JÚNIA ALENCAR DINIZ

NETWORK THEORY APPLIED TO DISCRIMINATE LAND USES AND EVALUATE MINED AREAS RECLAMATION

Dissertation submitted to the Soil Science and Plant Nutrition Graduate Program of the Universidade Federal de Viçosa, in partial fulfillment of the requirements for the degree of *Magister Scientiae.*

Adviser: Emanuelle Mercês Barros Soares

Co-advisers: Igor Rodrigues de Assis
Ivo Ribeiro da Silva
Marcelo Lobato Martins
Rafael Vasconcelos Valadares

VIÇOSA – MINAS GERAIS
2021

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T

D585n
2021

Diniz, Júnia Alencar, 1965-
Network theory applied to discriminate land uses and evaluate mined areas reclamation / Júnia Alencar Diniz. - Viçosa, MG, 2021.
1 dissertação eletrônica (95 f.): il. (algumas color.).

Inclui anexos.
Inclui apêndices.
Orientador: Emanuelle Mercês Barros Soares.
Dissertação (mestrado) - Universidade Federal de Viçosa, Departamento de Solos, 2021.
Referências bibliográficas: f. 80-82.
DOI: https://doi.org/10.47328/ufvbbt.2022.029
Modo de acesso: CRB6/2578.

1. Solos - Qualidade. 2. Plantas e solo. 3. Ferro - Minas e Mineração. 4. Bauxita - Minas e Mineração. 5. Áreas degradadas - Recuperação. 6. Teoria das redes complexas (Física). 7. Análise de componentes principais. I. Soares, Emanuelle Mercês Barros, 1979-. II. Universidade Federal de Viçosa. Departamento de Solos. Programa de Pós-Graduação em Solos e Nutrição de Plantas. III. Título.

CDD 22. ed. 631.42

Bibliotecário(a) responsável: Renata de Fátima Alves CRB6/2578

JÚNIA ALENCAR DINIZ

NETWORK THEORY APPLIED TO DISCRIMINATE LAND USES AND EVALUATE MINED AREAS RECLAMATION

Dissertation submitted to the Soil Science and Plant Nutrition Graduate Program of the Universidade Federal de Viçosa, in partial fulfillment of the requirements for the degree of *Magister Scientiae.*

APPROVED: October 25, 2021.

Assent:

Júnia Alencar Diniz
Author

Emanuelle Mercês Barros Soares
Adviser

To my father, from whom I would have liked to inherit at least half of his intelligence…

I dedicate.

ACKNOLEDGEMENTS

First, I would like to thank *Eckhart Tolle*, who taught me how to observe my mind and keep it under control. His teachings were fundamental for this work to flourish and bear fruit.

I also thank *Mariana Nascimento*, for having assisted me, after hours, on the day I needed it most; and *Maísa Carvalho*, for her dedication and patience in her professional care.

I am very grateful to my advisor *Emanuelle Soares*, for being so attentive and understanding in the face of the difficulties imposed on this research by the COVID-19 pandemic, and especially for her personal support, in such a humane and reassuring way.

To the *UFV*, for having welcomed me so many times and for always having been my greatest inspiration, and to the *Soil Department*, to which I am proud to belong, I give special thanks, and by extension, to my collaborators in this institution:

- to Professor *Ivo Ribeiro*, for going with me to the Physics Department at UFV, helping me to kick-start this work. Even without knowing it, he has always been my mentor, so that a good part of this dissertation, both in form and content, was influenced by his ideas;
- to Professor *Marcelo Lobato*, for having initiated me into Network Science and for giving me his approval on my work, even all the adaptations it suffered due to distance imposed on us by the pandemic;
- to Professor *Igor de Assis*, for the support and good will in collaborating with our work, from the first moment. And mainly for the agility in responding to all my requests, a quality that deserves to be highlighted at any time;
- to *Rafael Valadares*, an angel that appeared in the most difficult hour of this journey, guiding me with such efficiency, security and positivity. It is even difficult to find words to express such gratitude;
- to my friends and colleagues at *LIE* and *GEMOS* for the support, the welcome, the learning and the pleasant hours in their company;
- to *Hugo* and *Guilherme* from *GISC*, the complex system research group of the *Physics Department*, for their solicitude and patience in helping me design the first networks;

- to *CAPES*, for the scholarship. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001;
- to all my *colleagues* and the *staff of the Soil Department*, from whom I learned a lot.

I would especially like to thank professors *Fábio Rodrigues Pereira* and *Rafael da Silva Teixeira* for accepting my invitation to the defense panel and for their valuable contributions.

At last, but not least, I thank my *family members*, especially my *partner* and my *daughters*, for their unconditional support and valuable company during this time of social distance. To them I also ask forgiveness for the eventual lack of attention and for the mishaps imposed by academic life. Who to ask for help when your own mother is a graduate student?

The Occident is, indeed, at the early dawn of an age of a truly New Science.

[...] this volume will come to be regarded as one of the forerunning treatises of that coming Great Age of a New Science, when, at last, the war-drums shall be stilled and the battle-flags be furled in the Parliament of Man and the Federation of World.

Then only shall there be throughout all the continents one nation indivisible, one constitution and one law, one sovereignty, one family of humanity in at-one-ment. Then only shall there be Right Civilization, Right Government, Right Science.

W. Y. Evans-Wentz, 1957.

ABSTRACT

DINIZ, Júnia Alencar, M.Sc., Universidade Federal de Viçosa, October, 2021. **Network Theory applied to discriminate land uses and evaluate mined areas reclamation.** Adviser: Emanuelle Mercês Barros Soares. Co-advisers: Igor Rodrigues de Assis, Ivo Ribeiro da Silva, Marcelo Lobato Martins and Rafael Vasconcelos Valadares.

Mining is an activity that deeply impacts the soil-plant system, by totally removing the vegetation cover and the superficial layers of the soil, being mandatory the recovery of these areas, back to their original condition. For this to be possible, it is necessary to know in depth the current state of the system and compare it with an appropriate reference. However, soil-plant system is highly complex, both in natural ecosystems and in agro-ecosystems degraded by mining, so we need a tool that can handle all this complexity. Network Theory has stood out in the field of complex systems as a very versatile tool and is already present in soil science, although it is still little used in the mined areas reclamation. This theory allows visualizing a complex system as a whole, transforming structural densities into clusters (communities) and enabling the identification of the elements (nodes) most relevant to the structure of relationships (links). For this reason, our aim was to investigate the potential of Network Theory to discriminate the current state of mined areas under reclamation and to quantify the similarity between these areas and their references, in two different conditions: iron mining and bauxite mining. For this, we used data on soil attributes and vegetation parameters as a basis for the construction of weighted bipartite networks, composed of two classes of nodes: *area* and *attribute*. All networks were generated with *Gephi* software, version 0.9.2, with layouts produced by the *ForceAtlas2* energy model, which provided a very clear and intuitive interpretation of the data structure. Network Theory allowed the discrimination of areas through groupings and the identification of the most relevant attributes for their distinction. It was also possible to quantify these differences by a *Similarity Index* (*SI*) and a *Relative Similarity Index* (*RSI*), using the weights of the nodes in the complete networks and in the area-projections, respectively. For the calibration of the method, we compared our results with those obtained through Principal Component Analysis (PCA) by other authors, who worked with the same data. We conclude that Network Theory can be used to discriminate the different land uses in areas affected by mining, allowing for consistent results, comparable to

those of PCA. The method developed in this work to calculate the weight of the edges in the networks proved to be very accurate, because the weights kept the proportions contained in the original data values. However, better results can be obtained by adjusting the equations to the characteristic functions of each attribute, especially to non-linear functions. We also believe that it is possible to select quality indicators using Network Theory, but this issue requires further conceptual studies. The present work can be considered an initial approach for the development of a new method of data exploration that can help in understanding the processes and in monitoring the reclamation of mined areas.

Keywords: Bauxite mining. Iron mining. Soil quality. Soil-plant system. Weighted bipartite networks. PCA.

RESUMO

DINIZ, Júnia Alencar, M.Sc., Universidade Federal de Viçosa, outubro de 2021. **Teoria de Redes aplicada à discriminação de usos do solo e avaliação da recuperação de áreas mineradas.** Orientadora: Emanuelle Mercês Barros Soares. Coorientadores: Igor Rodrigues de Assis, Ivo Ribeiro da Silva, Marcelo Lobato Martins e Rafael Vasconcelos Valadares.

A mineração é uma atividade que impacta profundamente o sistema solo-planta, por remover totalmente a cobertura vegetal e as camadas superficiais do solo, sendo obrigatória a recuperação destas áreas, de volta a sua condição original. Para que isto seja possível, é necessário conhecer em profundidade o estado atual do sistema e compará-lo com uma referência adequada. O sistema solo-planta é, entretanto, altamente complexo, tanto nos ecossistemas naturais como nos agroecossistemas degradados pela mineração, sendo preciso, portanto, uma ferramenta capaz de lidar com toda esta complexidade. A Teoria de Redes tem se destacado na área de sistemas complexos como uma ferramenta muito versátil e já está presente na ciência do solo, embora muito pouco utilizada na recuperação de áreas mineradas. Esta teoria permite visualizar um sistema complexo como um todo, transformando densidades estruturais em agrupamentos (comunidades) e possibilitando a identificação dos elementos (nós) mais relevantes para a estrutura de relações (links). Por esta razão, nosso objetivo foi investigar o potencial da Teoria de Redes para discriminar o estado atual de áreas mineradas em recuperação e quantificar a similaridade entre estas áreas e suas referências, em duas condições diferentes: mineração de ferro e de bauxita. Para isto, utilizamos dados de atributos do solo e vegetação como base para a construção de redes bipartidas ponderadas, compostas por duas classes de nós: *área* e *atributo*. Todas as redes foram geradas com o software *Gephi*, versão 0.9.2, com layouts produzidos pelo modelo de energia *ForceAtlas2*, que proporcionou uma interpretação bastante clara e intuitiva da estrutura dos dados. A Teoria de Redes permitiu a discriminação das áreas através de agrupamentos e a identificação dos atributos mais relevantes para a sua distinção. Também foi possível quantificar estas diferenças através de um *Índice de Similaridade* (*SI*) e um *Índice de Similaridade Relativa* (*RSI*), utilizando os pesos dos nós nas redes completas e nas *projeções-área*, respectivamente. Para a calibração do método, comparamos nossos resultados com os obtidos através da Análise de Componentes Principais (ACP) por outros autores, que trabalharam com os mesmos dados.

Concluímos que a Teoria de Redes pode ser utilizada para discriminar os diferentes usos do solo em áreas mineradas, permitindo a obtenção de resultados consistentes, comparáveis aos da ACP. O método desenvolvido neste trabalho para calcular o peso das arestas nas redes mostrou-se muito preciso, pois os pesos mantiveram as proporções contidas nos valores dos dados originais. No entanto, melhores resultados poderão ser obtidos ajustando-se as equações às funções características de cada atributo, principalmente às funções não lineares. Acreditamos também ser possível a seleção de indicadores de qualidade por meio da Teoria de Redes, mas esta questão demanda estudos conceituais mais aprofundados. O presente trabalho pode ser considerado uma abordagem inicial para o desenvolvimento de um novo método de exploração de dados que possa auxiliar no entendimento dos processos e no monitoramento da recuperação de áreas mineradas.

Palavras-chave: Mineração de bauxita. Mineração de ferro. Qualidade do solo. Sistema solo-planta. Redes bipartidas ponderadas. ACP.

SUMMARY

1 INTRODUCTION

Although mined lands do not comprise extensive areas, mining can be considered one of the most impacting activities to the soil, with subsequent need of recovering the area to the same previous use (BORGES, 2013). Thus, the essential problem of rehabilitate mined areas is to find the best strategy to bring the ecosystem back as close as possible to its original condition (SER, 2002). For this, it is necessary to know in depth the current state of the system and compare it to a suitable reference.

Repeating this process regularly allows the proper monitoring of the recovery, ensuring its success (SER, 2002). Yet a good diagnostic tool is needed to determine the similarity between the area under recovery and its reference. Understanding the importance of each element in the system (e.g., soil and vegetation features) is also key to learning how to manage or monitor it.

The recovery of mined areas mainly involves the soil-plant system, just because the mining process almost always requires the total removal of vegetation cover and topsoil. However, soil-plant system is highly complex, both in natural, non anthropized areas, and in agro-ecosystems, such as those degraded by mining. So to achieve a good diagnosis, it is first necessary to obtain a large amount of information that is able to encompass all the complexity of relationships within the ecosystem of interest.

Complex systems exhibit peculiar characteristics such as non-linearity and high levels of connectivity between their parts, forming a *network* of interactions, which adapts to the environment and evolves over time (BEN ELI, 2019). So we need a tool that not only can handle a large amount of information, but is also suitable for working with nonlinear systems.

Network Theory emerged as a new tool for working with complex systems at the beginning of this century and has been prominent in major fields of knowledge such as physics, biology, economics, computer and social sciences, among others. Network theory allows for a mathematical representation of the interactions among elements in a system, through the graph theory and statistical physics. Therefore, it is possible to visualize the system as a whole, to measure network indices, and to grouping the elements (nodes) according to their linkages (BARABÁSI, 2016).

There are already several works in soil science that use this new tool, mainly in soil physics (CÁRDENAS *et al.*, 2010; MOONEY; KOROŠAK, 2009; SAMEC *et al.*, 2013) and soil microbiology (BARBERÁN *et al.*, 2012; MA *et al.*, 2016; PÉREZ-RECHE *et al.*, 2012; SIMARD, 2009), but it is still very little used in the recovery of mined areas.

The few papers found on mined areas reclamation (FANG *et al.*, 2020; ZHANG; WANG; LI, 2019) used the correlations between soil attributes as a basis for building the networks. Correlations are also indispensable for identifying patterns in Principal Component Analysis (PCA), a multivariate statistic widely used in soil science to identify clusters and select indicators for monitoring recovering areas. But in order to investigate the full potential of Network Theory in discriminating mined areas in reclamation, we set ourselves the challenge of building the networks with just the raw data or its averages, without previously using any other descriptive or inferential statistics. Most of these methods require the data to satisfy assumptions of normality, linearity, homoscedasticity (HAIR, 2009), that cannot always be satisfied, especially in complex systems like soil.

Our hypotheses were:

- Network Theory can be used in the exploratory data analysis on mined areas to: (i) group similar areas; (ii) discriminate land uses and quantify the degree of similarity between reference areas and rehabilitating ones; (iii) identify the soil attributes or vegetation parameters that are most important in distinguishing these areas.
- It is possible to obtain the above results using weighted bipartite networks, just transforming the original data matrix into an adjacency matrix, with weights ranging from 0 to 1.

Thus, our aim was to investigate the potential of Network Theory to discriminate the current state of areas impacted by mining, from data of soil attributes and vegetation parameters, and also to compare these areas with their respective references. For this, a *Similarity Index* (*SI*) was created, as a ratio between the node weights in the complete network, and a *Relative Similarity Index* (*RSI*), as a ratio between the node weights of the network projections.

To calibrate the method, our results were compared with those obtained through PCA by other authors, who worked with the same data. This study can be considered as a first approach to the development of a new method that can help in data mining and also in understanding and monitoring the reclamation process of mined areas.

2 THEORETICAL BACKGROUND

2.1 Complex Systems Science

"If physics is the science of matter and its interactions, the science of complex systems is its natural extension where both matter and interactions are seen in a much broader context." (HOLOVATCH; KENNA; THURNER, 2017, p. 17). Complex systems science (CSS) has its roots in non-linear physics and information theory (FILOTAS *et al.*, 2014) although some of its principles have been independently discover in particular domains (SIEGENFELD; BAR-YAM, 2019) such as biology, ecology, economics and sociology, give rising to a new interdisciplinary research.

The study of so different disciplines under a unifying scientific framework is possible because CSS focuses on the type of *relations* within systems constituents and not on the type of constituents in themselves (SIEGENFELD; BAR-YAM, 2019). This assessment allows for the study of common properties complex systems exhibit such as heterogeneity, hierarchy, self-organization, adaptation, non-linearity, and uncertainty, no matter the nature of their constituents (FILOTAS *et al.*, 2014).

CSS is applicable to systems of very different scales, sizes and functions, and for this reason, it is very difficult to formulate a strict definition of what a complex system is. In general, the term can be applied to "any system consisting of many interconnected parts which, as a whole, possesses properties that are not trivial aggregates of the properties of its separate constituents" (HOLOVATCH; KENNA; THURNER, 2017, p. 02).

Another definition of complex systems – a more useful one to physicists – is that which "a system is complex if its behavior crucially depends on its details" (HOLOVATCH; KENNA; THURNER, 2017, p. 02). To comprehend that, one must firstly to consider what is meant by *complexity.* There are several different ways to measure complexity, but they may be separated in three basic groups, based on the system's (i) difficulty of description; (ii) difficulty of creation or; (iii) degree of organization (LLOYD, 2001). For example, the complexity of a system's behavior can be measure by the length (the amount of information) of its description. Yet, this length depends on the level of detail – the scale – used to describe it. Therefore,

complexity depends on scale and so “characterizing a system requires understanding its complexity across multiple scales” (SIEGENFELD; BAR-YAM, 2019, p. 03).

Complexity tends to increase as the scale decreases, since one gets more information when looking at a system in more detail, and at each level, the length of the description grows longer. In contrast to random systems, correlated systems present this multi-scale structure, since various behaviors occur at different scales, and thus gradually increase their complexity as the scale decreases, which is an essential feature of complex systems (SIEGENFELD; BAR-YAM, 2019). We note that soil system fits in very well with this feature.

Nonetheless, measuring complexity can present great difficulties when applied to empirical studies, particularly for researches in natural systems, entailing physicists to concentrate their efforts on looking for the properties (manifestations of the complex structure) that characterize each system, by means of tools borrowed from mathematics, statistical physics or information theory. And this is possible because “even apparently disparate, complex systems can have astonishingly similar characteristics both in their structure and in their behavior” and therefore, one can expect the existence of some universal laws that govern their properties. (KWAPIEŃ; DROZDZ, 2012, p. 115).

Definitions of complex systems include two other notions as well: (i) the macrostate and microstates of the system dynamically update each other, that is to say, the interactions among the constituents parts (the microstates) influence the system’s collective behavior (the macrostate), and conversely, these interactions are modified by the macrostate, in the course of system’s evolution; (ii) “the states of constituents and interactions co-evolve over time", or in other words, “interactions change the states of constituents, and the states of the constituents change the *networks* of interactions between constituents” (HOLOVATCH; KENNA; THURNER, 2017, p. 03 our emphasis).

Networks play a central role in the science of complex systems, because they offer a way to describe the different types of interactions between its constituents (HOLOVATCH; KENNA; THURNER, 2017). According to Barabási (2016, section 1.2) “we will never understand complex systems unless we develop a deep understanding of the networks behind them”.

2.1.1 Soil as a complex system

A system is simply a set of elements and the array of connections among them, called relations. In an unordered system (like a pile of sediments) there is no specific order or structure within its elements, hence it can be described only by listing all its elements, and its respective properties, in isolation. However, if these parts are ordered in a specific way, a global pattern of organization emerges and the system works as a coherent whole (SYSTEMS INNOVATION, 2020).

Generally speaking, a complex system exhibits some basic characteristics such as: (i) **heterogeneity** : presence of many components, of different nature and behavior, on many different scales, organized in subsystems that form a hierarchical structure (ii) **nonlinearity**: the result of the combined effect amongst the high interdependent elements of the system can be greater or lesser than the simple sum of each part in isolation; (iii) **connectivity**: at high levels of connectivity the system stops being a set of parts to become a *network* of connections; (iv) **adaptation**: without centralized coordination and with a certain degree of autonomy elements can synchronize their states locally and the system develops on the macroscale through a process of evolution (FILOTAS et al., 2014; SYSTEMS INNOVATION, 2020);

As can be seen, soil presents all the features that allow us to describe it as a complex system: it is an ordered system (unlike sediments), composed of heterogeneous parts (mineral particles, soil organisms, water, plant roots and debris, nutrients etc) self-organized in multiple scales (from ions and molecules to aggregates, profiles and pedogenetic sequences), that build up a specific structure, that eventually fosters the emergence of numerous important functions (e.g., water purification, nutrient cycling and carbon sequestration). Soil is also a living system containing multiple life forms in interaction with its mineral particles. So that the system as a whole reacts to the influences and modifications of the environment, adapting and developing over time (SIMARD, 2009; STEENBOCK; VEZZANI, 2013).

Therefore, in order to understand how soil works, it is not enough to simply know the properties of its elements. It is necessary to consider the interactions between them, i.e. the relations that turn soil into a network. For this reason, we consider that Network Theory can be an adequate tool to evaluate process involving the soil system. The next topic briefly

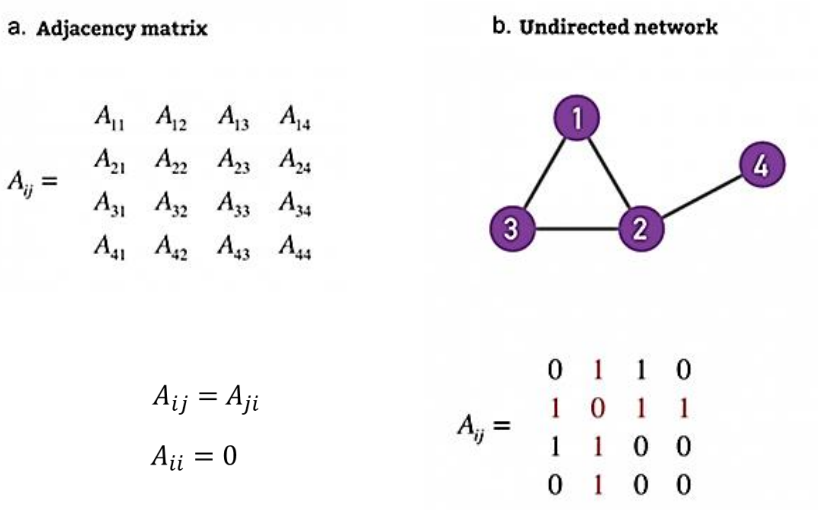
describes the concepts of Network Theory and the type of network we have used in this work. We hope this tool can help land managers and restorers to monitor recovering areas more efficiently, better understanding the relationships involved in this process.

