Clustering of Mining Paths

Miguel A. Cuba, Jeffery Boisvert, and Clayton V. Deutsch

In the SLM paradigm, multiple scenarios of the mining of the deposit are generated. From a practical perspective, this set of scenarios is unfeasible to be used in mine design due to the problem of having to analyze a large set of alternatives. Mine design is a complex process that involves aspects that cannot be fully automated or modeled with mathematical approaches, such as management decisions. Due to this complexity, only a reduced number of the more important alternatives are evaluated and only one of them is implemented. In this paper, a methodology to summarize the set of multiple scenarios into a reduced number of representative alternatives is presented. Major trends in the set of multiple scenarios are identified by implementing adapted clustering methodologies. An example is presented to illustrate implementation aspects.

Introduction

The implementation of the simulation paradigm in mine planning is impractical, as it requires the evaluation of a large number of mining sequence alternatives (Dominy, Noppé, & Annels, 2002). Similar to the Simulation paradigm, in the Simulated Learning Model (SLM) paradigm (Cuba, Boisvert, & Deutsch, 2010), a set o mining scenarios that account for the effect of the periodical evolution of the degree of knowledge of the deposit are generated. To make the SLM paradigm an alternative that can be used in mine design it is necessary to present the simulated mining scenarios in a condensed form. By identifying major trends in the set of mining scenarios, only a reduced set of options, that are representative of the simulated set, are necessary to be evaluated in the mine design process.

Clustering techniques are widely used in different fields of study, including biology, information retrieval, and business, to classify data based on certain similarities (Tan, Michael, & Vipin, 2006). Based on the type of implementation of the clustering techniques, clusters provide an abstract representation of the object elements they represent. In this paper, an approach to identify major trends in the set of simulated mining scenarios is proposed. This approach is based on adapted clustering techniques.

First, the set of generated mining regions in the SLM paradigm as set of object elements for clustering is discussed. Next, three semi-metric of comparison to cluster the mining scenarios are discussed. In the following section, a graph-network that represents the clustering structure of set of mining scenarios is presented. Next, the methodology is discussed. In the last section, an example to illustrate the implementation is discussed.

SLM Simulated Mining Scenarios as Dataset for Clustering

In many fields, including social sciences, biology, and business, clustering analysis is used to divide datasets into meaningful groups, named clusters, by capturing the natural structure present in the data (Tan, Michael, & Vipin, 2006). For implementation purposes, clustering techniques are divided into different categories based on their applications. There is extensive documentation about applications of different types of clustering methodologies, including (Hartigan, 1975), (Kaufman & Rousseeuw, 2005), and (Tan, Michael, & Vipin, 2006). In the case of simulated mining scenarios, the proposed methodology is built based on hierarchical clustering techniques, mainly because it is not required to know the number of clusters prior to its implementation. Another important advantage of hierarchical clustering is its ability to characterize the relationship between data elements at fine and coarse scales (Fielding, 2007). This second characteristic is of particular importance in the evaluation of the major trends of the mining scenarios. In hierarchical clustering, all the possible numbers of clusters are explored gradually (Kaufman & Rousseeuw, 2005). Hierarchical techniques are divided into agglomerative or divisive. In the agglomerative techniques, each data element is considered to be one cluster, and elements are merged until one cluster is formed. In the divisive techniques, all elements start in one cluster, which is divided until each element is in its own cluster. The characterization of clusters between the agglomerative and divisive techniques is usually different (Kaufman & Rousseeuw, 2005).

A simulated mining scenario consists of a set of regions positioned sequentially in the deposit that follow a mining path. Two important aspects that define the mining path are: 1) the geometric configuration of the mining regions and 2) the timing in which the mining regions are to be executed. The geometry of the mining regions in each period is quite variable, as it may consist of one region or a group of sub-regions (see Figure 1). The variability in the geometry the mined regions depends on the operating constraints of the mine sequencing algorithm implemented. The timing aspect defines the order sequence in which mining regions are positioned in the deposit. From the clustering perspective, the set of multiple scenarios of the mining of the deposit are a highly dimensional dataset. The majority of clustering techniques consider that the input set of objects consists of multidimensional points in the Euclidean space (Halkidi & Vazirgiannis, 2008). Because of the geometric complexity of the mining regions, the scenarios of the mining of the deposit cannot be adequately represented as point-like elements. The clustering techniques to be implemented are required to be adapted before being able to deal with the set of mining scenarios.

The goal of the proposed methodology is to summarize the set of simulated mining scenarios into a reduced set of generic mining paths or major trends that are more likely to occur, based on the mining and data collection strategies adopted. This reduced set of major mining trends presents a more reasonable set of alternatives to be evaluated in the mine design process than the whole set of simulated mining scenarios itself. In cluster analysis, prototypes are representative elements of the subset object elements that reside in each cluster. The major mining trends are considered as prototypes of the mining scenarios. In this paper, two important aspects in the definition of the major mining trends are: 1) the metric of comparison, and 2) the definition of the prototype. The metric of comparison is the measure of performance with which the accuracy of the prototype is quantified. The definition of the prototype considers the framework in which the relevant information of the major mining trends is structured.

