To examine the graph components that had weaker connectivity and partition them into two or more subcomponents, we needed to define a threshold for "weak connectivity" in this context. "Connectivity" alone refers to the minimum number of nodes that need to be removed to render a graph disconnected [13]. But connectivity does not allow us to distinguish between strongly connected graphs with leaves, and dumbbell-like graphs. For instance, an otherwise strongly connected graph could have connectivity = 1 because it has a single leaf (e.g., Fig. 1, brown). This is the same connectivity value as dumbbell-like graph, which has two or more separate, strongly connected components that are only connected to each other (bridged) by a single node (e.g., Fig. 1, blue graph). But only the dumbbell-like graph is a worthy candidate for partitioning. By removing edges to the bridge node (Fig. 1, red arrow), we could gain two new graphs, each with stronger connectivity than the original blue graph.

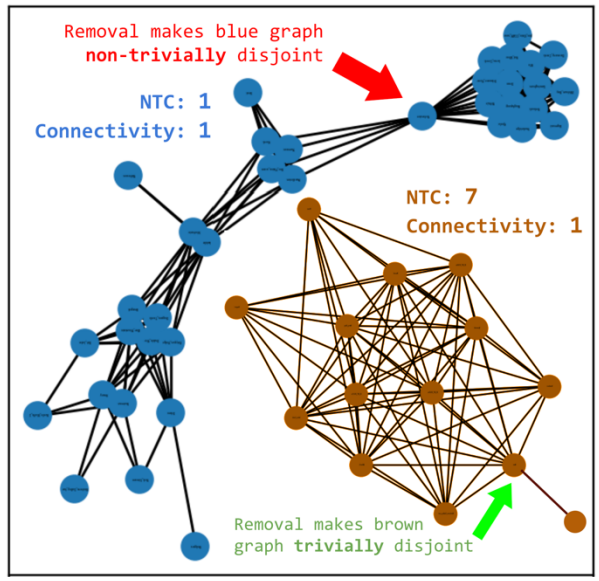

Figure 1: Two graph components from the original classification by [6]. ChemCam targets are represented as nodes, and similarity between targets as edges between nodes [6]. The brown graph is strongly connected, except for one leaf (added for clarity). The blue graph is weakly connected. The red arrow points to a "bridge" node. The simple connectivity of the two components is the same, but the non-trivial connectivity (NTC) differs.

The inability of simple connectivity to distinguish between these two different characters of graphs motivated our definition of "non-trivial connectivity" (NTC). To determine a graph's NTC, we first remove (or prune) the leaves, then compute the simple connectivity of what is left behind. As seen in Fig. 1, this process reveals that the brown graph has significantly higher NTC than the blue graph. Thus the metric of NTC, unlike simple connectivity, allows us to distinguish between these two different characters of graphs.

We used the following two conditions to identify components worth partitioning: A component was considered to have "weak connectivity" (and thus worth partitioning) if it (A) had NTC ≤ 1 , and (B) had more nodes than the median component size generated by the original algorithmic grouping in [6]. We partitioned the components that passed this query using the Kernighan-Lin Bisection algorithm (KL) as implemented in the python library NetworkX [10]. This algorithm finds a fast, approximate solution to the balanced graph cut problem [14]. KL works by initializing a pseudo-random partition and then swapping nodes between the sides of the parti-

tion, rewarding final partitions that require the fewest edges to be removed from the original graph [14]. To mitigate the influence of randomness, we bootstrapped KL with 1000 repetitions and 1000 maximum iterations per repetition, and picked the mode partition.

Results: The original classification graph generated 10 components with a median component size of 12 nodes [6]. Of the five components with size greater than or equal to the median, two had NTC ≤ 1. These had sizes of 37 and 53, and constituted the weakly connected components that were partitioned with KL. After partitioning, these two components with median size 45 (both NTC = 1) became five components with median size 18 (mean 17) and median NTC = 3 (mean 6), along with three new miscellaneous targets (no edges). The process of partitioning the size-53 component is illustrated in Fig. 2. The size-37 component's process was similar, resulting in two new groups and the three new miscellaneous targets.

Discussion: [6] obtained 10 groups in phase two, and then manually refined these groups up to 16 in phase three. By identifying and partitioning the two weakly connected components from phase two as described, we subdivided those two groups into five new groups, bringing the total number of groups up to 13 (10 original – 2 weak + 5 new = 13 total). In the process, we also added 3 miscellaneous targets to the collection of 13 already present from phase two of [6]. As seen in Fig. 3, weakly connected components had associated some targets that bore little visual similarity. But after partitioning, such targets were separated into new groups with stronger image association.

