Using Hierarchical Clustering to Distinguish Primary Craters from Secondary Craters to Improve Surface Age Determination

Ali Bramson

Keywords: primary and secondary craters, hierarchical clustering, networks, dendrogram

Background:

Crater statistics are a fundamental tool for learning about the geology and surfaces of planets, as well as the population of bodies in the solar system doing the impacting. Without physical samples, crater densities remain the only way we can estimate absolute ages of planetary surfaces [1]. Impacts generally are considered to happen randomly in time and space, so the more craters a surface has, the older it is. The size-frequency distribution (SFD) of craters follows an inverse power law [2], so the number of craters increases with decreasing crater diameter. This means smaller craters are more statistically significant when used for calculating surface unit ages [3]; however doubts have recently emerged about how these smaller craters should be interpreted.

Secondary craters form when the ejecta from a large crater re-impacts the surface. These secondary craters lead to higher than expected crater counts within a geologic instant [2], which presents a problem for age determination because it makes a surface appear artificially older. Most of these secondaries appear close to their primary in clustered patterns such as rays and so can be excluded from crater counts. However, they often can be hard to distinguish from background small primary craters. This is especially true if they re-impact far from their source (at high velocities), in which case they match the circular morphology and depth to diameter ratios of primary craters.

A detailed study of the Europan surface by Bierhaus et al. 2005 shows that secondary craters comprise up to 95% of craters with diameters < 1 km. The ability to extract these secondary craters from the background primary crater density is essential if we wish to accurately constrain the ages of surfaces. The amount and quality of imaging data from recent planetary missions (such as the Mars Reconnaissance Orbiter, Mercury Messenger, and Lunar Reconnaissance Orbiter) drive the need for a technique that can not only identify secondary craters, but also work on the very large data sets we are currently receiving.

Fortunately, there is a growing body of work in computer science, devoted to understanding networks and clustering, which can assist in the analysis of these data. In particular, social networking algorithms have been developed to better understand how people in different groups are connected [4], and I have experience from my undergraduate senior thesis in applying such techniques to astronomy (galaxy groups). I propose to combine my interest in computer science and passion for planetary science by applying hierarchical clustering techniques common to network science to find patterns in spatial locations of craters. Applying these techniques to analyze the patterns of crater locations is a logical project that could prove to be very insightful, helping us to distinguish primary craters from secondary.

The Algorithm:

The hierarchical clustering technique [5] iteratively builds a hierarchy of clusters by starting with all nodes (the craters in question, in my case, or people in the social science setting) as separate groups and connecting them until they are all in the same group (agglomerative). To connect these features, it creates a distance matrix, a matrix of "similarity", where each element will be the distance between two craters. In a social setting, the distance matrix could use number of friends in common as a measure of similarity. The agglomerative technique then merges the two nodes that are the closest into one cluster. These nodes are removed from the distance matrix while our new cluster is added in, and the process is iterated until all objects have been merged into one cluster.

Bierhaus et al. 2005 utilized aspects of this algorithm, combining them with other clustering and statistical techniques in his study of Europa's surface. However, Europa is a simpler case because it is sparsely cratered compared to other surfaces, making it easier to disentangle primary from secondary craters [3]. I propose to apply this algorithm to the more complicated Martian surface, using the latest database of impact craters on Mars from Robbins & Hynek 2012 [6,7].

Data:

I will use a new global Martian (covering >99% of the surface area of the planet) crater database from which is freely available via the U.S. Geological Survey's (USGS) Mars Crater Consortium, for my study [6,7]. This database contains 384,343 craters and is statistically complete down to diameters ≥ 1km. The database does have secondary crater classification for a select section of the database [1,8]. This will allow me to compare my technique to the manual classification of Robbins and Hynek. This comparison can be used to assess the accuracy of my automatic methods.

Expected Results:

I will use the output of the hierarchical clustering to yield a tree-like "dendrogram" to show the order in which the craters have been connected. Rosolowsky et al. (2008) have used dendrograms to determine the structure of molecular clouds. Where the branches of their dendrograms correspond to self-gravitating molecular cloud structures, the branches of my dendrograms would correspond to larger networks of craters. The shape of these branches can tell us about the spatial patterns of the craters; if the dendrograms are flat, with many connections at equal distances, the features are distributed homogenously. Long connections high in the dendrogram connecting groups of low-level links indicate the features spot the surface of the planet in a much less uniform manner (i.e. we can automatically identify clusters). I will also create dendrograms for a random distribution of objects of equal density as a comparison. I predict that the comparison to random will extract the primary craters that dot the surface randomly, leaving behind the probable secondaries that will appear in groups of low-level links that deviate from random.

The advantage of this technique is that it could be used in any field that involves clustering and connections e.g. creating dendrograms of asteroid proper orbital elements could lead to new asteroid groupings. We might also expect the clustering of surface features on Io to be connected with volcanic events or other asymmetries between the leading and trailing hemispheres of tidally locked moons. The broad applicability across many disciplines is a key part of the broader impact and intellectual merit of this research. While I am passionate about planetary science and this research's applicability within the field, the algorithms and techniques I use will be valuable to many fields.

References: [1] Robbins, S.J. & Hynek, B.M. 2011, *GRL*, **38**, L05201. [2] McEwen, A.S. & Bierhaus, E.B. 2006, *Annu. Rev. Earth Planet. Sci*., **34**, 535. [3] Bierhaus, E.B., Chapman, C.R. & Merline, W.J. 2005, *Nature*, **437**, 1125. [4] Mucha, P.J., et al. 2010, *Science*, **328**, 876. [5] Newman, M.E.J. & Girvan, M. 2004, *Phys. Rev. E*, **69,** 026113. [6] Robbins, S.J. & Hynek, B.M. 2012, *JGR*, **117**, E05004. [7] Robbins, S.J. & Hynek, B.M. 2012, *JGR*, **117**, E06001. [8] Robbins, S.J. & Hynek, B.M. 2011, *JGR*, **116**, E10003. [9] Rosolowsky, E.W. et al. 2008, *ApJ*, **679**, 1338.