Metric of Comparison

The definition of clusters depends on the metric of comparison used. A large part of the research on hierarchical clustering is focused on the derivation of the metrics of comparison, as they have a big impact on the clustering performance (Castro, Coates, & Nowak, 2004). The comparison between mining scenarios is done by considering both the geometric and the timing aspects. The variability in the lifetime of the mining scenarios adds some complexity to the comparison. For simplicity, the metric of comparison between mining scenarios is based on the individual comparison of the mining regions in each period within a specified range of periods that is common for the majority of the mining scenarios. In cluster analysis, the metric of comparison to be used have to satisfy certain conditions. In the case of a measure of dissimilarity, four conditions have to be met to make the metric of comparison licit:

1. The metric is zero or positive,
2. The magnitude of the metric is invariant regardless of the direction,
3. The metric is zero when two elements are identical,
4. When comparing three element objects, the magnitude of the metrics should behave as the sides of a triangle.

In certain cases, semi-metrics are used instead of full metrics. Semi-metrics have to satisfy only the first three conditions (Fielding, 2007). Since the elements to compare are volumetric objects, the fourth condition is difficult to meet. In this section, three alternatives of comparison are discussed. The first alternative is calculated in terms of the average distance between the respective discretized locations of the mining regions. In the second alternative, the mining regions are also discretized, but the measure of comparison is calculated in terms of the average distance between the discretized points and the other volume. The third alternative is calculated based on a metric of similarity. The viability of these three alternatives is discussed as follow:

The first alternative d_1 is calculated in terms of the averaged distances between the respective discretized points of the mining regions. It is expressed as:

$$d_1\left(V_a,V_b\right)=\frac{1}{n_a+n_b}\sum_{i=1}^{n_a}\sum_{j=1}^{n_b}\left\|\mathbf{u}_i-\mathbf{u}_j\right\|,\ \forall\mathbf{u}\in V_a,\mathbf{v}\in V_b, \tag{1}$$

where, V_a and V_b are the two mining regions being compared, $\mathbf{u}$ and $\mathbf{v}$ are the locations of discretized points within V_a and V_b, respectively, and n_a and n_b are the number of discretized locations in V_a and V_b, respectively.

Because of the discretization of the two mining regions, this alternative accounts for the differences in the shape of the two regions. However, in the case that the two regions have the same shape and are overlapped, the measure of this alternative is greater than zero, as all the individual distances of the combination of discretized points within the two regions are considered. Thus, because the similarity aspects are not properly accounted for in the comparison, this alternative is not an effective semi-metric.

Like the first alternative, the second alternative also considers the discretization of the two mining regions. The second alternative differs from the first alternative in the way the averaged distances between the discretized locations are calculated. In each mining region, the individual distances are calculated from its discretized locations to the volume of the other mining region. This calculation is done twice, one for each mining region. The second alternative d_2 is calculated based on the two sets of individual distances, d_a and d_b, as presented in equation (2). The calculation of the individual distances is approached as the minimum distance between each discretized location and the set of discretized locations of the other mining region. The second alternative is expressed as:

$$d_s\left(V_a,V_b\right)=\frac{d_a+d_b}{n_a+n_b}, \tag{2}$$

where,

$$d_a=\sum_{i=1}^{n_a}\min_{\mathbf{v}\in V_b}\left\|\mathbf{u}_i-\mathbf{v}\right\|,\ \forall\mathbf{u}\in V_a,$$

$$d_b=\sum_{j=1}^{n_b}\min_{\mathbf{u}\in V_a}\left\|\mathbf{v}_j-\mathbf{u}\right\|,\ \forall\mathbf{v}\in V_b.$$

The problem that the similarity between the two regions is not properly accounted for by the first alternative is solved in the second alternative. In the case that the two regions have the same shape and are overlapped, the measure of this alternative is zero, as all the individual distances are also zero. However, if the mining regions are fragmented, the second alternative may tend to underestimate the dissimilarity. In the example presented in Figure 2, because the minimum individual distances are considered, the d_a and d_b terms tend to have smaller values. The second alternative performs better in cases where the mining regions are not fragmented.

The third alternative d_3 is calculated based on the intersection volume between the two mining regions. The dissimilarity is calculated as one minus the standardized intersection volume, where the standardization factor is the average volume of the two mining regions. The third alternative is expressed as:

$$d_3\left(V_a,V_b\right)=1-\frac{2\left(V_a\bigcap V_b\right)}{V_a+V_b}, \tag{3}$$

Since the intersection volume is considered, this alternative solves the problem of the second alternative of reducing the dissimilarity when fragmented regions are compared. However, one potential drawback is that when the two mining regions are completely different, this alternative does not quantify how different they are, as it occurs in the other two alternatives. In case two mining regions are completely different the value of this alternative is one.

Due to the long-term scale the mining scenarios, the semi-metric used is not required to be strict in comparing the mining regions. The objective of the clustering, in each period, is to identify general regions in the deposit where the simulated mining scenarios target the zones where to mine. The definition of the clusters is sensitive to the type of the semi-metric used. In Figure 3, an example of two similar mining regions in a specific period is presented. In case the third alternative is used, only the effect of the sub-regions located in the middle of the deposit is quantified, and not the effect of the east sub-regions, as these two sub-regions has a very small overlapping zone. This aspect is also present, to a lesser extent, in case the second alternative is used. The first alternative is not affected. To make the comparison less strict, a buffer zone can be added to the mining regions. The buffer zone acts in a similar way as a bounding zone where the mining region can shift its position.