When KL finds a partition on a graph, the induced subgraphs in each partition are not guaranteed to be connected. This is why we can obtain > *2n* new groups (new connected components) when partitioning *n* weakly connected components (e.g., Fig. 2 B-1 & B-2, disjoint connected components of Partition B).

None of the five new groups qualify as weakly connected, as they all have NTC ≥ 2. However, with a slightly different dataset, partitioning could result in *new* weakly connected (albeit smaller) components. Iteratively partitioning such components could increase the total number of groups further and likely improve image association within groups. This approach may also increase the geologic usefulness of the classification, as the higher number of groups would be closer to that of the classification obtained in phase three of [6], used for interpretations by [e.g., 9, 11, 12].

Conclusions: Automatic partitioning of weakly connected components is an effective method for improving ChemCam target grouping by increasing the number of groups and the target-image association within groups, given a graph-based classification, such as [6]. Iteratively identifying and partitioning weakly connected components may help to match the quality of image-association within groups achieved in the manual third phase of [6].

Acknowledgments: NASA Mars Exploration program and CNES, France.

References: [1] Grotzinger et al. (2014) *Science, 343(6169),* [2] Grotzinger et al. (2015) *Science, 350(6257),* [3] Mangold et al. (2016) *JGR, 121*(3):*353-387*, [4] Cousin et al. (2017) *Icarus, 288:265-283*, [5] Sun et al. (2018) *Icarus, 321:866-890*, [6] Essunfeld et al. (2021) *LPSC 52 #2180*, [7] Wiens et al. (2012) *SSR, 170:167–227*, [8] Maurice et al. (2012) *SSR*, *170:95-166*, [9] Comellas et al. (2021) *LPSC 52 #2176*, [10] Hagberg, Schult, Swart (2008) *SciPy (12)*, [11] Comellas et al. (2022) *this conference,* [12] Morris et al. (2022) *this conference*, [13] Diestel (2005) *Graduate Texts in Mathematics, vol. 173 (12)*, [14] Kernighan, Lin (1970) *Bell System Technical Journal, 49:291–307.*

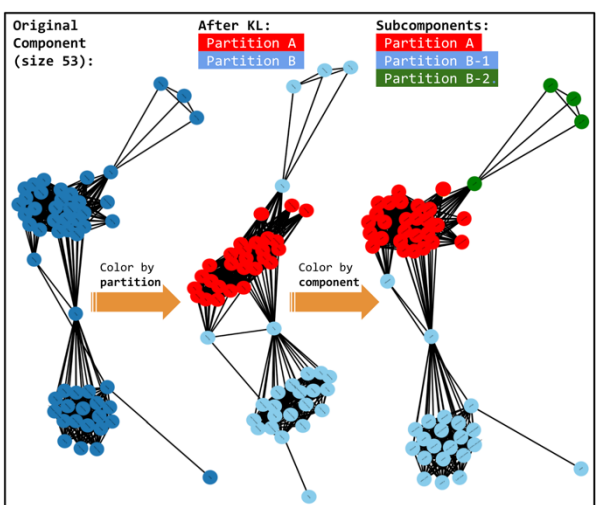

Figure 2: The weakly connected component of size 53 from the original classification graph by [6]. Left: the original component, uncolored. Middle: coloring applied by performing KL and obtaining partitions A (red) and B (light blue). Right: further coloring on disjoint connected components of partition B. (Partition A happened to be connected.)

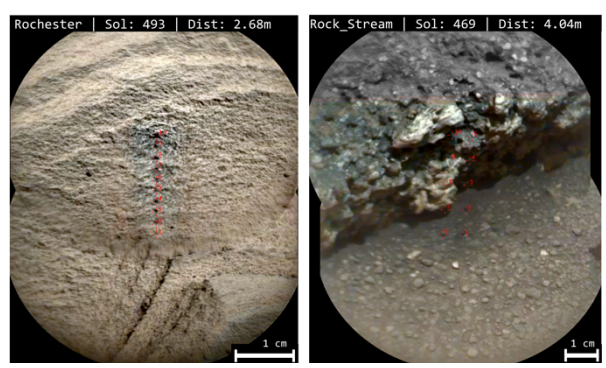

Figure 3: RMI mosaics of two ChemCam targets from the weakly connected component of size 37 [6]. Despite differences in texture and tonality, these targets were originally sorted into the same dumbbell-like group [6]. Because they were on opposite sides of this dumbbell-like component, they were separated into new groups by partitioning.