2.2 Network Theory

Network Science has emerged as a separate discipline in the early 21st century. Their key characteristics are interdisciplinarity and the use of empirical data, employing quantitative, mathematical and computational tools. The mathematical formalism is derived from graph theory and statistical physics, besides concepts borrowed from engineering and statistics. The field has a strong computational character, and a series of software tools are already available, enabling scientists and other practitioners to analyze the networks of their interest (BARABÁSI, 2016).

Networks can be represented by a *graph*, that is, a map of their wiring diagram, where the elements of the network are called *nodes* (or *vertices*), and the interactions between them called *links* (or *edges*). Networks can be constructed either by real or virtual elements, and they can be mathematically represented by an adjacency matrix (Figure 1).

Figure 1 – The adjacency matrix for undirected networks (unweighted).



Source: adapted from Barabási (2016).

Note: The degree (number of links) of a node in an undirected network (b) can be expressed by the sum over columns or rows on the adjacency matrix, as highlighted in red for node 2, which has three links.

The adjacency matrix of undirected networks is symmetric (Figure 1). For unweighted networks like these, if there is a link between two nodes, $A_{ij} = 1$; in contrast, if there is no links between them or no self-loops (i.e., nodes do not interact with themselves), $A_{ij} = 0$. For weighted networks (see item 2.2.2. below), $A_{ij} = w_{ij}$, and so there is a link with weight w_{ij} between the nodes, indicating the strength of the interactions among them.

2.2.1 Network properties

For great networks, a visual inspection of their graphs is not sufficient to understand their properties. Hence, we have to use the tools of the network science to explore them. There are many different properties in a network that can be measured, as following, according to Barabási (2016), for undirected networks:

▪ **Degree,** k_i: the number of links a node has to other nodes, where *i* is the i^{th} node in the network. For example, in Figure 1b, $k_1 = 2$; $k_2 = 3$; $k_3 = 2$; $k_4 = 1$.

▪ **Total number of links,** *L*:

$$L = \frac{1}{2}\sum_{i=1}^{N} k_i \qquad (1a)$$

where *N* is the number of nodes in the network (e.g., in Figure 1b, *N* = 4 and *L* = 4).

In a **complete graph,** also called a *clique*, each node is connected to every other node. Therefore, its total number of links is equal to L_{max}:

$$L_{max} = \frac{N(N-1)}{2} \qquad (2a)$$

If *L*‹‹ L_{max}, the network is considered sparse, like most real networks. For example, if the network of Figure 1b, was a clique, its L_{max} would be equal to 6.

▪ **Average degree**, $\langle k \rangle$:

$$\langle k \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i = \frac{2L}{N} \qquad (3a)$$

For example, in Figure 1b, $\langle k \rangle$ = 2.

▪ **Degree distribution**, p_k**:** the probability that a randomly selected node has degree k:

$$p_k = \frac{N_k}{N} \qquad (4a)$$

where N_k is the number of degree-k nodes, and $\sum_{k=1}^{\infty} p_k = 1$. For example, in Figure 1b, p_2 = 2/4 = 0.5.

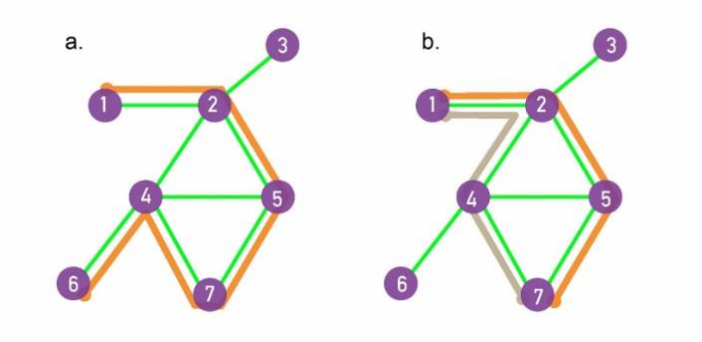
▪ **Path:** a route that runs along the links of the network. A path consists of *n+1* nodes and *n* links.

▪ **Path's length:** the number of links a path contains.

▪ **Shortest path** (or distance): the path with the fewest number of links, denoted by d_{ij} or simply *d* (Figure 2).

▪ **Average Path Length** $\langle d \rangle$ – the average of the shortest paths between all pairs of nodes.

Figure 2 – Examples of paths.



Source: Barabási (2016).

Note: (a) a path with length equal 5 (in orange); (b) two shortest paths between nodes 1 and 7, or the distance d_{17}, with length equal 3 (in orange and gray).

▪ **Diameter**, d_{max}: the longest shortest path in a graph. The diameter is the largest distance between any pair of nodes.

▪ **Clustering coefficient**: the degree to which the neighbors of a given node link to each other. The number of neighbors of a k_i node is equal to k_i, , that is to say, the neighbors of node *i* are the nodes connected to it.

▪ L**ocal clustering coefficient**, C_i:

$$C_i = \frac{2L_i}{k_i(k_i - 1)} \qquad (5a)$$

where L_i represents the number of links between the k_i neighbors of node *i.*

If none of the neighbors of node *i* link to each other, $C_i = 0$; if the neighbors of node *i* form a complete graph, $C_i = 1$. Since, C_i is the probability that two neighbors of a node link

to each other (Figure 3a). C_i measures the network's local link density: the higher the local clustering coefficient, the more densely interconnected the neighborhood of node *i*.

▪ **Average clustering coefficient,** $\langle C \rangle$: the degree of clustering of a whole network, defined as:

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i \qquad (6a)$$

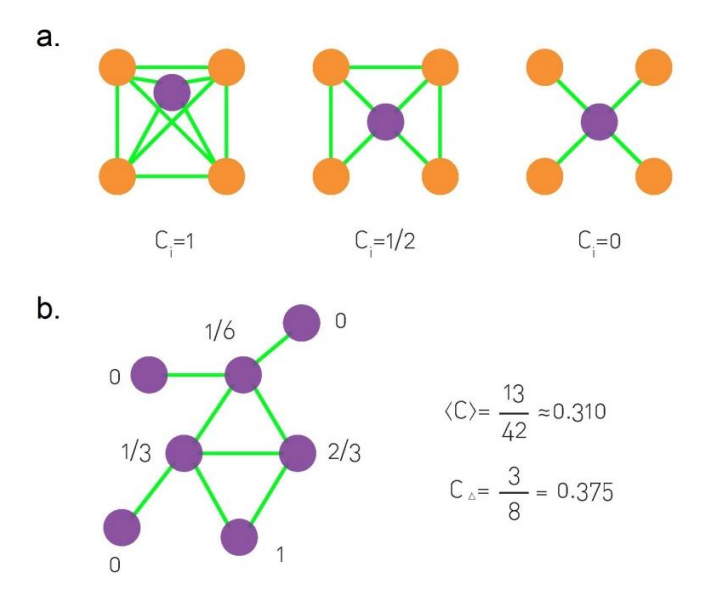
The average clustering coefficient is the probability that two neighbors of a randomly selected node link to each other (Figure 3b).

▪ G**lobal clustering coefficient**, C_Δ: measures the ratio between the number of closed triangles and open triangles (triplets) in a network, and is defined as:

$$C_\Delta = \frac{3\ x\ Number\ of\ triangles}{Number\ of\ connected\ triplets} \qquad (7a)$$

where a connected triplet is an ordered set of three nodes (Figure 3b).

Figure 3 – Examples of local, average and global clustering coefficients



Source: Barabási (2016).

Note: (a) examples of local clustering coefficient, C_i; (b) an example of network, with the local clustering coefficient of each node, calculated according to equation (5a); the average clustering coefficient $\langle C \rangle$ according to equation (6a); and the global clustering coefficient, C_Δ, according to equation (7a).

▪ **Connectedness**: a network is connected if all pairs of nodes are connected, and it is disconnected if there is at least one pair with $d_{ij} = \infty$. In a disconnected network, the subnetworks are called *components* or *clusters*. A link that connects two components is called a *bridge*, and if cut, disconnects the network.

▪ **Community:** a group of nodes more connected to each other than to nodes from other communities. They are subgraphs in a network, locally dense connected. A community is *strong* if each node inside it has more links within the community than to the rest of the graph. Conversely, it is *weak*.

▪ **Centrality**: refer to node pairs that belong (or not) to different communities. Centrality is small if two nodes are in the same community and high if two nodes belong to different communities.

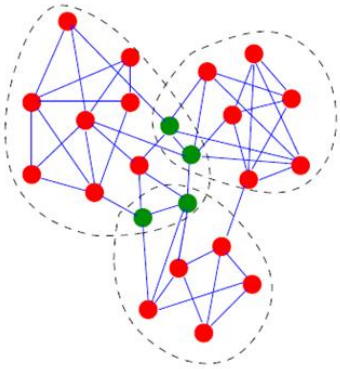
▪ **Betweenness centrality:** measures how often a node appears on shortest paths between nodes in the network**.**

▪ **Closeness centrality:** represents the average distance from a given starting node to all other nodes in the network.

▪ **Eccentricity**: the distance from a given starting node to the farthest node from it in the network.

▪ **Modularity**: measures how good is a given network partition, considering that a node rarely is confined to a single community, that is to say, communities overlap (Figure 4), and that exist different ways to divide a network. Beside its limitations, modularity offers a first understand of a network's community structure.

Figure 4 – Overlapping communities

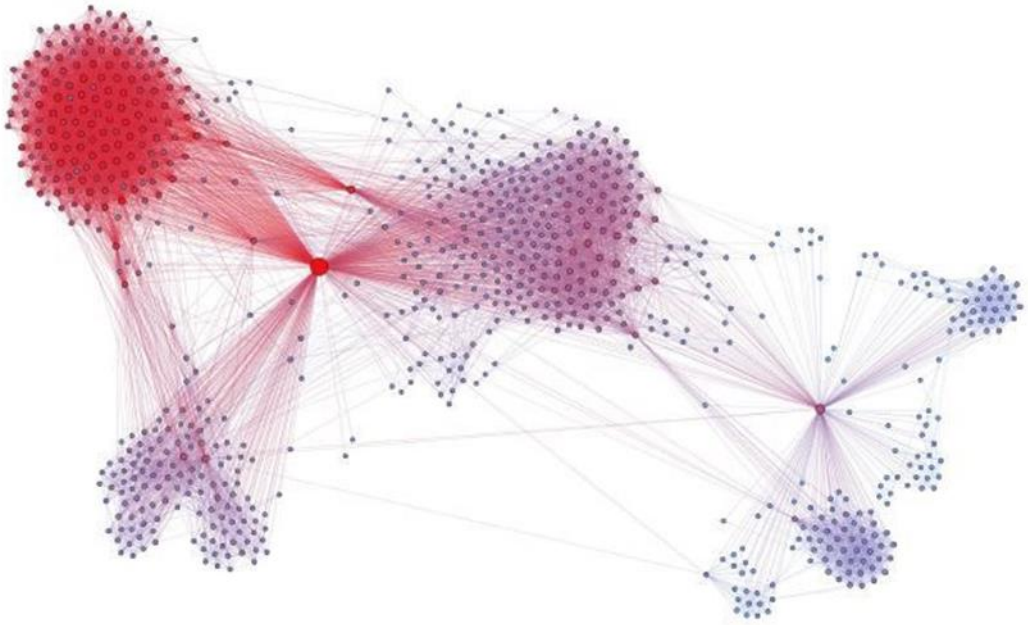


Source: Fortunato and Castellano (2007).

Note: the green vertices are shared between three groups.

The properties described above allow us to study networks systematically and figure out the features of a particular complex system. For instance, the most connected nodes are called *hubs* (Figure 5 shows a network with three hubs), assigning to a network a spoke character.

Figure 5 – Example of a graph containing three hubs, in a complex network.



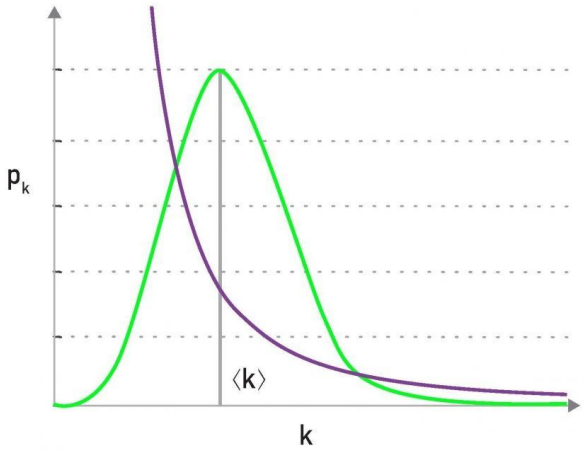
Source: Cota (2016).

Note: in this graph, nodes (points) represent friends and common friends of two profiles in the social network *Facebook*, and links (lines) represent the friendly relationships.

This feature influences the degree distribution, which plays a central role in network theory, since calculation of most network properties requires to know it. Besides, degree distribution determines many network phenomena, such as the network robustness – the resilience or attack tolerance – and the spreading phenomena – the transference of information (or, for instance, viruses) along the links (BARABÁSI, 2016).

Real-world networks have typically a fat tailed (a power law decay) of node degree distribution, called *sale-free*, that reflects strong inhomogeneities in the system's behavior (Figure 6). This type of distribution is characterized by a very short average distance between nodes, called *small-world* phenomena. Networks with scale-free structures (as the network shown in Figure 5), sometimes called *complex networks* may have "a strong impact on the dynamics of the system and may effect percolation properties or self-organization" (HOLOVATCH; KENNA; THURNER, 2017, p. 11).

Figure 6 – The degree distribution of random and the scale-free networks



Source: Barabási (2016).

Note: in green: the degree distribution of random network – with a Poisson or a Gaussian distribution; in purple: the degree distribution of a scale-free network. As noted, the degree of a randomly chosen node in a scale-free network can be significantly different from the average degree $\langle k \rangle$. Hence, $\langle k \rangle$ does not serve as an intrinsic scale.

Other important network indices are *clustering coefficient* and *assortativity*. In the first, a power law decrease in its values is a signal of a hierarchical organization in the network. The second, that can be measured for instance by using Pearson's correlation coefficient, is an extent to which similar nodes are linked with each other. Both of this properties have great impact on how interactions between nodes manifest and how a complex system evolves over time (HOLOVATCH; KENNA; THURNER, 2017).

The identification of communities can also contribute to understand the structure of networks and may help to uncover functional modules. Several types of community detection algorithms have been proposed, facing the challenge of finding good partitions in a reasonably fast way, especially for great networks (BLONDEL *et al.*, 2008). These algorithms are already available as software packages, although their efficient use and interpretation require knowing the assumptions built into them (BARABÁSI, 2016).

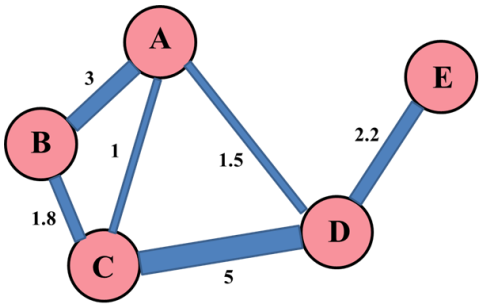
In summary, according to Barabási (2016), the presence of communities has a profound impact on network behavior. Besides, this author claims that the structure of the network is a reflection of a deep order, present in most complex systems, and this order results of deviations from random settings, requiring a deeper explanation.

2.2.2 Weighted networks

Incorporating the weights of the edges in the study of networks' properties can capture the richness of the data, depicted by its quantitative interactions. Weights allow to differentiate edges in term of their strength, intensity or capacity. For instance, they may represent the carbon flow between species in food webs, the number of synapses in neural networks, or the amount of traffic in transportation networks (OPSAHL; PANZARASA, 2009). Other examples include the actors–movies network; the scientists–papers network; the world cities hosting branches of multinational firms; and, moreover, the intensive use of bipartite networks in internet technology and applications, during recent years (PAVLOPOULOS *et al.*, 2018).

For unweighted (or binary) networks, the adjacency matrix indicates only the presence (Aij = 1) or the absence (Aij = 0) of a link. In weighted networks (Figure 7), links have a weight w_{ij} , representing a strength or flow parameter. This way, in the adjacency matrix, $A_{ij} = w_{ij}$ (BARABÁSI, 2016). In Figure 7 weights are nonnegative, but negative weights are also possible. For instance, they are used sometimes in sociological networks to represent animosity between individuals (NEWMAN, 2004).

Figure 7 – Example of weighted network.



Adjacency matrix

$$\begin{bmatrix} 0 & 3 & 1 & 1.5 & 0 \\ 3 & 0 & 1.8 & 0 & 0 \\ 1 & 1.8 & 0 & 5 & 0 \\ 1.5 & 0 & 5 & 0 & 2.2 \\ 0 & 0 & 0 & 2.2 & 0 \end{bmatrix}$$

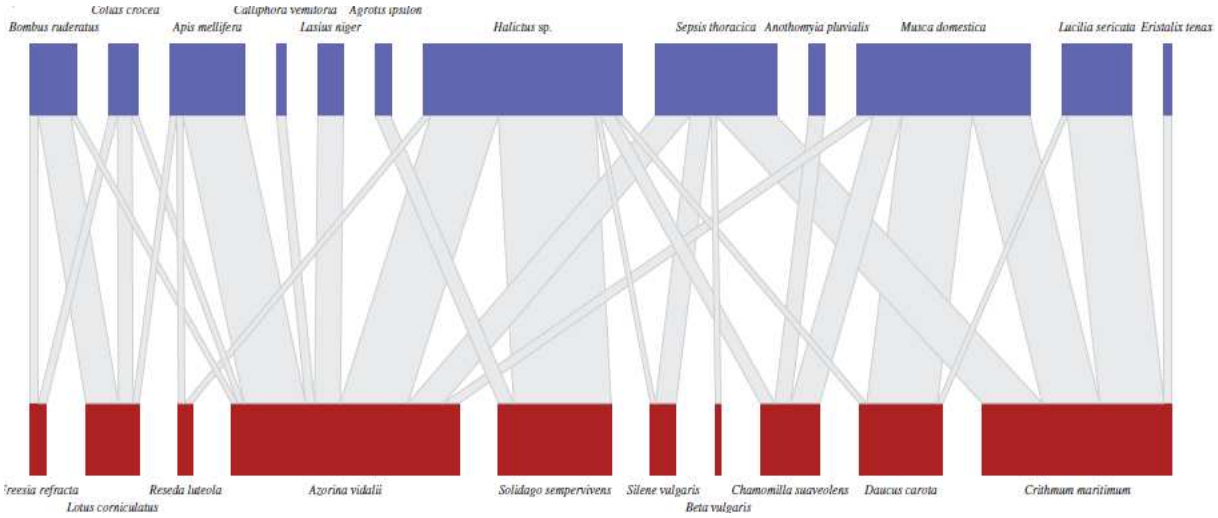
Source: the author.

Note: In this network, the adjacency matrix has only nonegative weights.

In this work, we have used weighted networks, in which weights were calculated based on the values of soil attributes analysis and vegetation parameters measures. Specifically, we

have used bipartite weighted networks, that can be represented, for example, as shown in Figure 8. More details about bipartite networks are found in the next topic.

Figure 8 – A bipartite network of 12 species of pollinators (blue nodes) visiting 10 plant species (red nodes).



Source: Beckett (2016).

Note: the width of the edges linking the nodes represents the number of pollinator–plant visitations, while the width of the nodes represents the total of visits made by a pollinator species or received by a plant species.

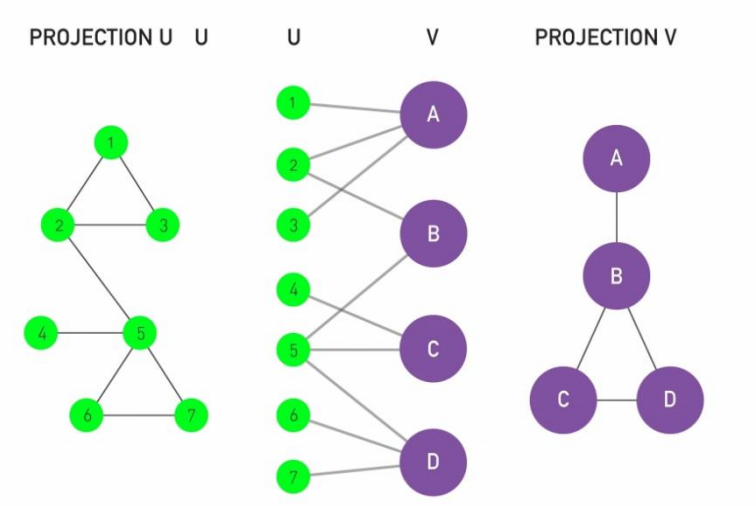
2.2.3 Bipartite networks

In bipartite (or two-mode) networks there are two distinct classes of nodes, such that nodes can only interact with nodes from the other class (BECKETT, 2016). The interesting thing is that this type of networks can be visualized as a whole (as shown in Figure 8 and in the network at the center of Figure 9), or visualized through its projection, in which only one of the classes is visible (Figure 9, left and right networks).