The three semi-metrics discussed are used to compare pairs of individual mining regions. Throughout the implementation of hierarchical clustering, intermediate clusters of mining regions are compared to each other, or to individual mining regions. Three typical approaches to compare clusters are: 1) MIN, 2) MAX, and 3) group average. These approaches are also referred to as graph-based definitions of cluster proximity (Tan, Michael, & Vipin, 2006). Each of these approaches is calculated based on the pairwise metric or semi-metric used to compare the individual object elements. MIN considers the two most similar object elements of the two clusters. MAX considers the two most different object elements of the two clusters. The group average approach considers the averaged distance between all the object elements of the two clusters. In the case of the mining regions, the group average approach performs better, as the whole set of elements within each cluster is considered.

Definition of the Mining Sequence Prototype

The clustering evaluation may become difficult if whole mining scenarios are considered object elements. This makes that the comparison between pairs of mining scenarios is quantified in a single value that comprises all the individual comparisons of their respective mining regions altogether. Thus, the variability in certain periods may mask similarity in other range of periods. The use of weighting factors to add flexibility to the comparison, giving more or less importance to specific range of periods, have the same effect, as patterns in the mining scenarios can be hidden due to not knowing how the mining scenarios behave on a period basis. In Figure 4, an example of two mining scenarios is presented. In a direct comparison, these two mining scenarios result different to some degree. The direct comparison hides specific aspects such as the similarity of the two mining scenarios in periods 1, and 4 to 6. This specific information could be obtained in the direct comparison by weighting specific periods, which adds more complexity to the evaluation in terms that different weighting configurations would have to be explored. Instead, a representation of the two mining scenarios in the form of a graph-network provides a more informative picture of how the two mining scenarios are related. The graph network is built based on the period base comparison of the respective mining regions.

Similar to the example presented, the mining scenarios behave like intertwined threads within the range of periods evaluated. In the proposed methodology, the mining sequence prototype has the form of a graph network, where the nodes represent a set of potential regions to be mined in each period, and the connections are associations of the mining sequence that are likely to occur (see Figure 5). The mining sequence prototype behaves as a flow-network, where the mining scenarios move throughout the network choosing among different the alternatives that are presented in each period. Since the decision to mine the first period consists of only one alternative, there is one node at the beginning of the graph-network.

The major trends are identified as branches in the graph-network in which nodes have more occurrences and more number of associations. In Figure 5, the main trends form a sub-network with the nodes that are more likely to occur. This sub-network represents a decision structure within the graph-network that is supported by a majority of the mining scenarios.

The nodes are calculated by clustering the mining regions in each period. The connections are calculated from the associations between the clustered mining regions that are present in the mining scenarios. In Figure 6, an example of the implementation of an agglomerative hierarchical clustering technique is presented. The dataset of the example consists of point data with two attributes. Three clusters are identified in the example. In hierarchical clustering, the number of clusters depends primarily on the definition of the clustering threshold. The definition of the number of clusters has a certain degree of subjectivity. Milligan & Cooper (1985) reviewed over thirty techniques for defining clusters to conclude that the performance of some criteria depends on the nature of the dataset. A small clustering threshold results in accurate clusters, and a large clustering threshold makes the clusters to be more generic. The difference between accurate and generic clusters is the degree of dispersion of the object elements in their corresponding clusters. In the proposed methodology, after deciding the semi-metric of comparison to use, the clustering of the mining regions depends on three aspects: 1) range of buffer zones, 2) definition of the clustering threshold, and 3) definition of minimum number of object elements per cluster. The range of the buffer zone makes the comparison between mining regions more generic, by artificially reducing the magnitude of the semi-metric of comparison, in relative terms. The clustering threshold defines how strict is the definition of clusters. The minimum number of object elements sets the minimum requirement to form a cluster. The clusters that do not satisfy this minimum requirement are categorized as outliers.

The mining sequence prototype is required to be simplistic. The simplicity is related to having a small number of nodes in each period. The presence of a large number of nodes tends to increase the complexity of the associations in the graph-network. The quantity of nodes in each period depends on how strict the clustering of the mining regions is implemented. The simplification of the mining sequence prototype by reducing the number of nodes tends to reduce the performance to characterize the behaviour of the mining scenarios. The performance of the mining sequence prototype is measured in terms of the dispersion of the mining regions in each cluster node of the graph-network. A measure of the dispersion δ of a node C is:

$$\delta_C = \frac{1}{n_c^2}\sum_{i=1}^{n_c}\sum_{j=1}^{n_c} d_{ij}\left(V_i, V_j\right),\ \forall V \in C, \tag{4}$$

where, d is a metric or semi-metric of comparison, V is the volume of the mining region in cluster C, and n_c is the number of object elements in cluster C. Alternative metrics based on the comparison of the cluster prototype and the corresponding mining regions, e.g., mean-absolute-error or mean-squared-error, are less practical in this case, since due to the volumetric nature of the mining regions, the cluster prototypes require the definition of additional parameters to be characterized.

The complexity of the graph network due to the variability of the mining scenarios depends highly on the real existing dataset. The mining scenarios have more freedom to vary as the uncertainty in the deposit is high. As more drilling data is added to the initial dataset, the simulated mining scenarios converge into a unique mining scenario. Another aspect that affects the variability of the mining scenarios is the mining strategy implemented. For example, the set of mining scenarios is more rigid if operating mining conditions do not allow too much fragmentation of the mining regions. The data acquisition strategy does not have much impact on the variability of the mining regions, as it depends on simulated information. In case the mining sequence prototype is complex in certain periods, the merging of consecutive mining regions may be considered as an option to reduce the number of clusters. The merging of consecutive mining regions reduces the restriction that only mining regions that are in the same period can be compared. For example, in Figure 4, if periods two and three are merged, the two mining sequences become identical. The cost of merging consecutive mining regions is the loss of detail in the mining sequence prototype.

The mining sequence prototype is similar to a road map of the mining sequence decisions to mine the deposit. It does not deliver geometric information of the major trends identified. The representation of the major trends as spatial information, similar to a mining sequence, is done by positioning sequentially the prototypes of the clustered mining regions, following the branches of the major trends in the mining sequence prototype. Because of the inherent dispersion of the clustered mining regions, the spatial representation of the major trends does not have a operating geometric configuration. The purpose of this representation is the identification of the most probable sequences of regions targeted throughout the lifetime of the mining project.

Methodology

The proposed methodology consists of three steps: 1) select initial clustering parameters, 2) refine the mining sequence prototype, and 3) representation of major trends. The description of each step is as follows:

Step 1:

Three clustering parameters to define are: 1) the semi-metric of comparison, 2) range of buffer zone, and 3) clustering threshold. Before implementing the preliminary clustering, it is required to select the semi-metric of comparison to use. This decision is based on the geometric configurations of the mining regions in the mining scenarios. In this paper, semi-metrics 2 and 3 are the better alternatives to choose. In case of fragmented mining regions, alternative 2 is recommended. Alternative 3 is preferred if the mining regions consists of one region. The range of buffer zone is used to make the comparison between mining regions more generic. This parameter is used to reduce the number of clusters. The clustering threshold is used to set the degree of strictness of the initial clustering. The initial mining sequence prototype is calculated based on the definition of the three clustering parameters. These are parameters are set constant for the period range evaluated. It's recommended to start with a strict configuration of the clustering parameters and explore less strict configurations focusing on reducing the complexity of the mining sequence prototype.

Step 2:

This step consists of refining the mining sequence prototype generated in step 1. The definition of constant clustering parameters may result in an inappropriate definition of clusters is certain periods. In this step, each period is visited and the mining regions are compared versus their corresponding cluster prototypes. From the visual evaluation of the clusters it is decided whether to make the clustering more generic to reduce the number of clusters or to make it more strict to reduce the dispersion of the mining regions within the clusters. The tuning of the clustering characterization is done only by varying the clustering threshold. The semi-metric of comparison and the range of the buffer zone are not modified. An appropriate number of clusters per period is four per period. The number of clusters does not include the outliers. In case the number of clusters cannot be reduced, it is considered to merge consecutive periods. The range of periods with more number of clusters is combined into one period.

Step 3:

The major trends in the mining scenarios are identified from the mining sequence prototype. A threshold value that specifies the proportion of mining regions to be represented, e.g., 75%, is used to identify the sub-network of major trends. A major trend consists of one combination of clustered mining regions following the sub-network decision structure. Each major trend can be represented spatially as if it was the mining sequence of one simulated mining scenarios. The graphical representation of a major trend allows illustrating one sequence of regions that is likely to be followed during the mining of the deposit.

The decision about the quality of the clusters, to properly classify the mining regions, depends on the expert judgement of the practitioner. The proposed methodology allows a flexible tuning of the clustering of the mining regions, via the definition of the clustering parameters.

Example

The initial dataset consists of twenty eight vertical drillholes placed over a regular grid of 50 x 50 m. Based on the initial dataset, the SLM paradigm is implemented to generate a set of one hundred mining scenarios. The infill drilling program implemented consists of eight infill drillholes in each period. The objective of the mining strategy is to preserve a production of 2500MT per period.

In the first step, because fragmented regions are allowed in the mining strategy, the third semi-metric is used to cluster the mining regions in each period. As part of the evaluation of the buffer zones, different set of values are explored. In Figure 7, four cases of buffer zones are presented, 20 to 35m. The clustering of the mining regions tends to be more generic as the range of the buffer zone increase. In the four cases, the clustering threshold is set to 0.4, based on a preliminary revision of the clustering of the mining regions. In Figure 8, the case of period 5 using a buffer zone of 20m is presented. Clusters 3 and 4 are combined if the clustering threshold is larger than 0.416. In this example, based on both prototypes, it is considered that these two clusters are too different to be combined. A reduction of the clustering threshold below 0.31 results in an increase in the number of clusters.

In the second step, the 25m buffer zone case is considered as a candidate for the mining sequence prototype. The main reason for this decision is that the shifting tolerance of the mining regions is set to half of the average of the initial drilling campaign. Beyond this limit, mining regions are considered completely different. Although the 30 and 35m cases present simpler graph-networks, the clustering of the mining regions is too generic with respect to the initial drilling data. The complexity of the 25m buffer zone case is high in periods 4 and 5. After merging these two periods, the clustering is simplified resulting in three clusters (see Figure 9). As the combined volumes are larger, the clustering is done based on a more strict set of parameters, e.g., clustering threshold 0.2. The impact in the graph-network is the simplification of the mining sequence prototype (see Figure 10) with respect to the initial candidate.

In the last step, the branches with major number of associations are evaluated in the mine design process. In Figure 11, one example of the branch with more number of occurrences is presented. Similarly, the other variations of this major branch can be identified from the mining sequence prototype.

Concluding Remarks

In this paper, an approach to condense the simulated mining scenarios of the SLM paradigm is proposed. This condensed form is more suitable to be used in the mine design stage. The possibility to use the SLM mining scenarios in the mine design stage allows accounting for uncertainty in the overall evaluation of the mineable reserves.

The proposed methodology is easy to implement, since it relies on agglomerative clustering techniques. The use of the different semi-metrics of comparison and buffer zones gives high flexibility in its implementation. The mining sequence prototype provides an easy interpretation of the mining alternatives present.