Bipartite networks have some particular topological features, described by Pavlopoulos et al (2018) as following:

- **Degree**: in a bipartite graph, the maximum degree of a node is equal to the number of nodes from the other class. Besides, the sum of the degrees of the first class is equal to the sum of the degrees of the opposing one.

Figure 9 – A bipartite network with two classes of nodes, U and V.



Source: Barabási (2016)

Note: as shown in the center, there are no direct U-U or V-V links. Networks on the left and right represents the two projections generated from the center network. The projections are obtained by connecting two nodes from the same class if they are linked to the same node in the other class.

- **Closeness centrality:** In a bipartite graph, a node has a minimum distance "1" from vertices of the opposing class and "2" from vertices of the same class. Moreover, all paths between nodes of the same class are of even length, "a property that rather complicates the calculation of several measures**"** (PAVLOPOULOS *et al.*, 2018, p. 4)**.**
- **Local and global clustering coefficient**: these two conventional measures for 1-mode networks cannot be applied to bipartite networks, since size 3 cycles are absent in the latter.
- **Bipartivity**: this measure indicates how close a given network is to being a bipartite one.
- **Internal links and pairs**: appear when unipartite networks are derived (as projections) from a given bipartite network. Although it is a usual approach for bipartite network analysis, the resulting graph is associated with a large loss of information.

In addition to the properties mentioned above, it is worth mentioning **modularity**, which for bipartite networks, can be used to identify communities (groups or modules) within the same type of nodes or for both types. There are also several algorithms that use the strengths between interactions, to maximize weighted modularity in bipartite networks (BECKETT, 2016).

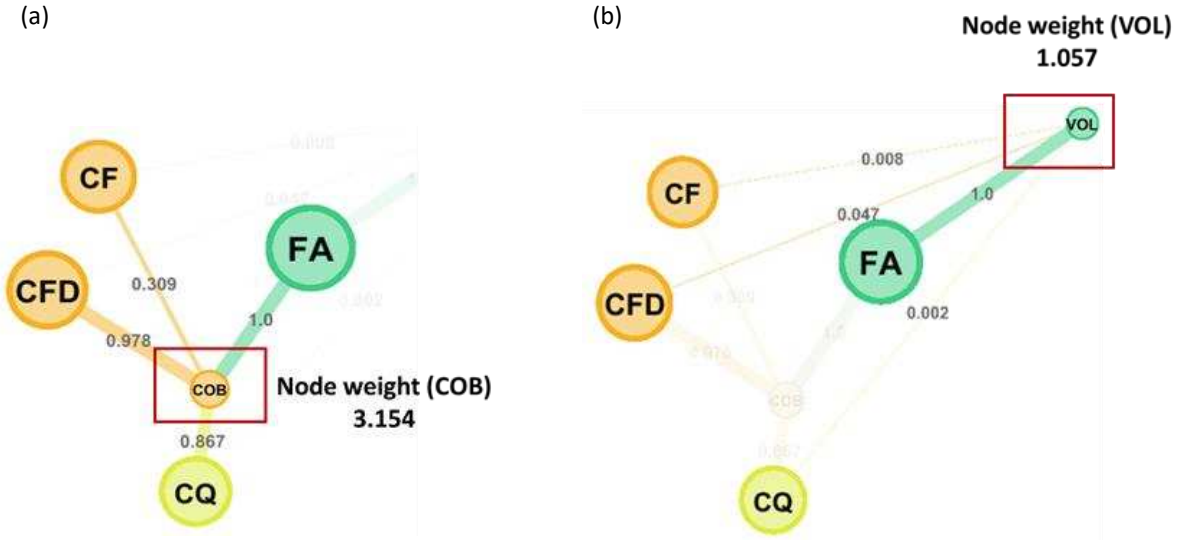
In the present study, the bipartite networks have two classes of nodes: (i) **area**, comprising reference areas, mined areas and rehabilitated ones; (ii) **attribute**, comprising soil attributes and vegetation parameters. This way, there are only links between classes (i) and (ii), although their projections connect the nodes from the same class, showing the interactions between their elements, for instance, a reference area and a rehabilitating one.

2.2.4 Network layout

To visualizing the networks, we have used *Gephi* software (BASTIAN; HEYMANN; JACOMY, 2009) – which transform networks (usually from 10 to 10,000 nodes) into maps – and *ForceAtlas2*, its default layout algorithm. *ForceAtlas2* works, fundamentally, making nodes to repulse each other (like charged particles) and edges to attract their nodes (like springs), creating a movement that results in a balance, placing each node depending on its connections with other nodes (JACOMY *et al.*, 2014).

The weight of the edges is also taken into account, and changes the force of attraction between the nodes, which may strongly impact the final result. Figure 10 shows an example of this, for two attributes (*COB* and *VOL*) and four areas (*FA, CF, CFD* and *CQ*).

Figure 10 – Example of a force-directed layout in a weighted network.



Source: the author.

Note: (a) and (b) show two highlighted nodes (*COB* and *VOL*, respectively) from the same network. The numbers on the edges represent their weights; the thickness of the edges is proportional to their weight.

As can be seen in Figure 10, the edges with the smaller weights have less attraction on the nodes *COB* and *VOL*, making them take the opposite direction (Figure 10a), or giving the impression that they are being pushed out of the graph (Figure 10b). Since the weight of each node results from the sum of the weights of its edges, the heavier nodes (e.g., COB) tend to be in the center of the network, while lighter nodes (e.g., VOL) tend to be in the periphery.

Therefore, the nodes with the smallest weights are those that show the greatest differences between the areas. For this reason, in the present work the attributes that presented the lowest weights in each network were considered as the most relevant for the distinction of those areas.

It is worth noting that the process of graph construction is not deterministic, and cannot be interpreted as a Cartesian projection, since the position of a given node cannot be taken by itself, but depends on the position of the other nodes (JACOMY *et al.*, 2014). For this reason, it is possible to generate different layouts from the same matrix, where the nodes occupy different places. However, the interactions between the nodes, as well as their weights (and also the edges weights) remain the same, following the same distribution, not affecting the quantitative results.

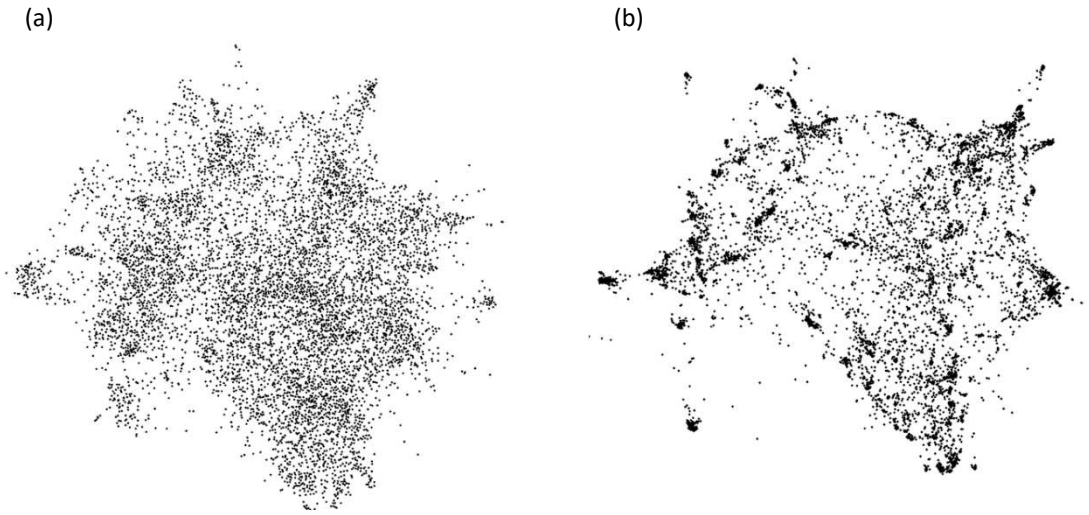
In essence, *ForceAtlas2* algorithm turns structural proximities into visual proximities, identifying groups with denser relations as communities. This collective proximity is measured by modularity, a measure proposed by Newman (2006), and optimized by force-directed layouts, since they produce visual densities that denote structural densities. The final configuration is expected to help the analysis of the data, allowing for a visual and intuitive interpretation of the its structure (JACOMY *et al.*, 2014).

The *LinLog mode* of *ForceAtlas2,* an energy model that uses a logarithmic attraction force, highlights Newman's modularity, and therefore results on a better placement of the communities (Figure 11). "This energy model has a strong impact on the shape of the graph, making the clusters tighter" (JACOMY *et al.*, 2014, p. 3).

In this work we have used these tools (mainly the *LinLog* mode of *ForceAtlas2* and the Modularity algorithm) to discriminate land uses and identifying the similarities (proximities) between rehabilitated mined areas and their references. We also used the node weights of area-projections to quantify the differences between the areas and the weighted degree

distribution of attribute-projections to determine the most relevant attributes for distinguishing the areas.

Figure 11 – Layouts with two different types of forces: *ForceAtlas2* (a) and the *LinLog mode* of *ForceAtlas2* (b).



Source: adapted from Jacomy et al. (2014).

3 MATERIALS AND METHODS

3.1 Bauxite mined area dataset (Dat-1)

For bauxite mining, the networks were built using data from a study carried out on a bauxite mined area in São Sebastião da Vargem Alegre district, Zona da Mata region, Minas Gerais State, Brazil. The predominant soil in this region is the Oxisol (LATOSSOLO VERMELHO-AMARELO distrófico típico, according to Sistema Brasileiro de Classificação do Solo – SiBCS), with clayey texture in the surface (0–0.20 m) soil layer. The climate classification is Cwa (according to Köppen), with a mean annual rainfall of 1,289 mm and mean annual temperature of 22.6°C (BORGES, 2013; BORGES et al. 2019).

Table 1 shows the 20 physical, chemical, organic and microbiological soil attributes evaluated in three different scenarios: (i) pre-mining: encompassing a native forest, Atlantic Forest (*Forest*) and a 10-year-old coffee crop (*PreMin*); (ii) post-reconfiguration: a reconstructed soil, six months after mining and reconfiguration (*PosMin*); and (iii) post-recovering: the recovered soil by coffee and cover crops, 19 months after recovering start.

In the first scenario, *Forest* was taken as a reference of the soil state in its original condition (non-mined) and *PreMin* as a reference of the soil before mining. In the second scenario, the coffee plants were slashed and the soil above the ore was removed and stockpiled for approximately 1 year. After mining, the area was reconfigured by unpacking and spreading the stockpiled topsoil throughout the mined site. This scenario was used to assess the impact of mining and as an initial reference for the recovery process. The third scenario, a recovery experiment installed eight months after mining and soil reconfiguration, comprised a *Coffeea arabica* crop using four fertilizations – no fertilization (*N*-); poultry litter fertilization (*P*-); chemical fertilization (*C*-); poultry litter and chemical fertilizations combined (*PC*-), and four cover intercrops: grass, *Brachiaria brizantha* cv. Piatã (*G*); leguminous, Estilosantes Campo Grande (*L*); grass + leguminous mix (*GL*); no intercrop (*NI*); encompassing, in total, 16 treatments: N-G, N-L, N-GL, N-NI, P-G, P-L, P-GL, P-NI, C-G, C-L, C-GL, C-NI, PC-G, PC-L, PC-GL, PC-NI.

Table 1 – Mean values of the soil attributes (0-0.20 m soil layer) at pre-mining sites with native forest (*Forest)* and coffee cultivation (*PreMin*), at six months after reconstruction the bauxite-mined site (*PosMin*), and at 19 months after reclamation with coffee plants under different fertilization and cover crops treatments.

		Reference areas			Treatments															
					No fertilizer				Poultry litter				Chemical				Poultry litter + Chemical			
Attribute	Unit	Forest	PreMin	PosMin	N-G	N-GL	N-L	N-NI	P-G	P-GL	P-L	P-NI	C-G	C-GL	C-L	C-NI	PC-G	PC-GL	PC-L	PC-NI
									Soil physiccal attributes											
BD	kg dm^{-3}	0.57	0.69	1.22	1.34	1.39	1.30	1.36	1.25	1.29	1.22	1.36	1.28	1.25	1.22	1.31	1.22	1.26	1.33	1.21
Mic	dm^3 dm^{-3}	0.25	0.28	0.24	0.34	0.37	0.33	0.33	0.34	0.35	0.35	0.35	0.33	0.34	0.37	0.32	0.38	0.29	0.36	0.34
Mac	dm^3 dm^{-3}	0.52	0.45	0.33	0.19	0.14	0.21	0.19	0.22	0.19	0.22	0.19	0.22	0.23	0.21	0.22	0.19	0.25	0.17	0.23
TP	dm^3 dm^{-3}	0.76	0.73	0.57	0.53	0.51	0.54	0.52	0.56	0.54	0.56	0.54	0.55	0.57	0.58	0.54	0.57	0.55	0.52	0.57
									Soil chemical attributes											
P	mg dm^{-3}	2.05	7.87	1.05	0.61	0.92	0.73	0.67	24.60	18.60	19.80	28.00	25.70	13.90	12.20	5.45	43.30	100.60	55.70	43.60
Mn	mg dm^{-3}	8.04	14.30	5.04	2.85	3.32	2.72	2.33	7.51	7.17	8.37	6.63	1.68	2.48	3.23	2.43	7.92	16.20	10.30	11.00
Fe	mg dm^{-3}	230.60	96.50	174.90	91.80	105.10	84.50	73.20	50.20	59.00	76.90	28.60	40.60	43.20	50.90	47.50	53.80	55.90	42.70	53.30
Zn	mg dm^{-3}	10.70	10.90	1.25	0.25	0.29	0.24	0.21	4.71	4.13	4.80	5.85	0.28	0.32	0.34	0.40	3.94	10.00	5.73	5.88
pH	-	4.62	5.64	5.09	5.09	5.25	5.05	4.98	5.76	5.81	5.41	5.72	6.02	5.80	6.19	5.98	6.25	6.71	6.31	6.03
Al3	$cmol_c$ dm^{-3}	1.56	0.10	0.14	0.12	0.08	0.05	0.05	0.01	0.03	0.04	0.01	0.00	0.01	0.00	0.00	0.03	0.00	0.00	0.01
HAl	$cmol_c$ dm^{-3}	14.70	7.22	3.91	5.29	5.59	4.74	4.87	4.38	4.38	5.86	3.25	3.53	4.36	4.16	3.54	3.51	2.50	3.29	4.25
EP	$cmol_c$ dm^{-3}	41.10	50.20	55.50	95.70	90.40	74.20	73.60	95.60	88.90	91.00	72.40	83.70	91.00	77.90	94.40	109.40	108.60	99.30	120.40
TEB	%	0.37	2.64	0.57	0.55	1.24	0.68	0.79	3.29	3.63	3.20	2.65	2.81	3.00	4.09	2.79	4.88	6.69	5.05	5.01
									Soil organic-biological attributes											
TOC	dag kg^{-1}	5.77	3.61	1.51	1.43	1.46	1.28	1.28	1.44	1.58	1.80	1.11	1.34	1.59	1.49	1.27	1.57	1.61	1.50	1.70
LOC	g kg^{-1}	3.10	2.84	0.59	0.77	0.99	0.84	0.83	1.24	1.21	1.34	0.90	0.96	1.14	1.01	0.87	1.40	1.56	1.27	1.41
TN	dag kg^{-1}	0.37	0.25	0.09	0.06	0.06	0.05	0.05	0.08	0.08	0.10	0.06	0.06	0.06	0.06	0.05	0.07	0.07	0.08	0.08
AciP	µg *p* -nitrophenol h^{-1} g	379.10	223.30	79.10	91.20	86.70	74.60	66.30	129.70	104.70	132.40	69.20	91.50	62.60	61.00	42.60	107.90	89.60	101.00	88.70
AlkP	µg *p* -nitrophenol h^{-1} g	55.90	38.20	0.00	11.60	19.60	17.20	5.24	43.40	31.30	32.10	31.50	25.60	20.30	19.80	14.30	37.30	40.10	22.70	24.60
Gluc	µg *p* -nitrophenol h^{-1} g	78.40	64.90	21.80	2.00	17.60	16.10	26.80	35.10	38.40	44.70	50.70	4.38	10.20	18.40	4.02	28.00	31.50	30.10	30.80
CCO2	µg of C-CO_2 h^{-1} g $soil^{-1}$	0.18	0.18	0.23	0.16	0.20	0.14	0.13	0.27	0.24	0.19	0.17	0.30	0.27	0.25	0.25	0.68	0.27	0.45	0.36

Source: adapted from Borges et al. (2019, Table S1)

Note: *TOC*; total organic carbon; *LOC*: labile organic carbon; *TN*: total nitrogen; *P*: available phosphorous; *Mn*: manganese content; *Fe*: iron content; *Zn*: zinc content; *pH*: pH in water; *Al3*: exchangeable aluminum, Al^{3+}; *HAl*: potential acidity, $H + Al^{3+}$; *EP*: soil equilibrium phosphorus; *TEB*: total exchangeable basis, K^{+}+ Ca^{2+}+ Mg^{2+}; *ASI*: aluminum saturation Index. *BD*: bulk density; *Mic*: microporosity; *Mac*: macroporosity; *TP*: total porosity; *AciP*: acid phosphomonoesterase enzyme activity; *AlkP*: alkaline phosphomonoesterase enzyme activity; *Gluc*: ß-glucosidase enzyme activity; *CCO2*: microbial respiration. *G*: grass intercrop; *L*: leguminous intercrop; *GL*: grass + leguminous mix intercrop; *NI*: no intercrop.

3.2 Iron mined area dataset (Dat-2)

For iron mining, the networks were built using data from a study carried out on iron mined areas in the Quadrilátero Ferrífero geologic province, Minas Gerais state, in Nova Lima district and surrounding areas, Brazil. The climate in this region is the humid subtropical (Cwa), according to Köppen classification with a mean annual rainfall of 1.390 mm and mean annual temperature of 21°C (ALVES, 2019).

Table 2 shows the 34 physical, chemical and biological soil attributes plus 2 vegetation parameters evaluated in four references and four recovering areas. The references encompass: a ferruginous rupestrian grassland with small shrubby vegetation on flat relief, named Sparse Ferruginous Rupestrian grassland (*FR*); a ferruginous rupestrian grassland with dense shrubby vegetation on steep slopes, named Dense Ferruginous Rupestrian grassland (*FR-D*); a quartzite rupestrian grassland with predominant grass instep relief (*QR*); and an Atlantic Forest with dense arboreal vegetation on steep slopes (*AF*) – all of them native vegetation, never disturbed by anthropic activities. The soils in these areas were classified as follows, according to WRB/FAO and [SiBCS – Brazilian System of Soil Classification]: Leptsol [NEOSSOLO LITÓLICO] in *FR-D* and *QR*; Plinthosol [PLINTOSSOLO PÉTRICO] in *FR*; and Cambisol [CAMBISSOLO HÁPLICO] in *AF*.

The rehabilitating areas encompass: a compensation area, 5 years under reclamation (*COM5*); a sterile pile, 15 years under reclamation (*SP15*); a pit, 15 years under reclamation (*Pit15*); and a sterile pile, 20 years under reclamation (*SP20*). *COM5* is an unmined area, adjacent to *AF*, only altered by deforestation and natural fires. *SP-15* is adjacent to *Pit-!5*, so that the waste from this pit was deposited in it.

There is no precise information about the recovery action taken in the areas mentioned above. It is only known that, in general, the waste material was laid out on top of them and seedlings were planted with standard fertilizer in a random manner, without defined spacing.

Table 2 formed the basis for the construction of all networks of iron mined areas, with the exception of *Network-3*, based on Table 3 (Annex A).

Table 2 – Averages of soil attributes and vegetation parameters assessed in reference environments and in areas undergoing reclamation in the Quadrilátero Ferrífero.