The mining sequence prototype can be used as a guideline to design cost-effective infill drilling campaigns. Infill drilling can be implemented to target the nodes in the mining sequence prototype to confirm or invalidate their occurrence. The goal of the infill drilling campaign is to reduce the possibilities of the mining sequence prototype to reduce the number of branches.

References

Castro, R., Coates, M., & Nowak, R. (2004). Likelihood based hierarchical clustering. *Signal Processing on IEEE Transactions* , 2308-2321.

Cuba, M. A., Boisvert, J. B., & Deutsch, C. V. (2010). A New Approach to Forecast Long Term Recoverable Reserves: The Simulation Learning Model. *Centre for Computational Geostatistics - Report Twelve* , p. 415.

Dominy, S. C., Noppé, M. A., & Annels, A. E. (2002). Errors and Uncertainty in Mineral Resource and Ore Reserve Estimation: The Importance of Getting it Right. *Exploration and Mining Geology , 11*, 77-98.

Fielding, A. H. (2007). *Cluster and Classification Techniques for Biosciences.* New York: Cambridge University Press.

Halkidi, M., & Vazirgiannis, M. (2008). Npclu: An Approach for clustering spatially extended objects. *Inteligent Data Analysis* , 587-606.

Hartigan, J. A. (1975). *Clustering Algorithms.* Toronto: John Wiley & Sons.

Kaufman, L., & Rousseeuw, P. J. (2005). *Finding Groups in Data.* Hoboken, New Jersey: John Wiley & Sons.

Milligan, G. W., & Cooper, M. C. (1985). An Examination of procedures for Determining The Number of Clusters in a Data Set. *Psychometrika* , 159-179.

Tan, P.-N., Michael, S., & Vipin, K. (2006). *Introduction to Data Mining.* Addison Wesley.

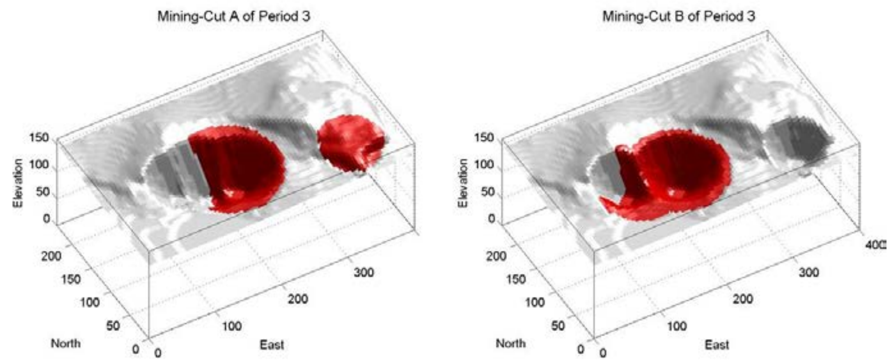

Figure 1: Two mining cuts, in the same period, with different geometric configuration

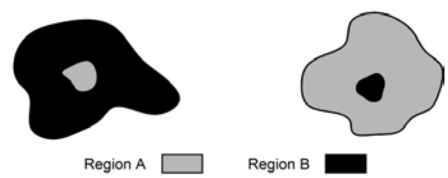

Figure 2: Case of two regions, A and B, where the second alternative semi-metric underestimates dissimilarity

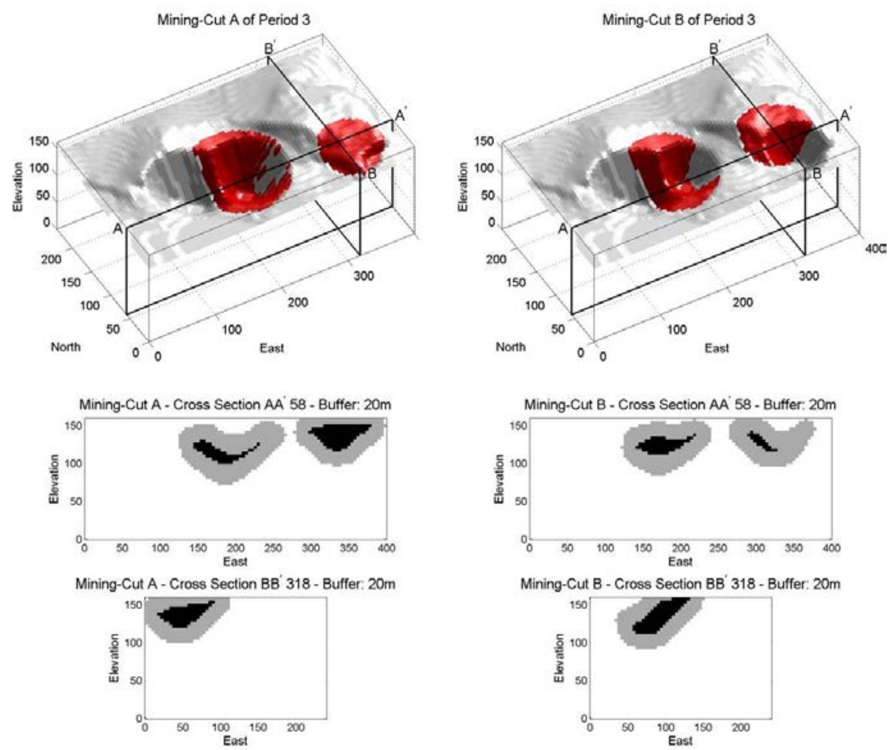

Figure 3: Sketch of the use of buffer zones in the comparison of two similar mining regions; the mining regions in the cross-sections are shown in black and the buffer zones in gray

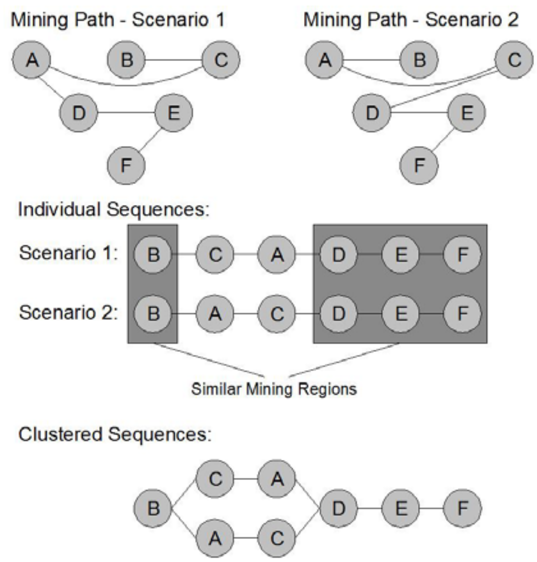

Figure 4: Example of a graph cluster representation of two mining scenarios

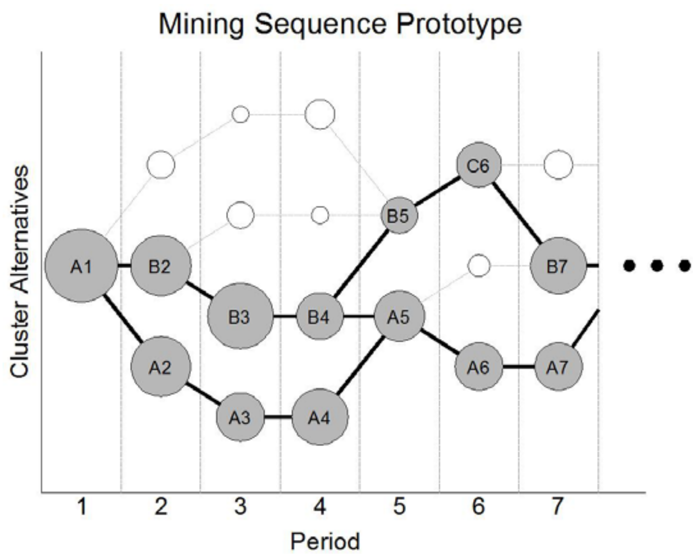

Figure 5: Sketch of a graph-network representation of a mining sequence prototype; the size of the circles represents the occurrence of the mining regions in each period, and the bold connected branches represent the major mining trends