		Reference areas				Recovery areas			
Attribute	unit	FR	FR-D	QR	AF	COM5	SP15	Pit15	SP20
		Soil physical attributes							
Ston	$g\ kg^{-1}$	372.20	485.58	187.06	45.13	154.27	147.59	160.45	139.14
FS	$kg\ kg^{-1}$	0.06	0.12	0.17	0.07	0.05	0.23	0.23	0.08
FSS	$kg\ kg^{-1}$	0.36	0.42	0.60	0.45	0.64	0.75	0.80	0.67
Silt	$kg\ kg^{-1}$	0.31	0.30	0.43	0.38	0.59	0.52	0.57	0.59
Clay	$kg\ kg^{-1}$	0.16	0.19	0.22	0.47	0.30	0.06	0.04	0.16
BD	$g\ cm^{-3}$	1.43	1.36	1.00	0.73	0.91	1.69	1.91	0.88
PD	$g\ cm^{-3}$	3.18	3.35	2.43	2.05	2.68	3.55	3.56	2.61
TP	$m\ m^{-3}$	0.55	0.60	0.59	0.63	0.66	0.52	0.46	0.66
AW	$kg\ kg^{-1}$	0.13	0.14	0.23	0.21	0.24	0.18	0.17	0.25
pH		4.52	4.47	4.84	4.15	5.32	5.46	6.31	4.69
	-	*Soil chemical attributes*							
K	$mg\ dm^{-3}$	35.25	59.56	96.00	77.25	123.56	29.56	27.00	60.56
Ca	$cmol_c\ dm^{-3}$	1.21	2.34	0.70	0.39	2.54	1.14	1.17	2.14
Mg	$cmol_c\ dm^{-3}$	0.15	0.49	0.19	0.13	0.48	0.20	0.19	0.58
Al3	$cmol_c\ dm^{-3}$	0.68	0.78	1.10	2.56	0.49	0.03	0.00	0.25
HAl	$cmol_c\ dm^{-3}$	14.38	16.51	6.60	13.81	6.35	2.68	1.07	6.44
CEC	$cmol_c\ dm^{-3}$	15.82	19.50	7.74	14.52	9.69	4.10	2.50	9.31
ASI	%	31.05	24.16	49.70	78.10	16.37	1.88	0.00	10.28
EP	$mg_c\ L^{-1}$	25.13	25.78	27.48	19.66	26.64	40.81	45.54	44.64
PEP	-	0.12	0.11	0.10	0.12	0.08	0.13	0.95	0.13
As	$mg.kg^{-1}$	6.91	0.00	9.64	19.53	0.00	0.00	0.00	3.06
Co	$mg.kg^{-1}$	5.77	5.69	0.00	9.96	13.92	6.37	10.59	12.32
Pb	$mg.kg^{-1}$	38.34	30.28	2.69	20.54	24.06	26.07	30.26	22.08
Ni	$mg.kg^{-1}$	2.63	1.00	2.78	90.15	114.80	2.23	2.31	60.09
Zn	$mg.kg^{-1}$	40.94	31.62	0.00	67.28	67.42	19.63	32.09	35.15
Mn	$mg.kg^{-1}$	628.90	391.50	48.50	450.11	681.37	3746.98	3573.97	2782.60
Fe	$g.kg^{-1}$	236.61	245.94	19.45	87.39	88.90	242.29	265.25	143.92
Al	$g.kg^{-1}$	23.15	15.57	17.00	28.15	22.26	4.85	5.05	14.77
		Soil organic-biological attributes							
TOC	$dag.kg^{-1}$	8.81	9.79	3.57	7.41	3.34	0.58	0.42	3.02
CCO2	$mg\ C\text{-}CO_2.kg^{-1}$	106.02	144.29	171.76	199.01	161.98	96.32	61.04	182.19
SBR	$mg\ C\text{-}CO_2.kg^{\perp}.h^{\perp}$	0.21	0.29	0.34	0.39	0.32	0.19	0.12	0.36
MBC	$mg.kg^{-1}$	398.88	565.83	133.10	1080.92	187.44	147.26	114.28	372.78
MBN	$mg.kg^{-1}$	29.07	48.03	26.40	100.14	36.41	20.21	13.88	34.11
qMic	%	0.50	0.61	0.39	1.47	0.56	7.56	2.80	1.28
qCO2	$x_c 10^{-3}$	0.54	0.48	6.91	0.37	1.94	3.73	1.30	1.32
		Vegetation parameters							
COV	%	30.88	97.77	86.74	100.00	99.89	81.59	62.69	100.00
VOL	m^{-3}	15.02	91.12	3.92	1922.95	73.14	14.87	29.19	1486.06

Source: adapted from Alves (2019)

Note: *FR*: ferruginous rupestrian grassland with thin vegetation; *FR-D:* ferruginous rupestrian grassland with dense vegetation; *QR*: quartzite rupestrian grassland; *AF*: Atlantic Forest; *COM5*: compensation area, 5 years under reclamation; *SP15*: sterile pile, 15 years under reclamation; *Pit15*: pit, 15 years under reclamation; *SP20:* sterile pile, 20 years under reclamation; *Ston*: stoniness; *FS*: fine sand; *FSS*: fine sand + silt; *BD*: bulk density; *PD*: particle density; *TP*: total porosity; *AW*: available water content; *pH*: pH in water; *K*: available potassium, K^+; *Ca*: exchangeable calcium, Ca^{2+}; *Mg*: exchangeable magnesium, Mg^{2+}; *Al3*: exchangeable aluminum, Al^{3+}; *HAl*: potential acidity, $H + Al^{3+}$; *CEC*: cation exchange capacity; *ASI*: aluminum saturation index; *EP*: equilibrium phosphorus; *PEP*: available phosphorous/equilibrium phosphorus; *As*: semi-total arsenic; *Co*: semi-total cobalt; *Pb*: semi-total lead; *Ni*: semi-total nickel; *Zn*: semi-total zinc; *Mn*: semi-total manganese; *Fe*: semi-total iron; *Al*: semi-total aluminum; *TOC*: total organic carbon; *CCO2*: microbial respiration; *SBR*: soil biomass respiration; *MBC*: microbial biomass carbon; *MBN:* microbial biomass nitrogen; *qMic*: microbial quotient; *qCO2*: metabolic quotient; *COV*: vegetation coverage; *VOL*: vegetation volume.

3.3 Networks structure and layout

The interactions between areas (reference areas and rehabilitating ones) and soil attributes (or vegetation parameters) were established as weighted bipartite networks, undirected, divided in two classes: (i) *area*; and (II) *attribute*. For some networks, the *area-projection* and/or the *attribute-projection* are also generated, depending on the goal of the analysis in each one.

All networks were generated using *Gephi* software, version 0.9.2, using the *ForceAtlas2* graph layout, either the standard mode or the *LinLog* mode, depending on the desired effect. Some networks were rotated, just to make them fit the paper or to give them a better aesthetic effect, without prejudice to their content. For some networks, repetitions were made with the purpose of confirming the results or obtaining layouts that better fit the data or present a better aesthetic effect.

We have run two network statistics: *Average weighted degree*, to obtain weighted degree distribution graphs, and *Modularity* (calculated according to Blondel et al (2008)) , in order to identify communities (or groups of nodes). Nodes that have the lowest weights in the *attribute-projections* and visually impacted the spatialization of the networks, extending beyond each community, were considered as the most relevant for discriminating the areas. This selection was made considering the weighted degree distribution (reported by *Gephi*), in which the weights appear separated into distinct groups (examples in Appendix D).

The appearance of each network (and their projections) was modified by changing the color of the nodes, according to the object of interest: the type of fertilization, the soil attribute class (physical, chemical or organic-biological), the weighted degree, the number of communities, the bipartite class or in distinguishing specific nodes. In this way, several figures were created, in order to analyze different aspects of the same network. The edges have been shown as curved or not, just to achieve a better aesthetic effect.

The size of nodes and the thickness of the edges (or links) were established as proportional to the values of their total weight: the greater the weight, the greater the size of the node and the thickness of the edge. However, in some networks we chose to show the nodes of the attributes in a larger size than the real one, so that they would be more visible.

3.4 Weights and similarity indices calculation.

Data shown in Tables 1, 2 and 3, forming the basis for the construction of the networks. For each network a table was assembled, starting from Tables 1, 2 or 3, according to the areas or attributes of interest. Thus, an *attribute-area* matrix (M_N) was prepared for each *Network-N* (where *N* is the network number), whose elements are the weights of the edges (w_{ia}) which connect each attribute (*i*) with each area (*a*), according to the equation:

$$w_{ia} = \frac{V_i}{V_L} \qquad (1)$$

where V_i is the value of attribute *i* for each area *a*, taken from Tables 1, 2 or 3;

V_L is the largest value among values V_i, in each row of Tables 1, 2 or 3.

Note*:* equation (1) adjusts the weights so that they are in the range between 0.000 and 1.000. It is worth noting that when the edge weight is 0.000 there is no link between the attribute and the area.

In order to compare any area (a) with some other area taken as reference (A), new matrices (M_{Nr}) have been created, composed of relative edge weights (w_{ir}), calculated for each edge of the matrix, according to the following equation:

$$w_{ir} = 1 - |w_{ia} - w_{iA}| \qquad (2)$$

where w_{ia} is the edge weight, calculated according to equation (1)*;*

w_{iA} is the edge weight w_{ia} for the area taken as reference.

Note: since $w_{iA} = w_{ia}$ for referential areas, the relative edge weights (w_{ir}) of these areas are always equal 1.000. Networks built with w_{ir} have exactly the same nodes of *Networks-N* and were called *Networks-Nr.*

All matrices were transformed into adjacency lists and exported to *Gephi* software, in the form of *Excel csv* spreadsheets. For a better understanding of the method used in this work, the matrices for *Network-1* (Matrix M_1).and *Network-1r* (Matrix M_{1r}), both from the bauxite mined area, were presented as an example in Appendix A. All other matrices were constructed in a similar way.

A *Similarity Index,* $SI_{(A-a)}$, was created to measure the similarity between the area taken as reference (*A*) and any other of interest (*a*), and was calculated as following:

$$SI_{(A-a)} = \frac{Wn_{(a)}}{Wn_{(A)}} \quad (3)$$

where $Wn_{(a)}$ is the node weight Wn, calculated by *Gephi* for the node *n of* area *a*;

$Wn_{(A)}$ is the node weight Wn, calculated by *Gephi* for the node *n of* area *A*;

Note: Wn is calculated by *Gephi* linearly and corresponds to the simple sum of the edge weights (w_{ia} *or* w_{ir}) in the columns (*area*) of the matrices M_N or M_{Nr}, respectively.

In the case of *area-projections*, the similarity index was named to *Relative Similarity Index,* $RSI_{(A-a)}$, and equation (3) was adapted to:

$$RSI_{(A-a)} = \frac{Wnp_{(a)}}{Wnp_{(A)}} \quad (4)$$

where $Wnp_{(a)}$ is the node weight Wn calculated by *Gephi* for area *a* in the area-projection *p;*

$Wnp_{(A)}$ is the node weight Wn calculated by *Gephi* for area *A* in the area-projection *p.*

Note: Wnp is calculated by *Gephi* from the internal links arising from the projection and seems to take into account the proportions between the weights w_{ia} *or* w_{ir} in each row of the matrix. In this way, the weights are automatically weighted by each other and the ranking between node weights Wnp is sometimes slightly different from that between node weights Wn. This is because Wnp is just not a simple sum of edge weights w_{ia} *or* w_{ir}, better reflecting the complexity of the data.

In networks where quality criteria (*c*) were used to distinguish the areas, the edge weights were named w_{iq}, the relative quality weight of the edges of Matrix M_{Nq}, calculated according to *equation 5*, as follows:

$$w_{iq} = 1 - |w_{ia} - w_{ic}| \quad (5)$$

where w_{ia} is the weight of the edge, calculated according to equation (1)*;*

w_{ic} is the weight (w_{ia}) of the edge taken as reference according to one of the established quality criteria (*c*): *more is better, less is better* or *optimal value.*

Note: since $w_{ic} = w_{ia}$ for the reference edge, their relative quality weight (w_{iq}) are equal to 1.000. However, for networks of bauxite mined areas, when the *optimal value* taken from the literature did not match any of the values in Table 1, the edge weight (w_{ic}) was calculated according to equation (6), below:

$$w_{ic} = \frac{V_{lit}}{V_L} \qquad (6)$$

where V_{lit} is the value taken from the literature;

V_L is the largest value among values V_i, in each row of Tables 1, 2 or 3.

Note: in this case, none of the edges in the matrix have a value equal to 1.000, but that edge whose w_{ia} is closest to the literature value will have the w_{iq} value closets to 1.000.

In some cases, a network derived from the original (*Network-N*) was created, called *Network-Np*. This was done by deleting some nodes (attributes) from the original network and keeping only the attributes selected in PCA. For most of these networks no layouts were generated because the goal was only to compare the weighted degree distribution of areas in *Network-N* with their distribution in *Network-Np.*

Some *Networks-Np* are used to calibrate the method developed in this work. These networks were built with exactly the soil quality indicators used by Borges et al (2019) to calculate the *SQI* (*Soil Quality Index*) and by Alves (2019) to calculate the *RQI* (*Recovery Quality Index*). The weights of nodes calculated by *Gephi,* both for the complete network (W_n) and for the *area-projection* (W_{np}), were compared with the *SQI* and *RQI* values found by these authors, for the bauxite and iron mined areas, respectively.

Some results were shown in the bar charts as a fraction of the largest value, for adjustment of the weight scales, so that comparison was possible. In these cases, they were called *relative weights.*

3.5 Networks built with Dat-1

For this dataset, the classes of the bipartite networks, with their respective nodes are: (i) **area**: *Forest, PreMin, PosMin, N-G, N-L, N-GL, N-NI, P-G, P-L, P-GL, P-NI, C-G, C-L, C-GL, C-NI, PC-G, PC-L, PC-GL, PC-NI*; (ii) **attribute**: *TOC, LOC, TN, P, K, Mn, Fe, Zn, pH, Al, HAl*, *EP, TEB, BD, Mic, Mac, TP, AciP, AlkP, Gluc, and CCO2*.

A total of 10 networks were constructed with Dat-1, divided into two groups, according to the type of evaluation: *Group 1*, to evaluate the similarity between areas; and *Group 2,* to evaluate the relative soil quality and to calibrate the method. Each network was generated with a different purpose, according to Table 4, below.

Table 4 – Networks built with Dat-1.

Group	Network	Areas	Attributes	Purpose
1	**1**	Forest, PreMin, PosMin	all	evaluate the impacts of mining
	1r	Forest*, PreMin, PosMin	all	evaluate the similarity with Forest*
	2	all	all	evaluate the effect of fetilization
	2r	all	all	evaluate the similarity with Forest*
	3	16 treatments	all	evaluate the differences between treatments
2	**4**	PreMin, PosMin, 16 treatments	all	assessing soil quality
	5	all	all	assessing the physical quality of soil
	6	all	all	assessing the chemical quality of soil
	7	all	all	assessing the organic-biological quality of soil
	2p	all	Indicators selected with PCA	method calibration

Source: the author.

Note: * area taken as reference. In *Network-4, Forest* was excluded because chemical quality criteria were used based on the maintenance of coffee farming.

To calculate the weights of the edges in *Group 1*, only equations (1) and (2) were used and in *Group 2*, the equations (5) and (6).

For *Networks 4, 5, 6* and *7*, the quality criteria were established for each soil attribute, according to the literature. For *TOC, LOC, TN, EP, TEB, Mic, Mac, TP, AciP, AlkP, Gluc,* the criterion *more is better* was adopted, because we understand that higher values of these attributes are related to better soil quality conditions. Conversely, for *Al3, HAl* and *BD*, the criterion *less is better* was adopted, because we understand that lower values for these

attributes are related to better soil quality. For *pH, Fe, Mn, Zn, P* and *CCO2*, a reference value was established, according to the *optimum value* criterion, that is, the best value would be the one that best represents the quality of the soil for the conditions evaluated. Thus, for *CCO2*, the optimal value was considered to be that of *PreMin* (in *Network-4*), and of *Forest* (in Network-7) as they are the most stable ecosystem, in relation to the other areas, of their respective networks.

For *pH, Fe, P*, *Mn* and *Zn, ,* the *optimum value* was determined according to Guimarães, Alvarez and Ribeiro (1999, p. 25, 29, 296, 298). Thus, the values considered as *optimum value* for each attribute (in each chart of the work cited above) was: for *pH,* the highest value of the class "good" (or adequate); for *Fe*, the upper limit of the "medium" class (or critical level); for *P*, the upper limit of the "medium" class, as a function of the *EP* value; for *Mn* and *Zn*, the upper value of the class "good". For these last three attributes, the optimal value was set to the maintenance condition of coffee farming.

The network used to calibrate the method (*Network-2p*) were built with exactly the soil quality indicators selected by Borges et al. (2019) in the PCA: *LOC, TEB, BD* and *AciP*. The calculation of edge weights (w_{iq}) was done by multiplying w_{ia} by the "Weights in the SQI" (BORGES *et al.*, 2019, Table 4) used by these authors*: LOC* = 0.30; *BD* = 0.30; *AciP* = 0.29 and *TEB* = 0.11.

Through the *SI* and *RSI* values it was possible to graphically observe the differences between the results obtained with Network Theory and with SQI, in two different ways: from the complete network and from its *area-projection*.

3.6 Networks built with the Dat-2

For this dataset, the classes of the bipartite networks, with their respective nodes are: (i) **area**: *FR, FR-D, QR, AF, COM5, SP15, Pit15* and *SP20* (ii) **attribute**: *Ston, FS, FSS, Silt, Clay, BD, PD, TP, AW, pH, K, Ca, Mg, Al3, HAl, CEC, ASI, EP, PEP, TOC, As, Co, Pb, Ni, Zn, Mn, Fe, Al, CCO2, SBR, MBC, MBN, qMic, qCO2, COV* and *VOL.*

A total of 11 networks were built, divided in three groups: *Group 1***,** showing the relationships between the areas, according to equation (1); *Group 2***,** the similarity between rehabilitating areas and each reference, according to equation (2)*;* and *Group 3*, the

qualitative relationships between rehabilitating areas and each reference, according to equation (5). The purpose in each network is shown in Table 5, bellow.

Table 5 – Networks built with Dat-2.

Group	Network	Areas	Attributes	Purpose
1	**8**	referencial areas	all	discriminate the reference areas
	9	all	all	discriminate all areas
	10	all plots of rehabilitating areas	all	evaluate the dispersion of data
2	**11**	FR* rehabilitating areas	all	Compare the rehabilitating areas with the reference
	12	FR-D* rehabilitating areas	all	
	13	QR* rehabilitating areas	all	
	14	AF* rehabilitating areas	all	
3	**11p**	FR* rehabilitating areas	Inicators selected with PCA	Method calibration
	12p	FR-D* rehabilitating areas		
	13p	QR* rehabilitating areas		
	14p	AF* rehabilitating areas		

Source: the author.

Note: * area taken as reference.

To evaluate the importance of each attribute class (physical, chemical, organic-biological and vegetation) in *Netwok-9*, the area-projections and attribute-projections were generated keeping only the attributes of each class at a time.

The networks in *Group 2* and *Group 3* were composed with the same areas established by Alves (2019) in his comparison groups. This author compared each of the reference areas (*FR, FR-D, Q* and *AF*) with the four areas in recovery (*COM5, SP15, Pit15* and *SP20*), in order to compose four groups, and used the indicators selected through PCA for each group, to calculate the RQI. We did the same for the *Group 3* networks, but in *Group 2* we used all 36 attributes evaluated. This strategy was used so that the results of the original networks (*Networks 11* to *14*) could be compared with those obtained in the indicators networks (*Networks 11p* to *14p*).

No layouts were generated for *Group 3* networks: we considered only the weighted degree distribution of the bipartite networks and their projections, keeping in the original networks (Networks *11* to *14*) only the attributes selected by Alves (2019) with PCA. The quality criterion used by this author was only the *optimal value*, coinciding with the values of Table 2 for each reference area. That is, in this case, the *optimal value* criterion corresponds to the concept of similarity used by us for referential areas. Thus, the weight of the edges remained the same, at first, since all edges of the references remained with a weight of 1.000.

For the calibration of the method, the edge weights were multiplying by the weights calculated by Alves (2019, p.111, 112, 114 and 116) from PCA for each attribute. Thus, the results obtained with the Network Theory for the *Group 3* networks were compared with the RQI, in the same way as was done for the bauxite mined areas.

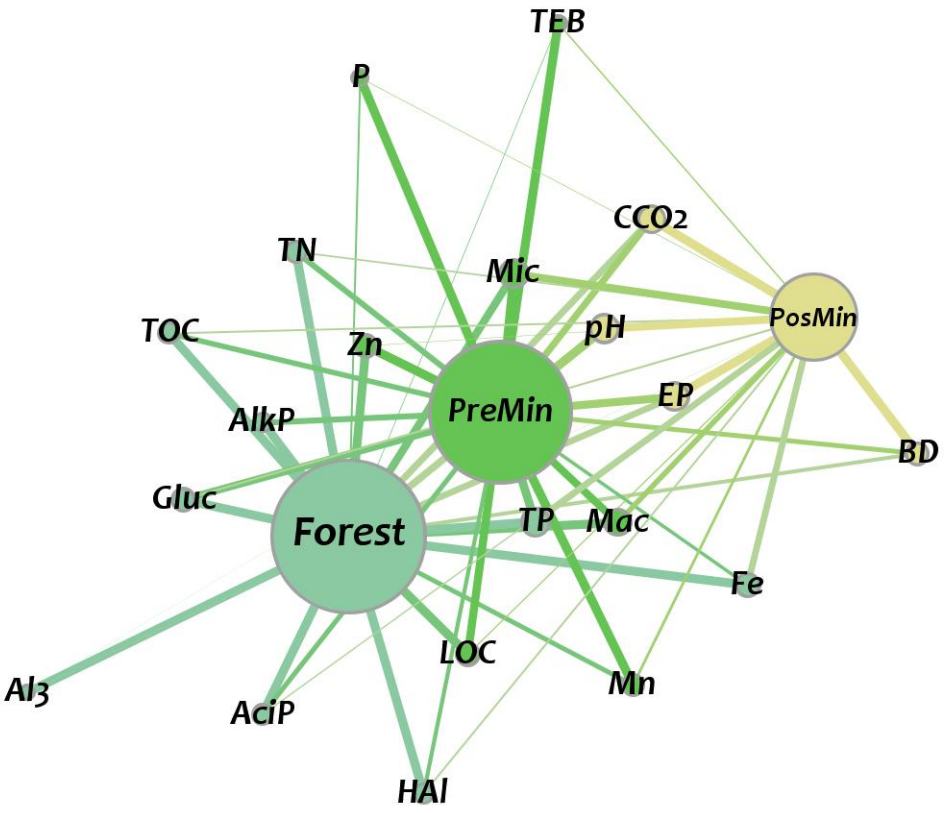
4 RESULTS AND DICUSSION

4.1 Bauxite mined area – *Group 1* networks

4.1.1 Mining impact assessment

Network-1 shows the relations between *Forest, PreMin, PosMin* and the soil attributes, evidencing the greater proximity and similarity between the unmined areas (*Forest* and *PreMin*) in relation to *PosMin*. (Figure 12). Three communities, or modularity classes – [0], [1] and [2], grouping the attributes with the highest edge weights (w_{ia}) in each area (see Matriz M_1, Appendix A).