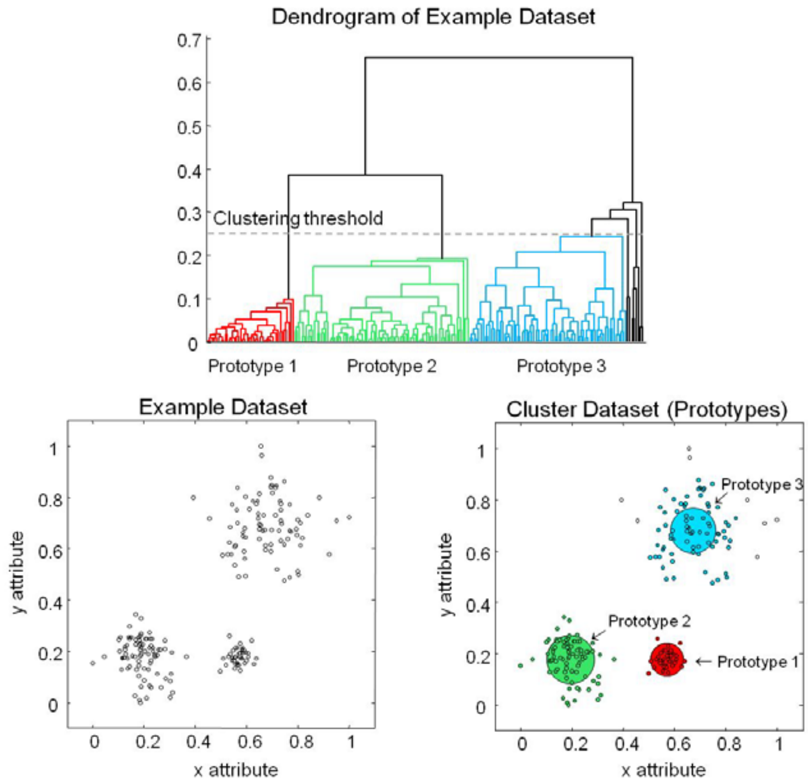

Figure 6: Example of agglomerative hierarchical clustering

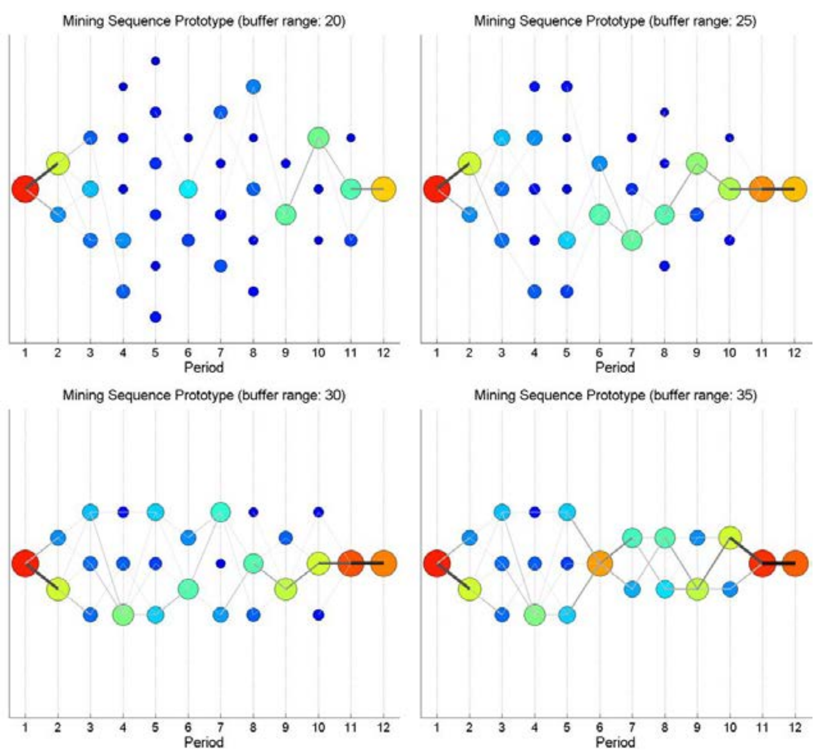

Figure 7: Four cases of initial mining sequence prototypes with different buffer regions and clustering threshold of 0.4; the size and color of the nodes represent the number of mining regions in each cluster

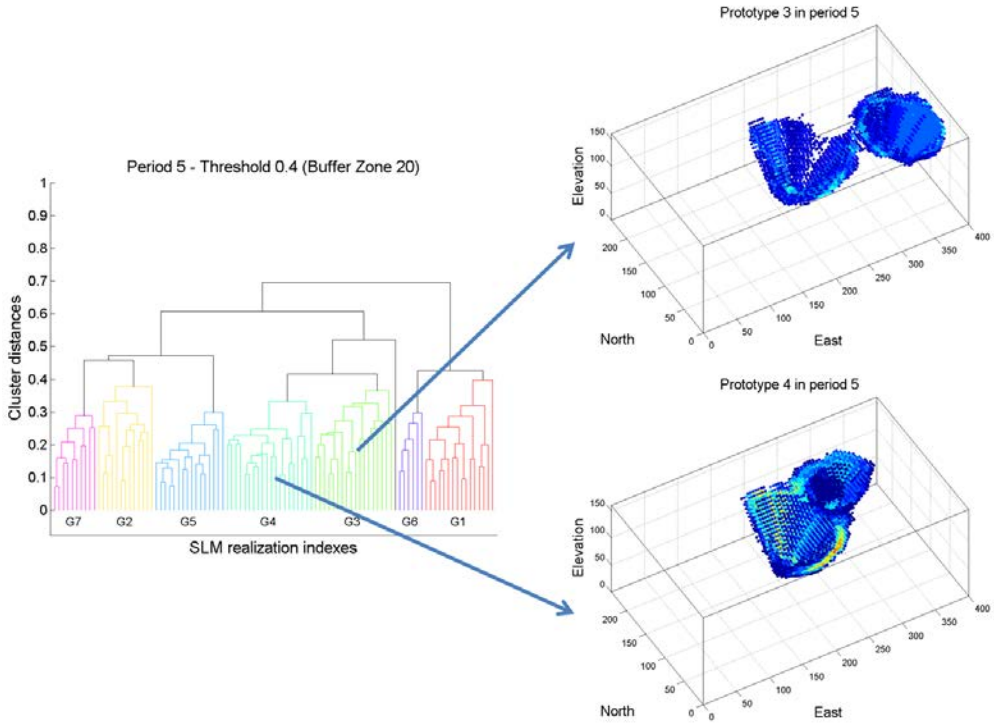

Figure 8: Definition of clusters 3 and 4 in period 5 with clustering threshold 0.4

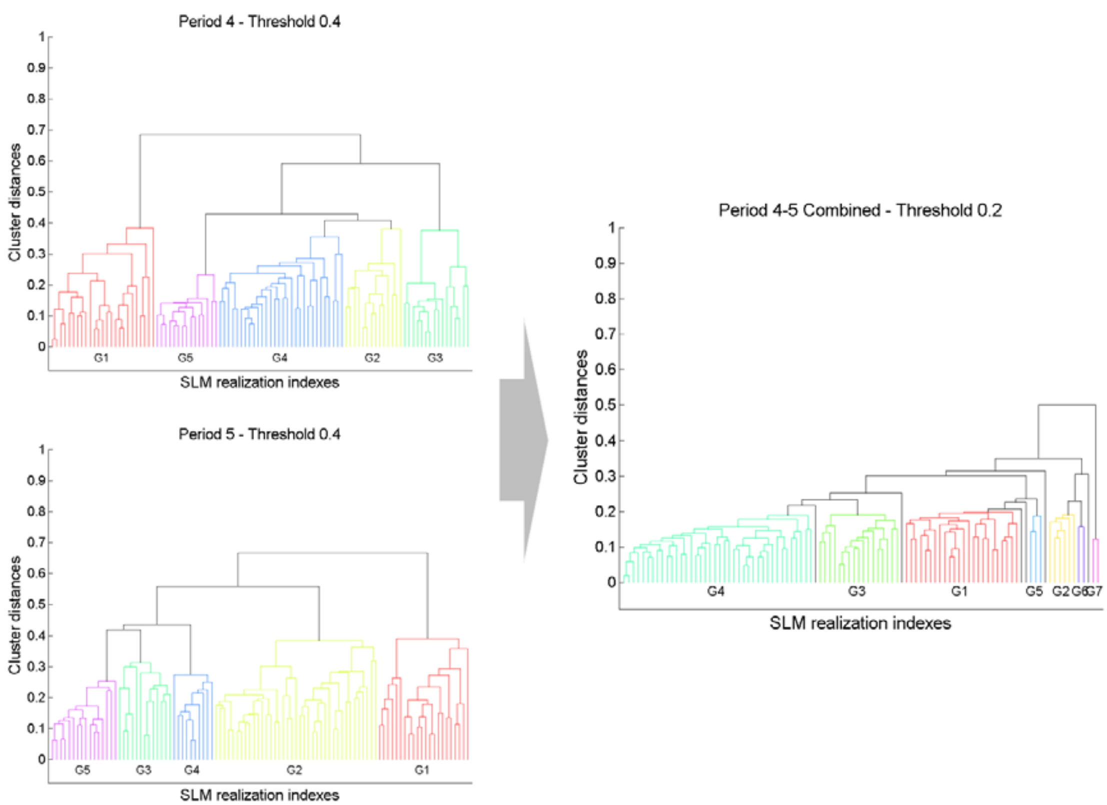

Figure 9: Effect of combining consecutive periods 4 and 5

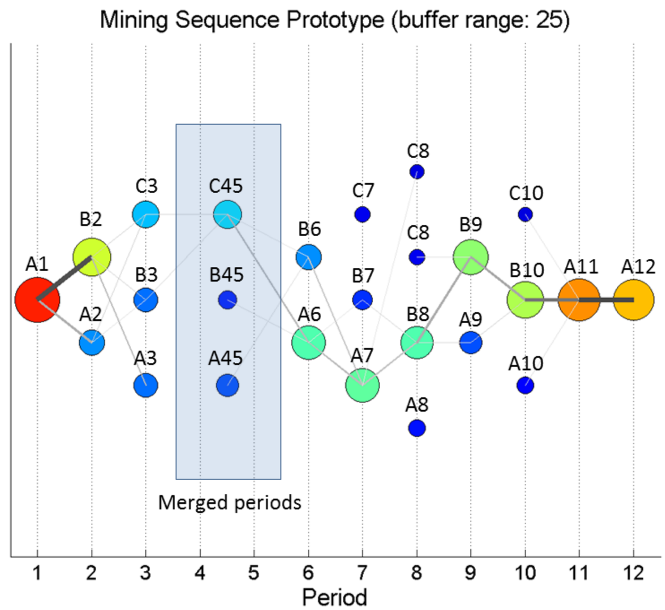

Figure 10: Mining sequence prototype

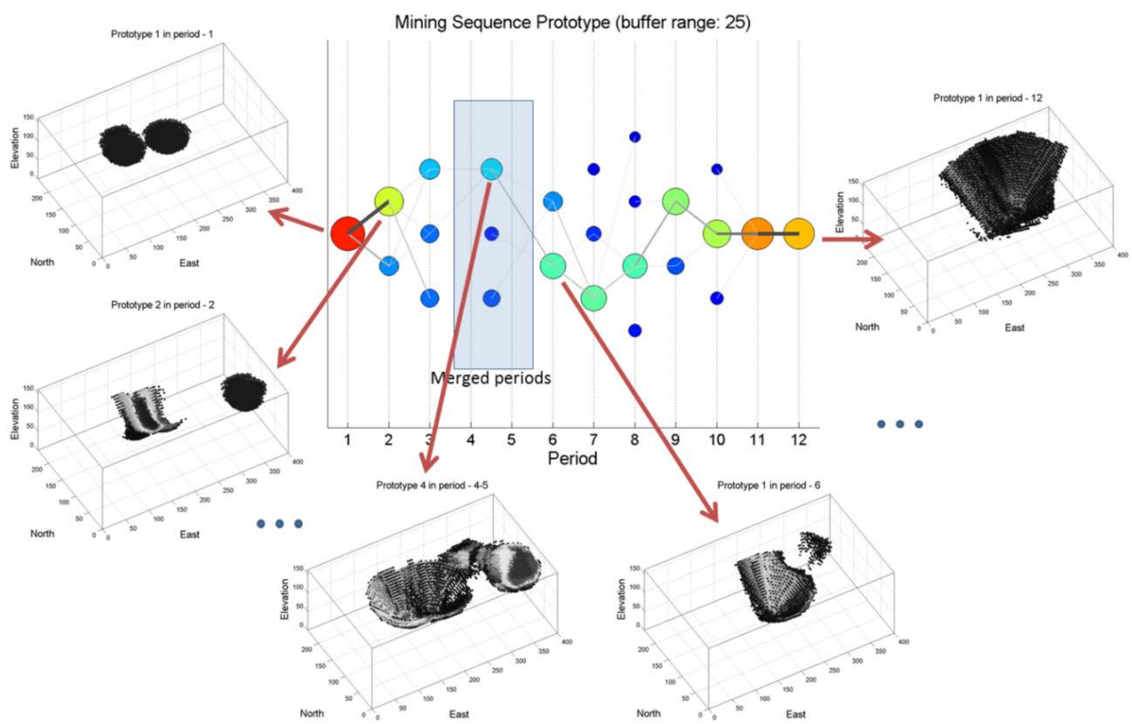

Figure 11: Representation of a major mining trend