Figure 12 – The three communities of *Network-1.*



Source: the author.

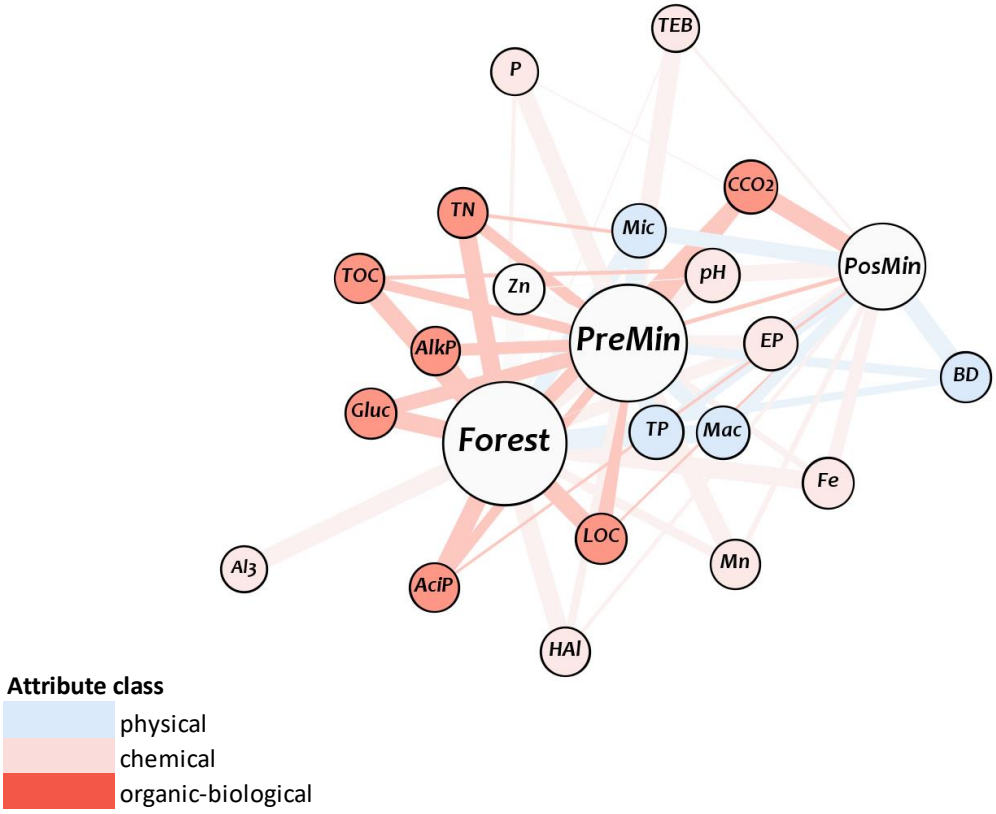
Modularity class	
[0]	*Forest, TOC, TN, Fe, Al3, HAl, TP, AciP, AlkP* and *Gluc*
[1]	*PreMin* , *LOC, P, Mn, Zn, TEB, Mic* and *Mac*
[2]	*PosMIn, pH, EP, BD* and *CCO2* .

Note: *Forest:* Atlantic forest; *PreMin*: 10-year-old coffee crop, before mining; *PosMin*: reconstructed soil, six months after mining.

The attributes that appear on the periphery of the network (and seen to project out of it) are the most important in distinguishing each area. The thickness of the edges proves this, for example for *TEB* and *HAl.* These attributes are among the least weighted (see Appendix B), precisely because the weight of nodes is obtained by adding the weights of edges. It is interesting to note that, although *BD* does not have such a low weight when compared to others, it is the attribute that most differentiates *PosMin*, evidencing the soil compaction caused by mining.

Figure 13 shows the same network, now with the attributes colored according to each class: physical, chemical and biological.

Figure 13 – The highlighted attribute classes of *Network-1.*



Source: the author.

Note: nodes of the attributes were represented in a larger size than the real one. *Forest:* Atlantic forest; *PreMin*: 10-year-old coffee crop, before mining; *PosMin*: reconstructed soil, six months after mining.

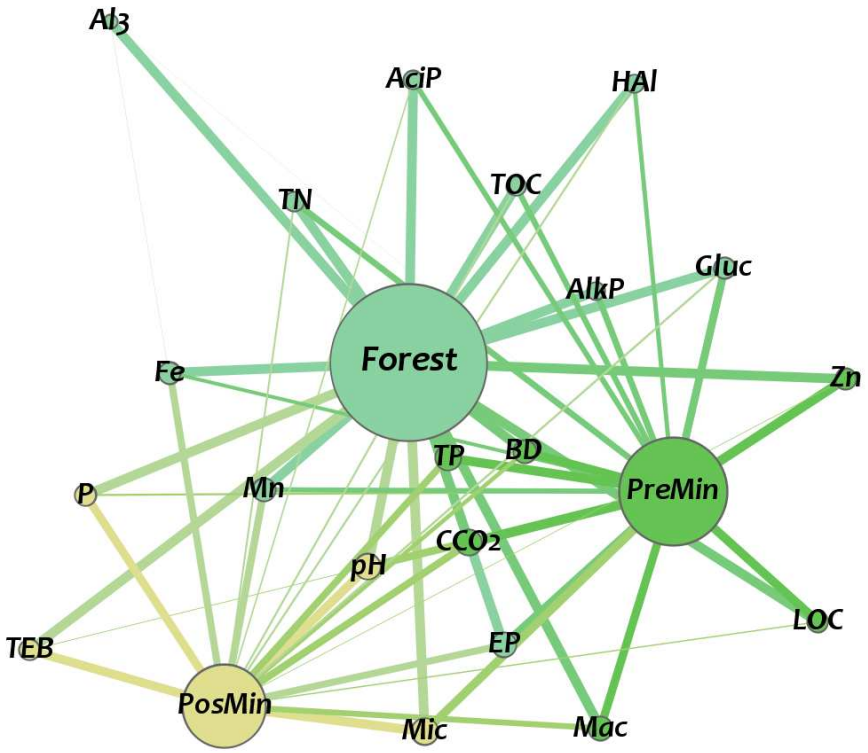
It is evident, then, the importance of the organic-biological attributes (in red) for the non-mined areas, especially for *Forest,* with the exception of *CCO2.* This also demonstrates how soil organic matter (SOM) is drastically impacted by mining, as the literature attests

(LORENZ; LAL; SHIPITALO, 2006; SHRESTHA; LAL, 2008). This result also highlights the high value of *CCO2* found in *PosMin*, a possible indication of imbalance in this ecosystem.

4.1.2 Similarity of *PreMin* and *PosMin* to *Forest*

Network-1r has the same nodes of *Network-1* and it also has three communities, but has *Forest* as reference area (Figure 14).

Figure 14 – The three communities of *Network-1r.*



Modularity class

[0]	*Forest, TOC, TN, Mn, Fe, Al3, HAl, EP, AciP, AlkP* and *Gluc*
[1]	*PreMin , LOC, Zn, BD, Mac, TP* and *CCO2*
[2]	*PosMIn, P, pH, TEB* and *Mic*

Source: the author.

Note: *Forest:* Atlantic forest; *PreMin*: 10-year-old coffee crop, before mining; *PosMin*: reconstructed soil, six months after mining.

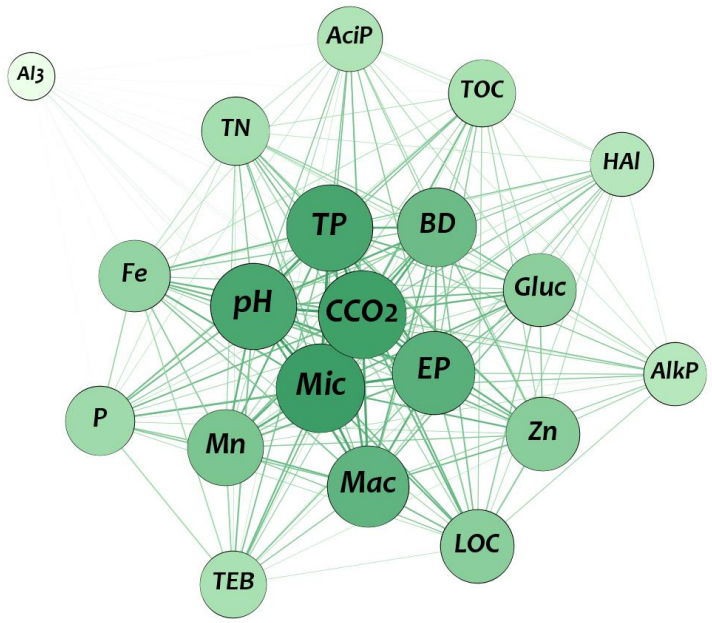
In this type of network, all edges in the reference area have weight 1.000 (see *Matrix* M_{1r}, Appendix A) and so the interpretation of communities is different from the previous network. Here, the attributes at the periphery of the network, in communities [1] and [2], are

those that have a weight closest to that of the reference area, as can be seen by the thickness of the edges, for example in *LOC, Gluc* and *TEB.*

Figure 15 shows the attribute-projection of *Network-1r* and the ranking of the attribute weights (Figure 15b). Those who best distinguish *Forest* from other areas are the ones with the lowest weights, found in the periphery of the projection (nodes that have the highest weights are in the center).

Figure 15 – *Network-1r attribute-projection* and its node weights.

(a)



(b)

Al3	20.777
AlkP	**27.762**
HAl	28.055
AciP	**28.714**
TEB	29.510
TOC	**29.689**
TN	**30.114**
P	30.651
Fe	31.685
Gluc	**32.280**
LOC	**32.453**
Zn	32.652
Mn	33.769
BD	34.953
Mac	36.059
EP	36.660
pH	37.974
TP	38.203
CCO2	**38.942**
Mic	39.242

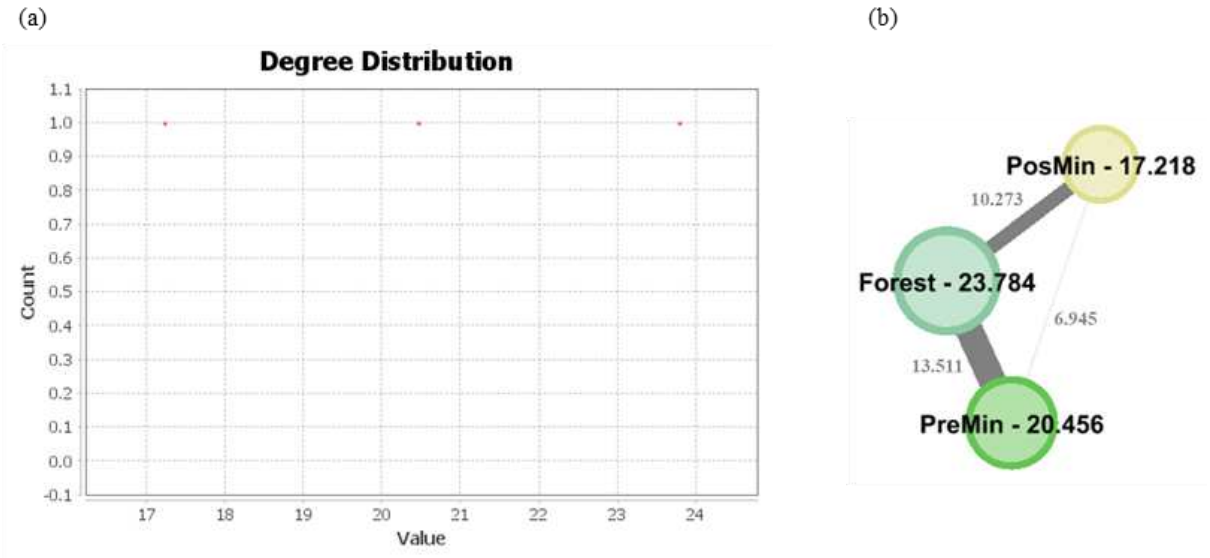
Source: the author.

Note: (a) the attribute-projection of *Network-1r*; (b) node weights of *Network-1r attribute-projection* calculated by *Gephi,* in ascending order (in emphasis, the organic-biological attributes). *TOC*; total organic carbon; *LOC*: labile organic carbon; *TN*: total nitrogen; *P*: available phosphorous; *Mn*: manganese content; *Fe*: iron content; *Zn*: zinc content; *pH*: pH in water; *Al3*: exchangeable aluminum, Al^{3+}; *HAl*: potential acidity, $H + Al^{3+}$; *EP*: soil equilibrium phosphorus; *TEB*: total exchangeable basis, $K^{+} + Ca^{2+} + Mg^{2+}$; *BD*: bulk density; *Mic*: microporosity; *Mac*: macroporosity; *TP*: total porosity; *AciP*: acid phosphomonoesterase enzyme activity; *AlkP*: alkaline phosphomonoesterase enzyme activity; *Gluc*: ß-glucosidase enzyme activity; *CCO2*: microbial respiration.

As expected, most of the organic-biological attributes are among the first, confirming their importance in this natural environment. However, *Al3* is the most evident attribute in this distinction, because quantity of Al3 is much higher in Forest than the other two areas, as can be seen in Table 1.

To calculate the *Relative Similarity Index* (*RSI),* the area-projection of *Network-1r* was generated, shown in Figure 16.

Figure 16 – *Network-1r area-projection* and its weighted degree distribution.



Source: the author.

Note: (a) the weighted degree distribution of *Network-1r*, reported by *Gephi*; (b) the *Network-1r area-projection*, showing node weights (in black) and edge weights (in gray), calculated by *Gephi. Forest:* Atlantic forest; *PreMin*: 10-year-old coffee crop, before mining; *PosMin*: reconstructed soil, six months after mining.

The *Similarity Indices* (*SI*) calculated for *Network-1r,* according to equation (3), were:

$$SI_{(Forest-PosMin)=0.514}$$

$$SI_{(Forest-PreMin)=0.676}$$

The *RSI* calculated for *Network-1r,* according to equation (4), were:

$$RSI_{(Forest-PosMIn)=0.724}$$

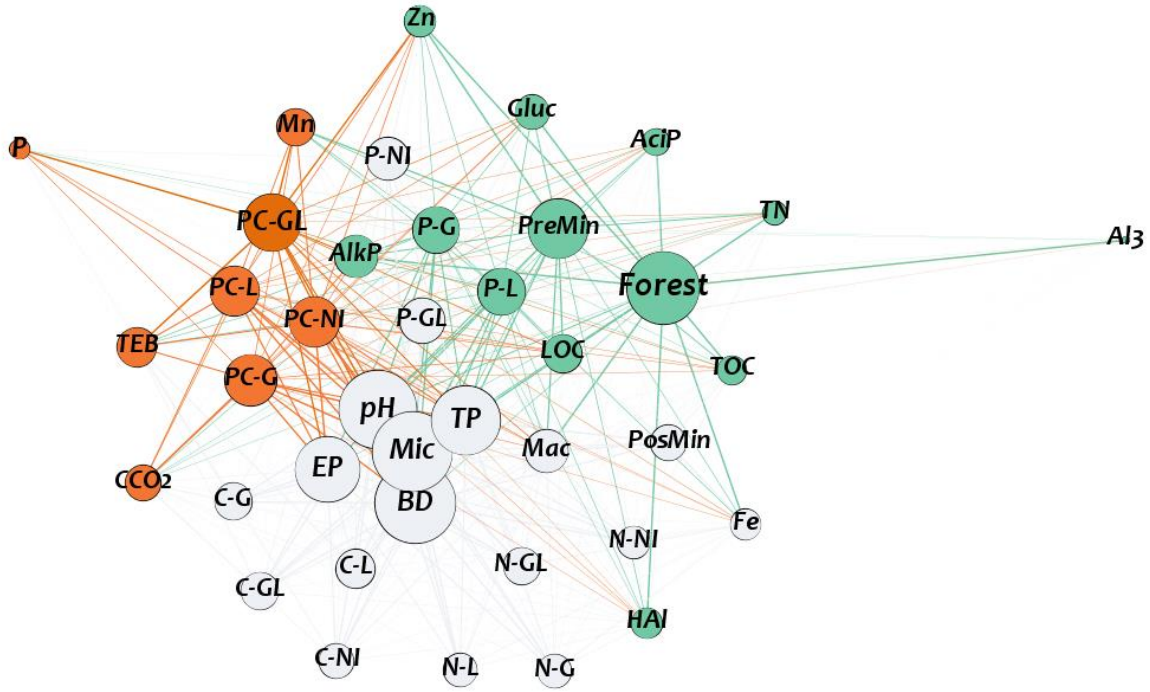
$$RSI_{(Forest-PreMin)=0.860}$$

The *SI* value tell us that the similarity between *PosMin* and *Forest* is only 51,4 %; and between *PreMin* and *Forest* is 67,7%. However, the *RSI* are higher for both, although approximately the same proportions maintained. Therefore, numerically two different results can be obtained depending on whether one uses the node weights in the complete network (as in *SI*), or the respective weights in their projections (as in *RSI*). These differences will be considered later in the calibration of the method (see item 4.2.3).

4.1.3 Fertilization influence on the distinction of areas

Network-2 shows the relations between all areas under study and, like *Network-1*, constitutes an initial exploration of the data, based on equation (1). The most important attributes for distinguishing the areas belong to communities [0] and [1], which have distinct profiles (Figure 17), similar to the two factors considered by Borges et al. (2019) in the PCA: Factor 1 and 2, respectively. Factor 1 was related by these authors to SOM, just as we can do with community [0], because it has the most organic-biological attributes; and Factor 2 was related to nutrient availability, represented in community [1] by *P, Mn* and *TEB*.

Figure 17 – The three communities of *Network-2.*



Modularity class	
[0]	*Forest, PreMin, P-G, P-L, TOC, LOC, TN, Zn, Al3, HAl, AciP, AlkP* and *Gluc*
[1]	*PC-G, PC-GL, PC-L, PC-Nl, P, Mn, TEB* and *CCO2*
[2]	*PosMIn, N-G, N-GL, N-L, N-Nl, P-GL, P-Nl, C-G, C-GL, C-L, C-Nl, Fe, pH, EP, BD, Mic, Mac* and *TP*

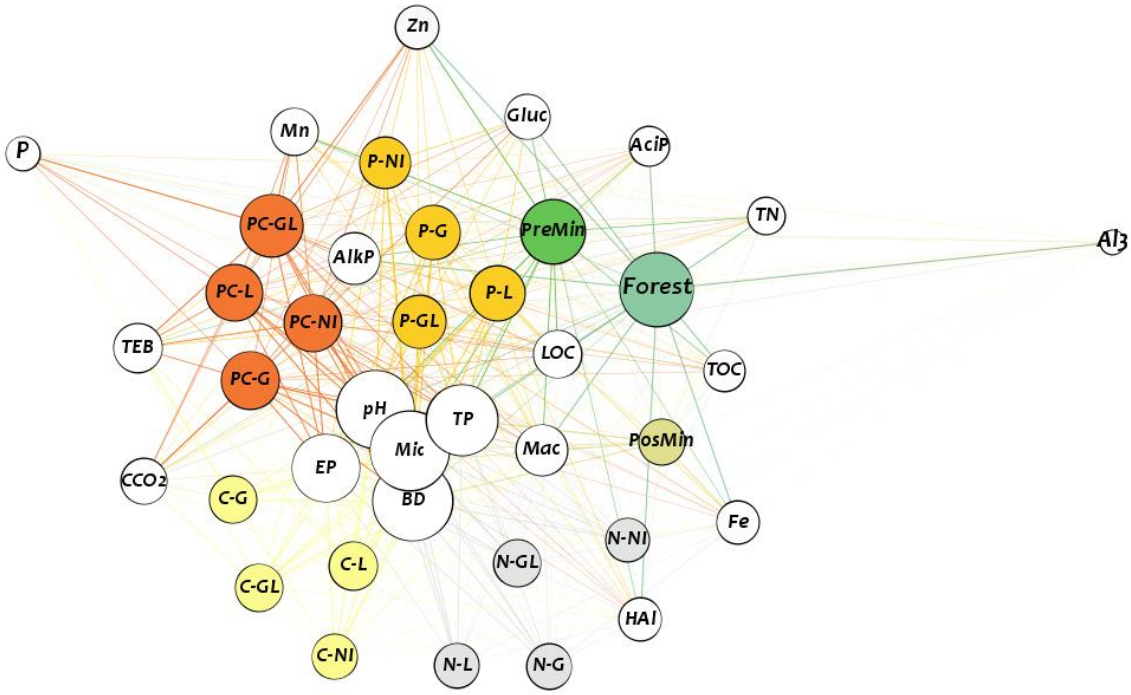
Source: the author.

Note: *Forest:* Atlantic forest; *PreMin*: 10-year-old coffee crop, before mining; *PosMin*: reconstructed soil, six months after mining; *N-*: no fertilization; *P-*: Poultry litter fertilization; *C-*: chemical fertilization; *PC-*: poultry litter and chemical fertilizations combined; *G*: grass intercrop; *L*: leguminous intercrop; *GL*: grass + leguminous mix intercrop; *Nl*: no intercrop.

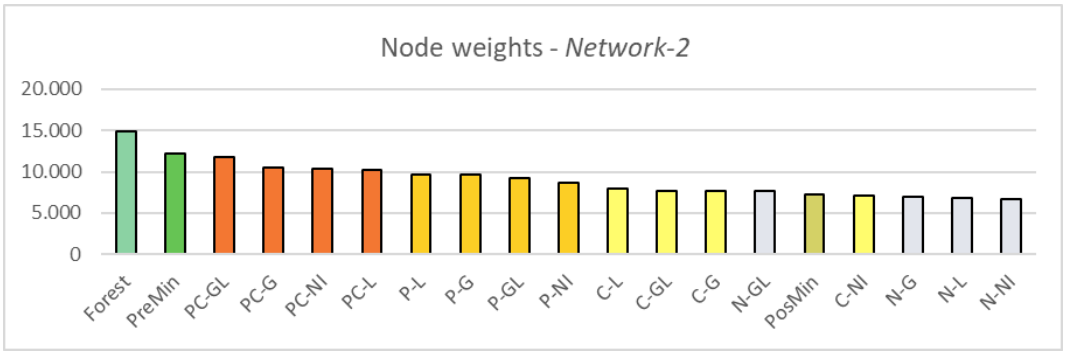
Figure 18a shows the same network, with the areas highlighted by colors. It is interesting how the areas that received the same type of fertilization are visually grouped, a result also obtained by Borges et al. (2019, p. 871, Figure 1) with the PCA. The distribution of node weights (Figure 18b) was also consistent with these results, which was not observed when the areas were colored according to the type of plant intercrop (see Appendix C).

Figure 18 –Highlighted reference areas and treatments of *Network-2.*

(a)



(b)



Fertilization type

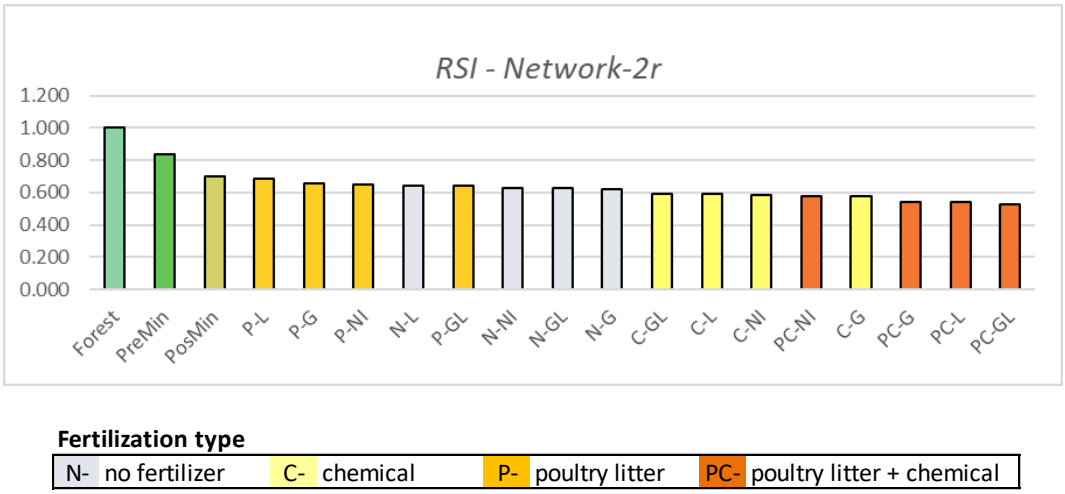
N-	no fertilizer	C-	chemical	P-	poultry litter	PC-	poultry litter + chemical

Source: the author.

Note: *Forest:* Atlantic forest; *PreMin*: 10-year-old coffee crop, before mining; *PosMin*: reconstructed soil, six months after mining. *G*: grass intercrop; *L*: leguminous intercrop; *GL*: grass + leguminous mix intercrop; *NI*: no intercrop.

Network-2r presents a different view on the effect of fertilization on the distinction of areas. This network is in some respects similar to the previous one, but has *Forest* as reference area (Appendix D). As can be seen in Figure 19, chemical fertilization increased the dissimilarity of the rehabilitated areas in relation to *Forest*, while organic fertilization alone brought them closer together. In fact, the amount of fertilizer also caused the same effect, since the most fertilized areas (PC- group) were those that were most distant from the reference area.

Figure 19 – *Relative Similarity Indices* of *Network-2r.*



Source: the author.

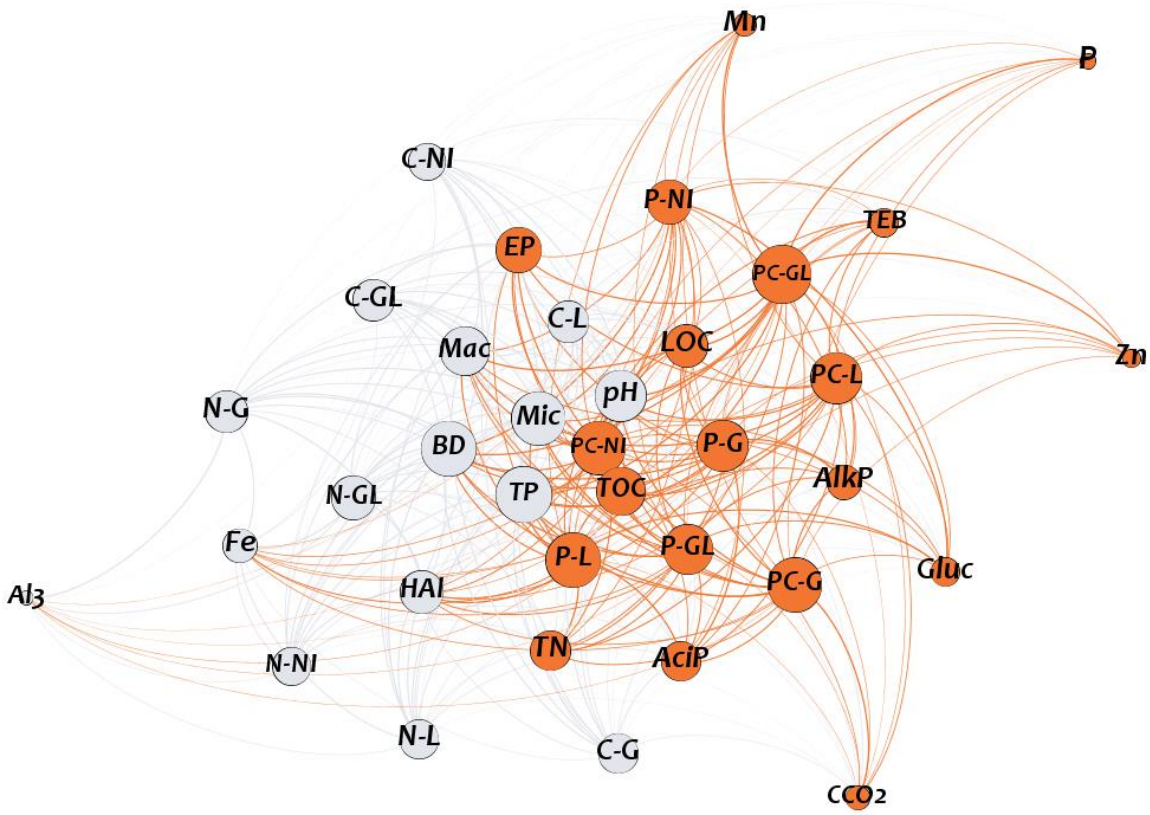
Note: *Forest:* Atlantic forest; *PreMin*: 10-year-old coffee crop, before mining; *PosMin*: reconstructed soil, six months after mining. *G*: grass intercrop; *L*: leguminous intercrop; *GL*: grass + leguminous mix intercrop; *NI*: no intercrop.

4.1.4 Importance of organic fertilization

Network-3 shows the relations between the 16 treatments, without any reference. The areas that received chemical fertilizers were completely separated from those that received organic fertilizer (Figure 20). This trend was also clearly observed in practice, for the *N-G, C-G, P-G* and *PC-G* treatments (see Annex B).

The two communities in this network also reflect this condition: all organic-biological attributes are part of class [0], along with the areas that received organic fertilization.

Figure 20 – The two communities of *Network-3*.



Modularity class

[0]	*P-G, P-GL, P-L, P-NI, PC-G, PC-GL, PC-L, PC-NI, TOC, LOC, TN, P, Mn, Zn, EP, TEB, AciP, AlkP, Gluc* and *CCO2*
[1]	*N-G, N-GL, N-L, N-NI, C-G, C-GL, C-L, C-NI, Fe, pH, Al3, HAl, BD, Mic, Mac* and *TP*

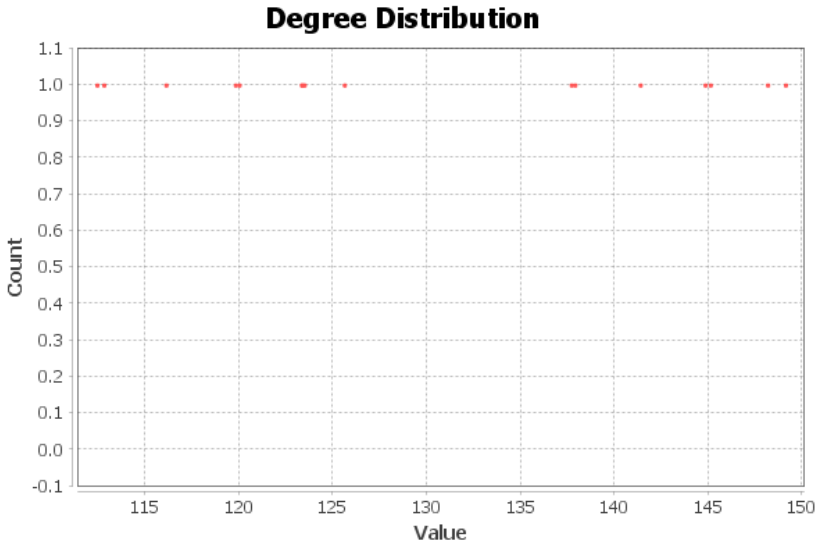
Source: the author.

Note: *N-*: no fertilization; *P-*: Poultry litter fertilization; *C-*: chemical fertilization; *PC-*: poultry litter and chemical fertilizations combined *G*: grass intercrop; *L*: leguminous intercrop; *GL*: grass + leguminous mix intercrop; *NI*: no intercrop.

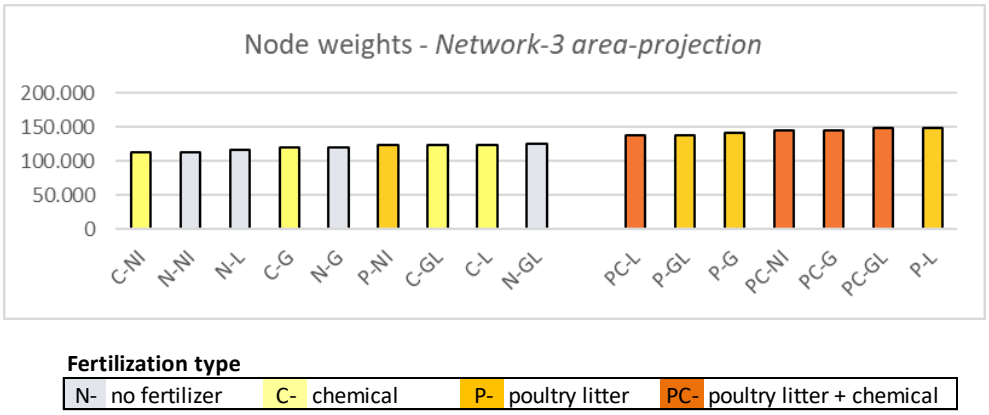
However, the weighted degree distribution of *Network-3 area-projection* shows a more refined result (reflecting the complexity of the data), placing *P-NI* in the opposite group (Figure 21). This result was also obtained by Borges et al. (2019) when they found that, with the exception of *PC-NI,* which received heavy fertilization, the areas without cover crop did not differ statistically from each other.

Figure 21 – The weighted degree distribution of *Network-3 area-projection.*

(a)



(b)



Source: the author.

Note: (a): weighted degree distribution of *Network-3 area-projection*; (b) node weights of *Network-3 area-projection* in ascending order. *G*: grass intercrop; *L*: leguminous intercrop; *GL*: grass + leguminous mix intercrop; *NI*: no intercrop.

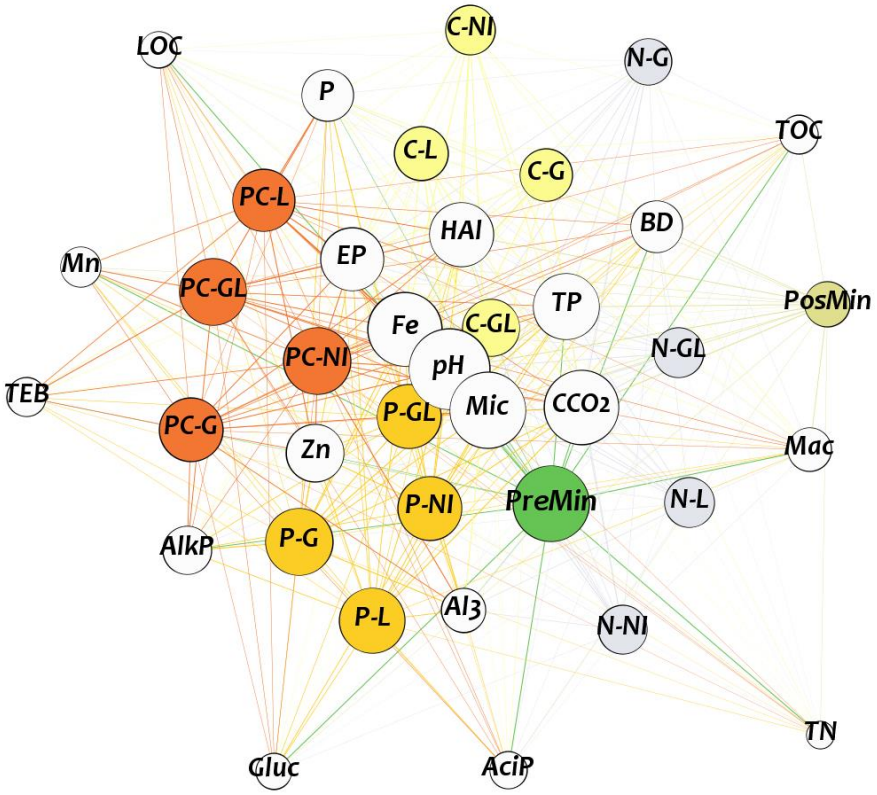
4.2 Bauxite mined area – *Group 2* networks

4.2.1 Soil quality in coffee-growing areas

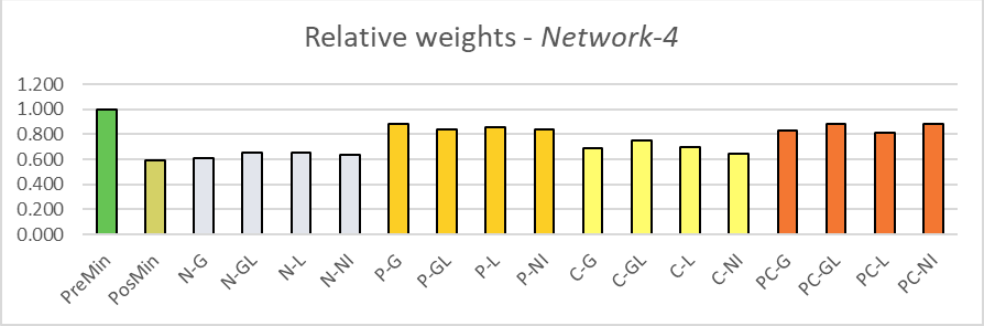
Network-4 shows the qualitative relations between *PreMin*, *PosMin* and the 16 treatments and, as can be seen, the clustering trend among the treatments that received organic fertilizer (*P*- and *PC*- groups) was maintained (Figure 22).

Figure 22 – Highlighted areas of *Network-4,* colored by type of fertilization.

(a)



(b)



Fertilization type

N-	no fertilizer	C-	chemical	P-	poultry litter	PC-	poultry litter + chemical

Source: the author.

Note: *PreMin*: 10-year-old coffee crop, before mining; *PosMin*: reconstructed soil, six months after mining; *N-*: no fertilization. *G*: grass intercrop; *L*: leguminous intercrop; *GL*: grass + leguminous mix intercrop; *NI*: no intercrop.

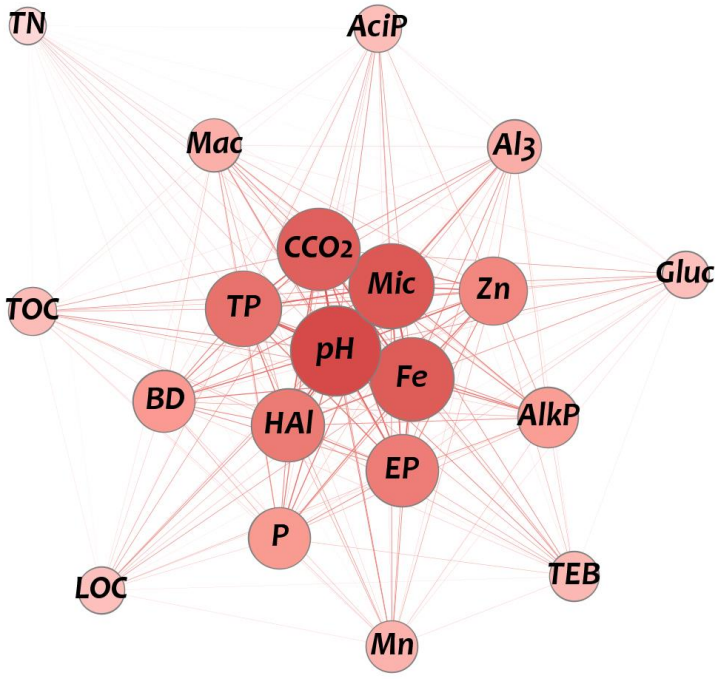
However, the soil quality in the treatments that received *P-* and *PC-* fertilizations with intercrops, when compared to *PosMin*, proved to be higher to than that found by Borges et al (2019) in their IQS: 45% in our evaluation, versus 23% in the IQS. In addition to not considering

all the attributes evaluated, the IQS assigned a much lower weight to *TEB* (possibly due to the inclusion of *Forest* in the PCA) than to the other indicators. This caused to soil quality in the treatments to be evaluated at a disadvantage, since *TEB* is a very relevant attribute in agricultural systems. In assessing the impact of mining on soil quality, there was also a great difference: 41% in our evaluation, versus 65% in the IQS, possibly because the absence of *Forest* in *Network-3*.

The *attribute-projection* of *Network-4* (Figure 23) also confirms our conclusions about the importance of organic fertilization and *TEB* in distinguishing the areas: the most relevant attributes (those with the lowest weights) were the organic-biological (*TN, LOC, Gluc, Acip* and *TOC*), followed by *TEB*, with the lowest weight among the chemical attributes (Figure 13b).

Figure 23 – Network-4 attribute-projection and its weighted degree distribution

(a)



(b)

TN	69.324
LOC	90.971
Gluc	91.089
AciP	92.214
TOC	94.526
TEB	97.554
Mn	102.414
Mac	105.509
Al3	106.791
AlkP	121.857
P	124.678
BD	125.161
Zn	139.670
EP	148.725
HAl	150.203
TP	156.763
CCO2	171.528
Fe	173.590
Mic	177.062
pH	188.000

Source: the author.

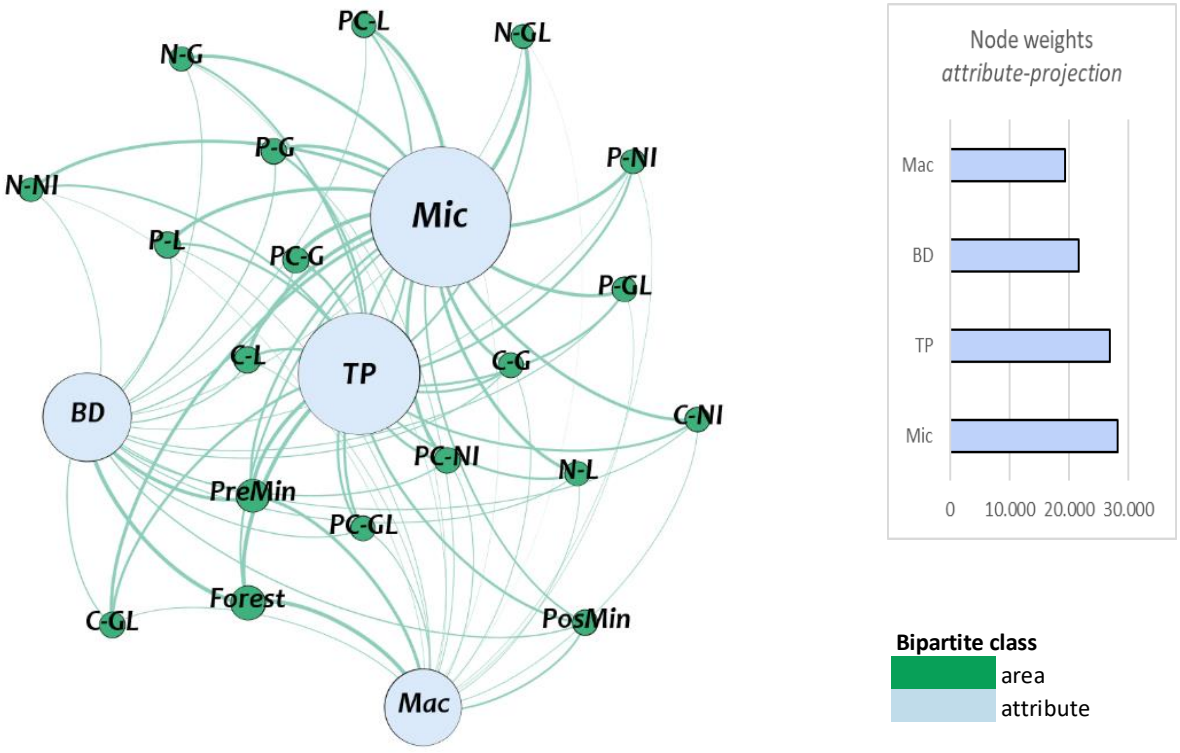
Note: (a) the attribute-projection of *Network-4*; (b) node weights of *Network-4 attribute-projection*, in ascending order; in emphasis, the organic-biological attributes. *TOC*; total organic carbon; *LOC*: labile organic carbon; *TN*: total nitrogen; *P*: available phosphorous; *Mn*: manganese content; *Fe*: iron content; *Zn*: zinc content; *pH*: pH in water; *Al3*: exchangeable aluminum, Al^{3+}; *HAl*: potential acidity, $H + Al^{3+}$; *EP*: soil equilibrium phosphorus; *TEB*: total exchangeable basis, $K^{+} + Ca^{2+} + Mg^{2+}$; *BD*: bulk density; *Mic*: microporosity; *Mac*: macroporosity; *TP*: total porosity; *AciP*: acid phosphomonoesterase enzyme activity; *AlkP*: alkaline phosphomonoesterase enzyme activity; *Gluc*: ß-glucosidase enzyme activity; *CCO2*: microbial respiration.

4.2.2 Physical, chemical and organic-biological quality of all areas

Networks 5, 6 and *7* show the qualitative relationships between all areas under study, in their physical, chemical and organic-biological aspects, respectively. Figure 24 shows the complete graph of *Network-5* (a) and the results of *Network-5 attribute projection* (b). As can be seen, among the physical attributes, the most relevant for distinguishing the areas were *Mac* and *BD.* They have stronger relationship with *Forest* and *PreMin,* precisely because they present higher Mac values and lower BD values, when compared with the other areas. This result confirms the effects of soil compaction caused by the mining process.

Figure 24 – *Network-5*: physical quality.

(a) (b)

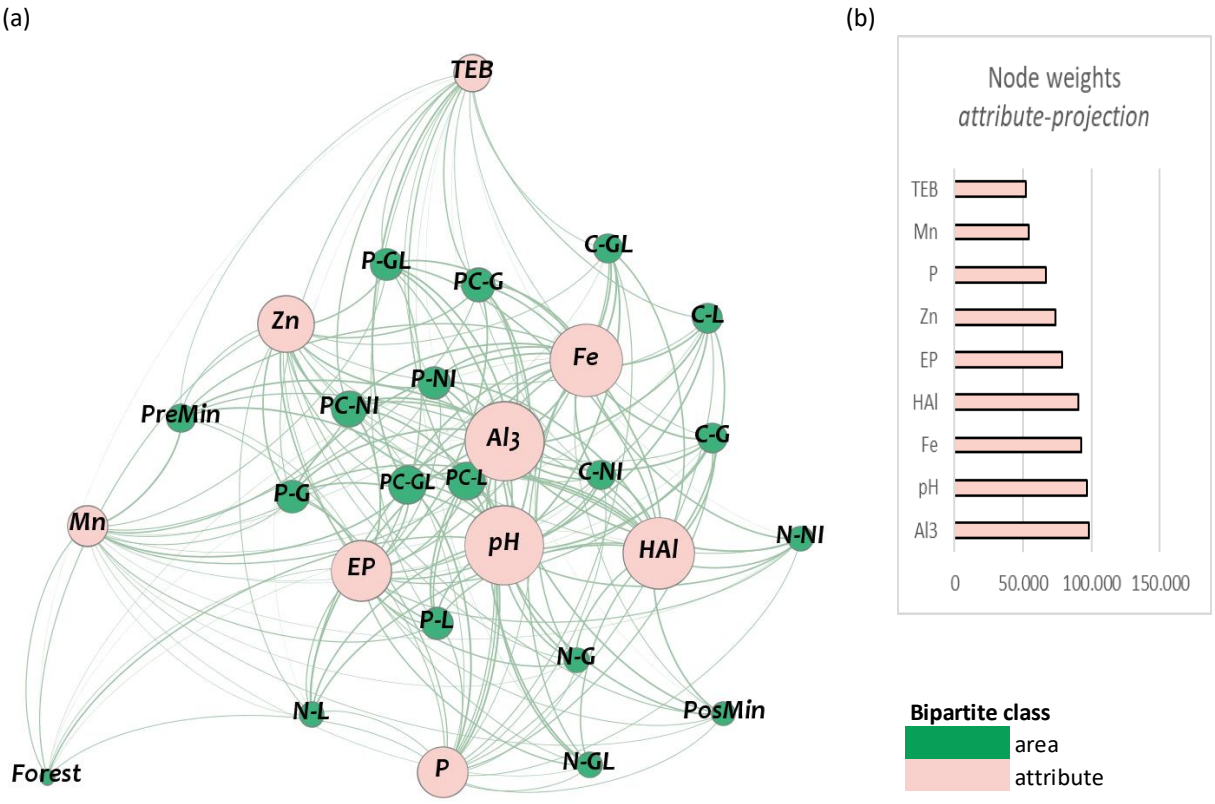


Source: the author.

Note: (a) the complete graph of *Network-5*; (b) node weights of attribute-projection of *Network-5*, calculated by *Gephi,* in ascending order. *BD*: bulk density; *Mic*: microporosity; *Mac*: macroporosity; *TP*: total porosity.

It is interesting to note how *Forest* appears shifted outwards in *Network-6* (Figure 25). This is because its attributes presented very low values, since the chemical quality criteria were established for the maintenance of coffee crop.

Figure 25 – *Network-6*: chemical quality.



Source: the author.

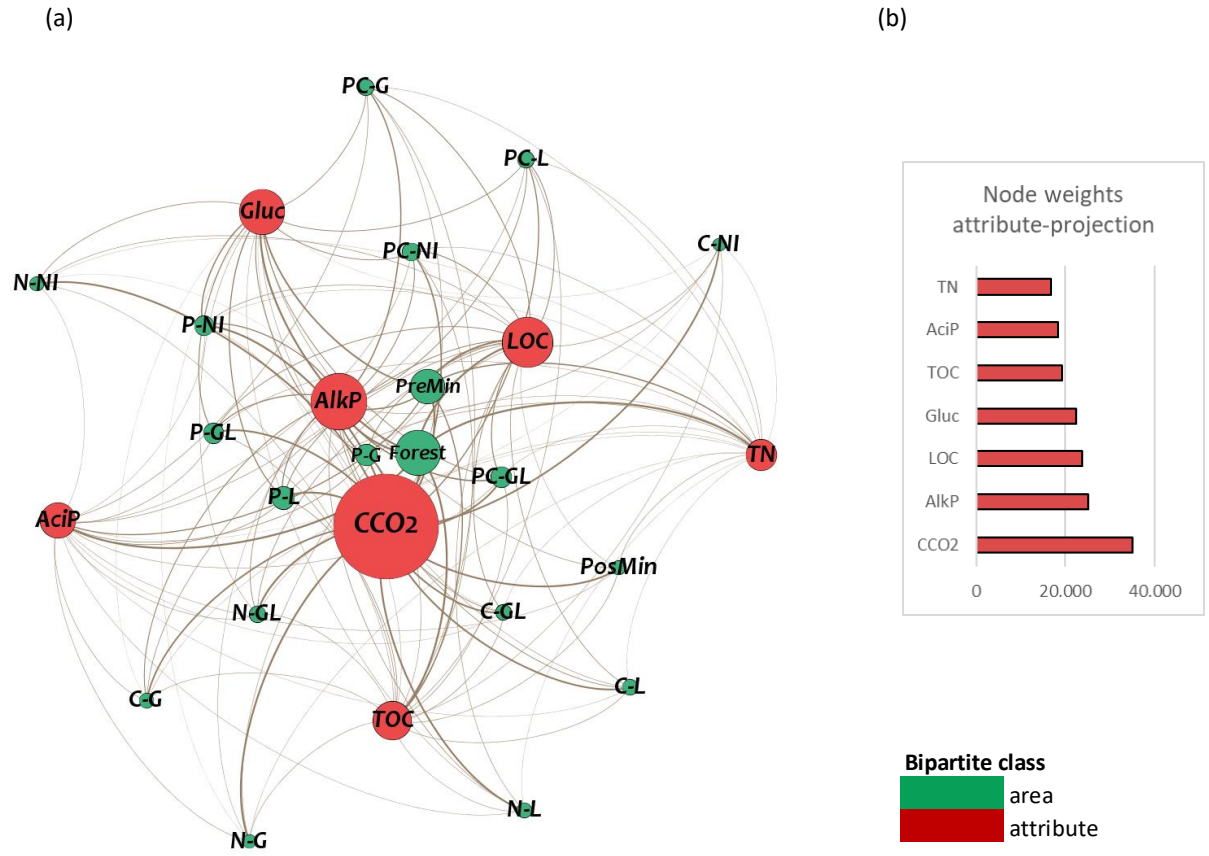
Note: (a) the complete graph of *Network-6*; (b) node weights of *attribute-projection* of *Network-6*, calculated by *Gephi,* in ascending order. *P*: available phosphorous; *Mn*: manganese content; *Fe*: iron content; *Zn*: zinc content; *pH*: pH in water; *Al3*: exchangeable aluminum, Al^{3+}; *HAl*: potential acidity, $H + Al^{3+}$; *EP*: soil equilibrium phosphorus; *TEB*: total exchangeable basis, $K^{+} + Ca^{2+} + Mg^{2+}$.

Thus, the most important attributes to distinguishing the areas were *TEB* and *Mn*. In relation to *Mn*, the lowest values were presented by the areas that did not receive fertilization (*N*-) or that received only chemical fertilization (*C*-). For *TEB*, the lowest values were presented by the areas that did not receive any fertilization: *Forest, PosMin* and *N*- treatments. The network layout shows the opposition between these areas and these attributes quite clearly.

On the other hand, the attribute with the highest weight was *Al3*, because it was rated as less is better. With the exception of *Forest*, all the other areas presented hight weights for this attribute.

When we evaluated the organic-biological quality of the areas, Forest turns to the center of the network, and now has the largest size among the areas, presenting exactly weight 1.000 for all attributes (Figure 26). In other words, Forest can be considered a true benchmark for soil organic quality.

Figure 26 – *Network-7:* organic-biological quality.



Source: the author.

Note: (a) the complete graph of *Network-7* (b) node weights of attribute-projection of *Network-7*, calculated by *Gephi,* in ascending order. *TOC*; total organic carbon; *LOC*: labile organic carbon; *TN*: total nitrogen; *AciP*: acid phosphomonoesterase enzyme activity; *AlkP*: alkaline phosphomonoesterase enzyme activity; *Gluc*: ß-glucosidase enzyme activity; *CCO2*: microbial respiration.

TN, AciP and *TOC* were the attributes most important to distinguishing the areas, just because *PosMin* and all 16 treatments showed very low weights for these attributes.

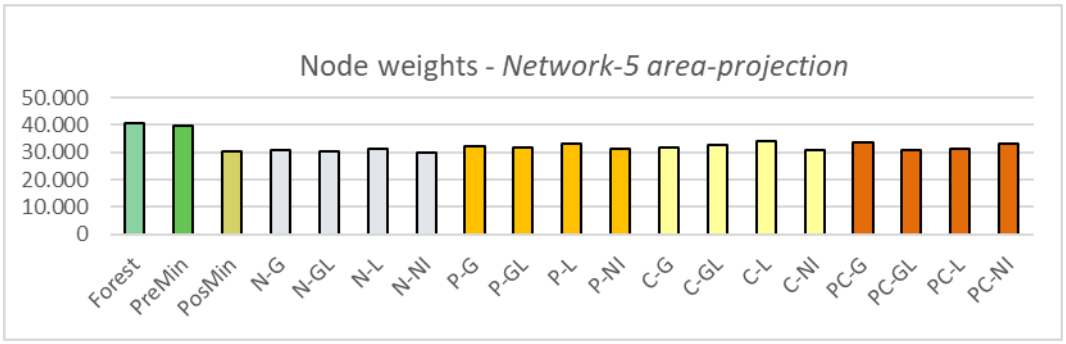
In short, with the exception of *TOC*, all the attributes selected by us in Networks 5, 6 and 7 are among the quality indicators selected by Borges et al. (2019). It is worth remembering that *TOC* and *TN* also had high weights in PCA and were not selected for practical or conceptual reasons.

Figure 27 shows the performance of each area within each of soil attribute class, considering the weights of nodes in each networks 5, 6 and 7. The same was done for the weighs in each *area-projection* (Appendix F). Comparing the first three graphs (a, b and c), it can be seen that the difference between the treatments and the reference areas was smaller in the physical quality aspect. As for the chemical quality (Figure 27b), the difference is notable in relation to *Forest*, as commented before. On the other hand, the result seems coherent in

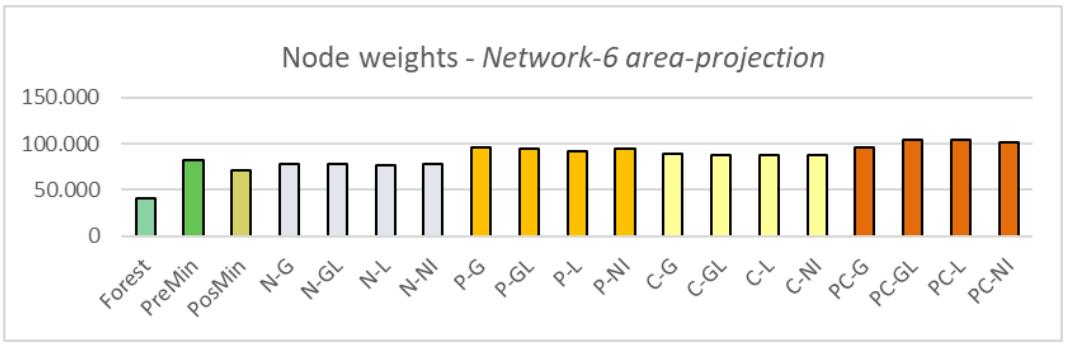
relation to the type and quantity of fertilizers. But, contrary to what was seen in *Network-4,* considering only the chemical aspect, treatments fertilized outperformed *PreMin.*

Figure 27 –Node weights of *Network-5, 6* and *7* compared.

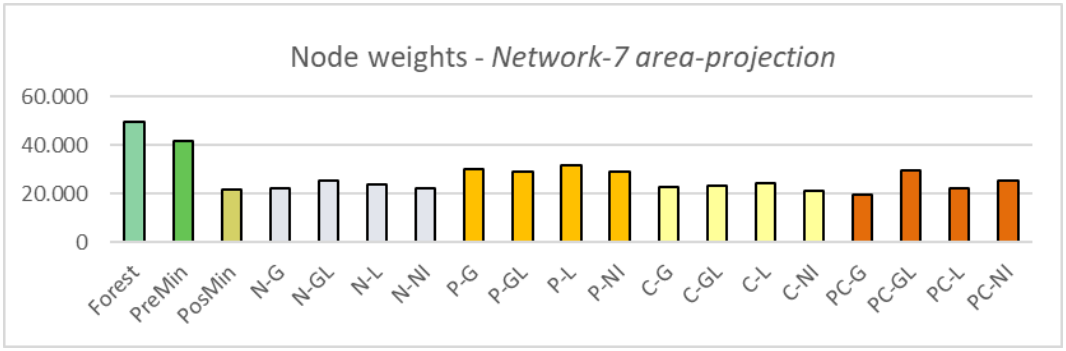
(a)



(b)



(c)



Source: the author.

Note: *Forest:* Atlantic forest; *PreMin*: 10-year-old coffee crop, before mining; *PosMin*: reconstructed soil, six months after mining. *G*: grass intercrop; *L*: leguminous intercrop; *GL*: grass + leguminous mix intercrop; *NI*: no intercrop.

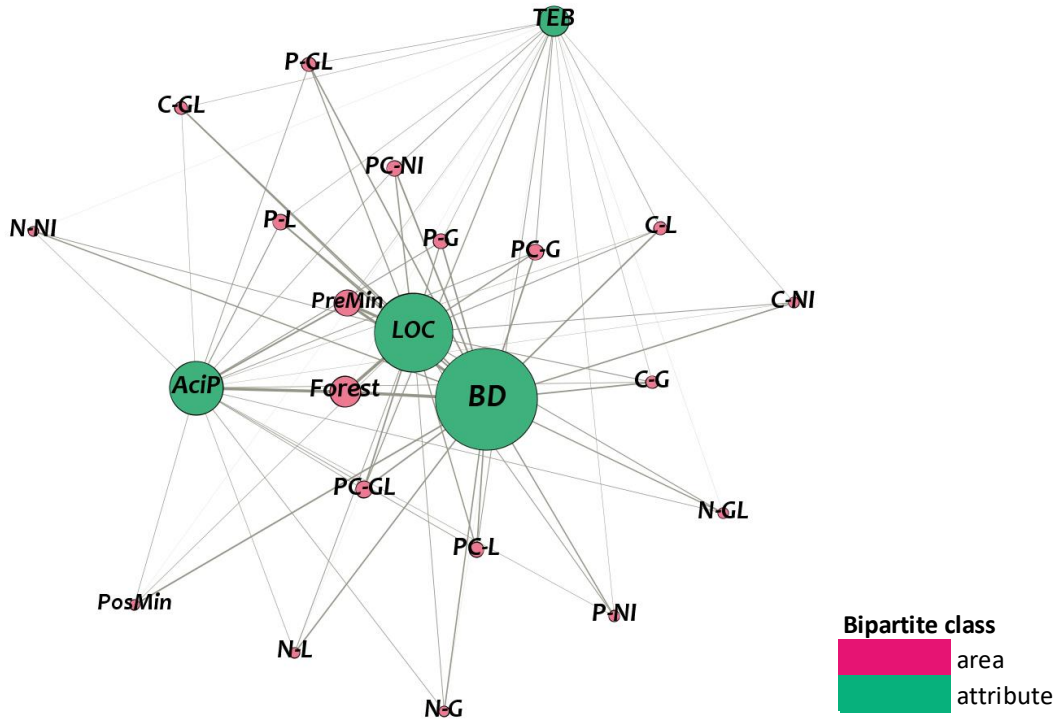
Interestingly enough, the organic-biological quality (Figure 27c) followed the same trend as the results obtained with Network-2p (Figure 28a). This is because Network-2p contains only four attributes: besides *TEB* (which had a much lower weight than the others,

favoring *Forest*), *LOC, AciP* and *BD* are strongly related to the amount of organic matter present in the soil (JOHN *et al.*, 2005) resulting in the great similarity between those bar charts.

4.2.3 Method calibration

Network-2p shows the comparison between all areas under study using the soil quality indicators selected by Borges et al. (2019), namely: *LOC, TEB, BD* and *AciP* (Figure 28).

Figure 28 –*Network-8.*



Source: the author.

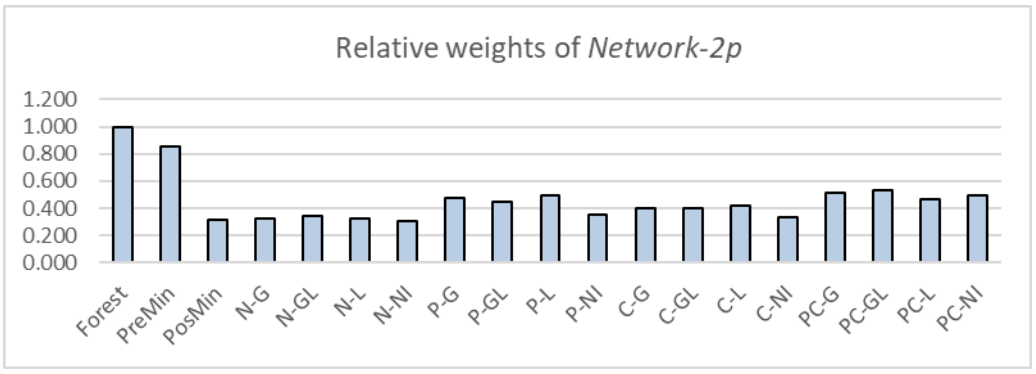
Note: *Forest:* Atlantic forest; *PreMin*: 10-year-old coffee crop, pre-mining; *PosMin*: reconstructed soil, six months after mining; *N-*: no fertilization; *P-*: Poultry litter fertilization; *C-*: chemical fertilization; *PC-*: poultry litter and chemical fertilizations combined; *G*: grass intercrop; *L*: leguminous intercrop; *GL*: grass + leguminous mix intercrop; *NI*: no intercrop. *LOC*: labile organic carbon; *TEB*: total exchangeable basis; *BD*: bulk density; *AciP*: acid phosphomonoesterase enzyme activity.

The network layout confirms the opposition of *TEB* (more related to fertilized areas) to *LOC, AciP* and *BD*, closer to *Forest*. The lower weight of *TEB* (its much smaller size) is due to its low weight in SQI: 0.11.

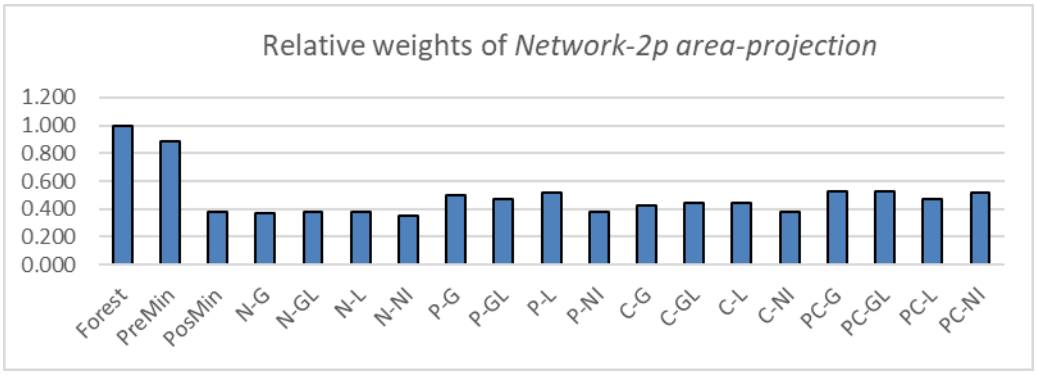
Figure 19 shows the relative weights of *Network-2p* and *Network-2p area-projection*. Node weights in the complete network (Figure 29a) fit SQI better than the weighs in the *area-projection* (Figure 29b), because the SQI calculation results from a summation, in the same way as the calculation of node weights in the complete network.

Figure 29– Relative weights of *Network-2p* and of *Network-2p area projection*.

(a)



(b)



Source: the author.

Note: *Forest:* Atlantic forest; *PreMin*: 10-year-old coffee crop, before mining; *PosMin*: reconstructed soil, six months after mining; *N-*: no fertilization; *P-*: Poultry litter fertilization; *C-*: chemical fertilization; *PC-*: poultry litter and chemical fertilizations combined; *G*: grass intercrop; *L*: leguminous intercrop; *GL*: grass + leguminous mix intercrop; *NI*: no intercrop.

However, some areas showed larger differences from SQI, for example *PreMin* and *P-L* (9% and 7%, respectively). This is because the scores of soil quality indicators calculated by Borges et al. (2019) do not correspond exactly to the weights calculated by us, according to equation (2), for the quality criterion *less is better*, used for BD. Our method proved to be more efficient in this case, as the difference between the weights maintained the same proportions as the differences between the values in the original data (see Appendix G).

In the case of *Network-2p area-projection* (Figure 29b), the differences due to the calculation of the *BD* scores were added to the differences due to the network approach. The projections take into account the particular variations (the proportions between the weights) in each row of the matrix, so that the results do not correspond to a simple sum of weights, as explained above. As some of this results turned out to be more accurate than the results

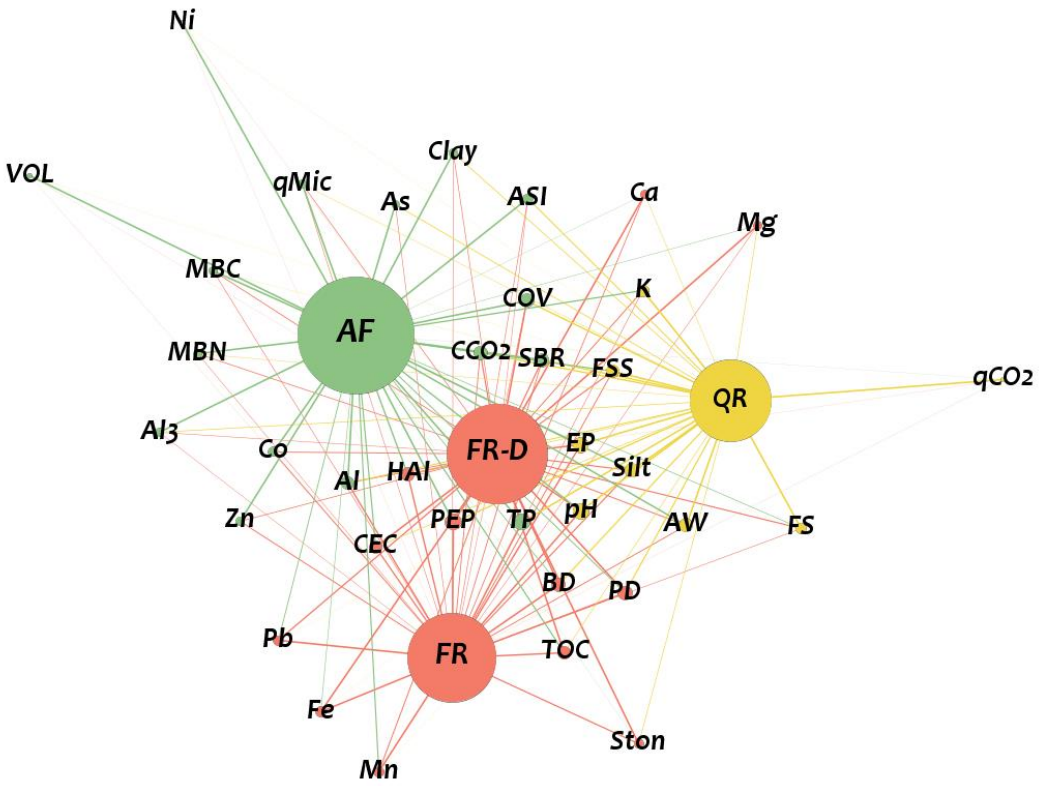
obtained with the complete network (for instance, those shown in Figure 21b), we have chosen to show either result, in case quantitative or non-quantitative comparisons have been made with the results presented by Borges et al (2019).

4.3 Iron mined area – *Group 1* networks

4.3.1 Distinction between the reference areas

Network-8 shows the relations between the four reference areas: *FR, FR-D, QR* and *AF*. Three communities, or modularity classes – [0], [1], and [2], were identified (Figure 30), coinciding with the groupings find by Alves (2019) with the PCA.

Figure 30 – The three communities of *Network-8.*



Modularity class

[0]	*AF, Clay, TP, Al3, ASI, As, Co, Ni, Zn, Al, CCO2, SBR, MBC, MBN, qMic, COV* and *VOL*
[1]	*QR, FS, FSS, Silt, AW, pH, K, EP* and *qCO2*
[2]	*FR, FR-D, Ston, BD, PD, Ca, Mg, HAl, CEC, PEP, TOC, Pb, Mn* and *Fe*

Source: the author.

Note: *FR:* ferruginous rupestrian grassland with thin vegetation; *FR-D:* ferruginous rupestrian grassland with dense vegetation; *QR*: quartzite rupestrian grassland; *AF*: Atlantic Forest.

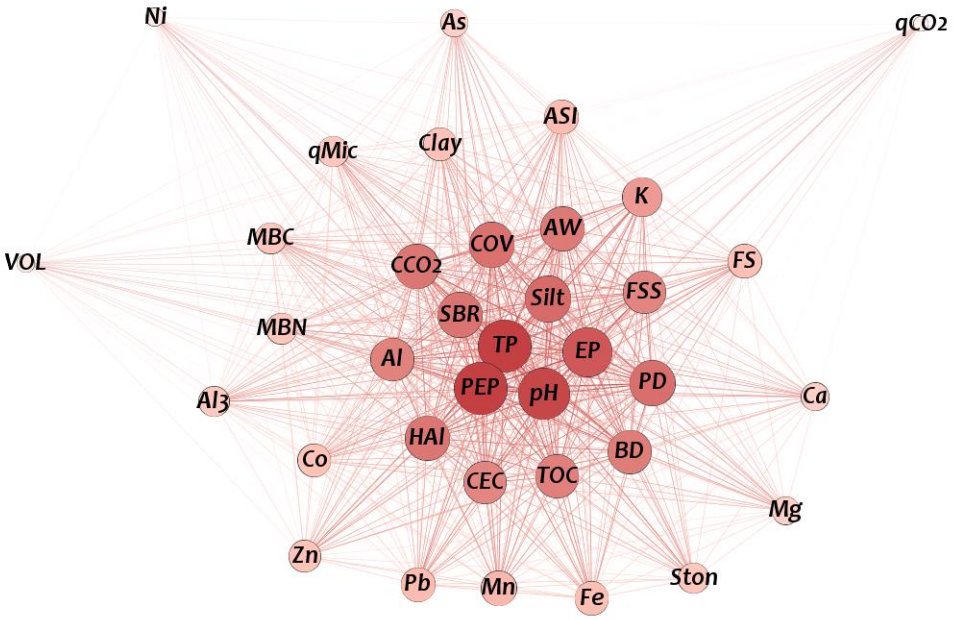
However, it is worth pointing out that the position of nodes in the networks is within a topological space, and not in Cartesian space, as are the dimensions of the PCA. This means that the positions of the nodes are only relative, not absolute.

According to that author, there was no significant difference between *FR* and *FR-D*, a result that fits with the fact that they belong to the same community. There was also some correspondence between the attributes clustered in each community and the vectors closest to each area in the PCA, including their relative positions: the attributes selected in the PCA are also in opposite positions in *Network-1* (e.g. *Zn* and *Pb* opposite *FS* and *FSS*; *BD* and *PD* opposite *VOL* and *Ni*).

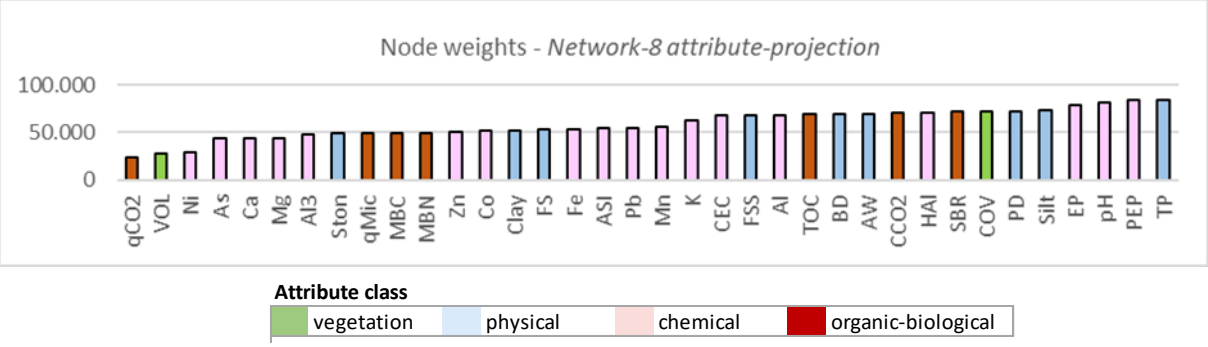
Figure 31 shows the results for *attribute-projection* of *Network-8*.

Figure 31 –*Network-8* attribute-projection.

(a)



(b)

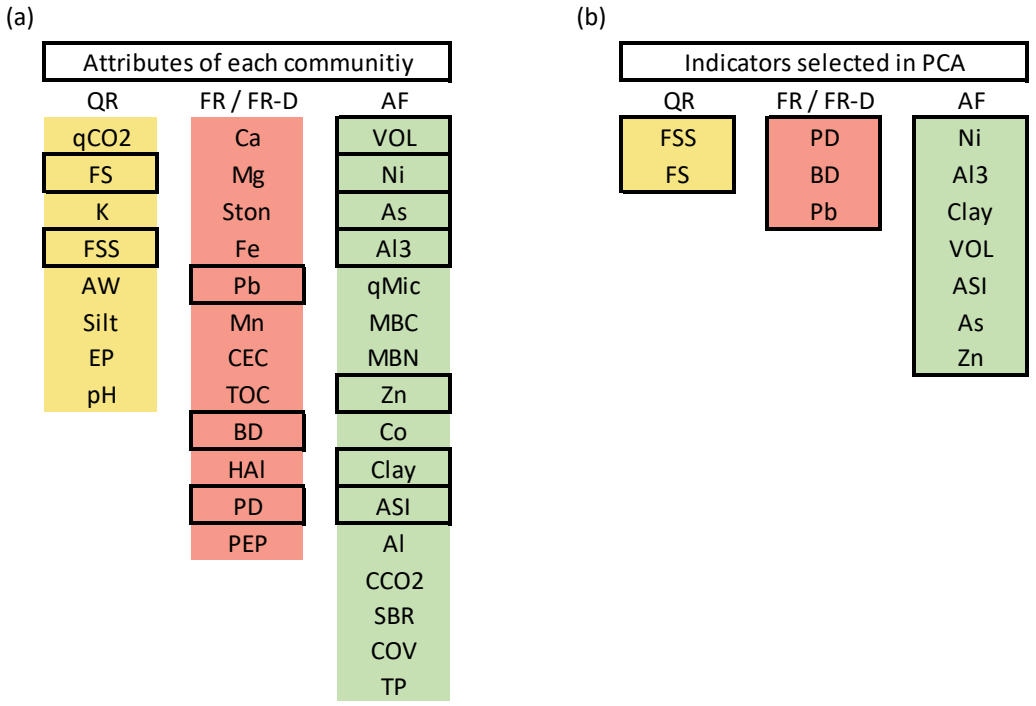


Source: the author.

Note: (a) the *attribute-projection* of *Network-8*; (b) node weights of *Network-8 attribute-projection* in ascending order, colored by attribute classes.

The most important attributes, the first two groups in the weighted degree distribution (Appendix H), correspond, approximately, to the two outermost node layers of the network projection. Comparing this result with the attributes selected with the PCA (Figure 32), we note that, despite the similarities, there are two important differences: (i) no organic-biological attributes were considered in the selection with PCA, while in the ranking of *attribute-projection* weights (Figure 31b) four of them are among the first; (ii) the importance of the attributes in distinguishing the areas is also not the same.

Figure 32 –Attributes belonging to each community in *Network-1* and the selected attributes in the PCA, in order of importance, in each column.



Source: the author.

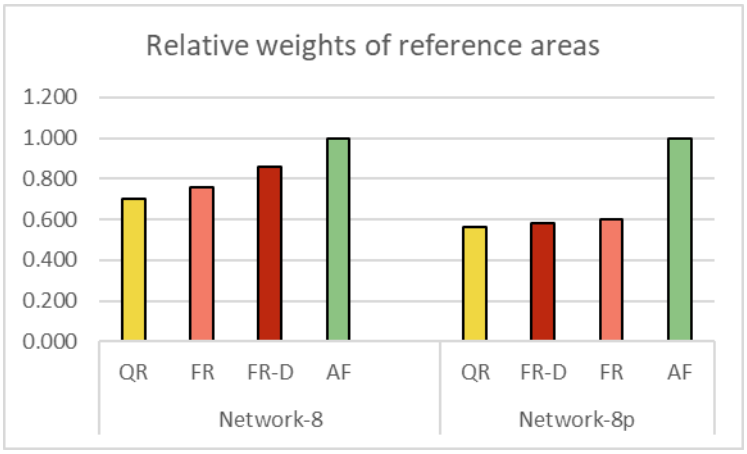
Note: (a) attributes belonging to each community in *Network-8,* in ascending order of their weights in each column; in emphasis, the attributes selected in the PCA; (b) attributes selected in the PCA by Alves (2019), in descending order of their weights in each column.

These differences are possibly due to the fact that PCA is based on the structure of correlations between attributes, which was not the case in our method, where the variables have not received any prior treatment. Furthermore, the attributes selected with PCA cover almost the entire ranking of node weights of *Netwok-8 attribute-projection* (Figure 31b), which clearly indicates that the selection criteria are not the same. However, the attributes selected with PCA for each group all belong to their respective community in the *Network-8.*

This means that the two methods are tuned with regard to cluster distinction, but the criteria for selecting indicators need to be further studied.

It is worth noting that, when comparing the ranking of weights of *Network-8* with those of *Network-8p* (constructed only with the PCA indicators), we note that the weights of *Network-8p* did not maintain the same trend as the weights of *Network-8*, since *FR* and *FR-D* reversed their positions, also presenting much lower weights than *AF* (Figure 33). In our view, the function of indicators should be to represent the original dataset while retaining the relationship structure of the data, but this did not happen in *Network-8p*, possibly, because the edge weights were not adjusted according to the weights of each indicator in the PCA.

Figure 33 – Relative weight of areas in *Network-8* and *Network-8p.*



Source: the author.

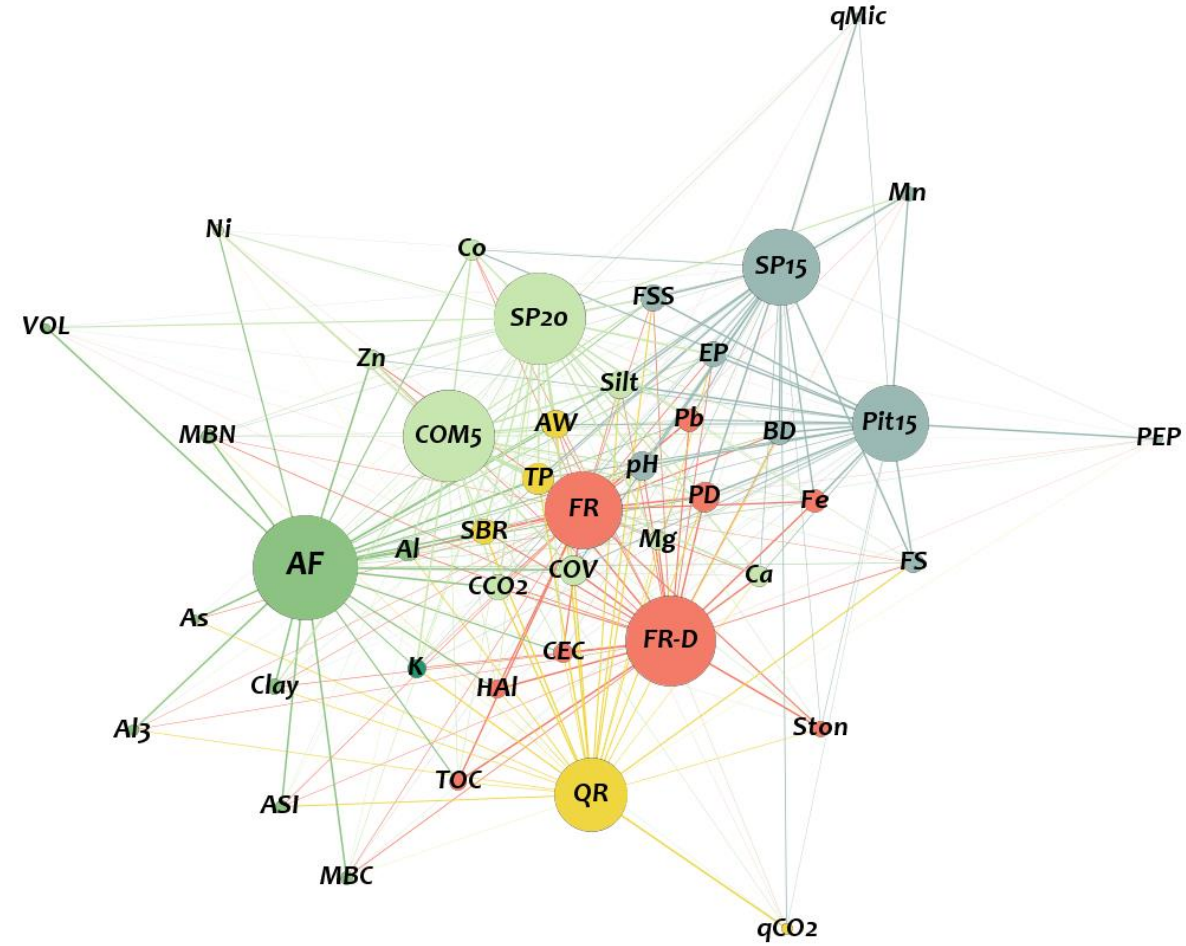
Note; *FR:* ferruginous rupestrian grassland with thin vegetation; *FR-D:* ferruginous rupestrian grassland with dense vegetation; *QR*: quartzite rupestrian grassland; *AF*: Atlantic Forest.

However, the problem may lie in the criteria used for choosing the indicators. For example: (i) qCO2 was the attribute with the lowest weight among all (Figure 31b), and also the most relevant in the discrimination of QR (Figure 32a) and yet it was not contemplated in the PCA selection, despite this attribute being an important indicator of stress and ecosystem disturbances (PINTO; MARIA; CASTILHOS, 2008) (ii) for the composition of the RQI, Alves (2019) chose to kept at least one of the vegetation parameters, even when they were excluded through the criteria she adopted for the selection of indicators in the PCA. And this is an important issue that deserves further study, so that one can improve the selection criteria or even investigate the suitability of PCA *alone* for this type of decision.

4.3.2 Distinction between reference areas and recovering ones

Network-9 shows the relations between all areas under study of Dat-2 and its five communities are described in Figure 34.

Figure 34 – The five communities of *Network-9*.



Modularity class

[0]	*FR, FR-D, Ston, PD, HAl, CEC, TOC,* and *Fe*
[1]	*QR, TP, AW, SBR* and *qCO2*
[2]	*AF, clay, Al3, ASI, As, Al, MBC, MBN,* and *VOL*
[3]	*SP-15, Pit-15, FS, FSS, BD, pH, EP, PEP, Mn* and *qMic*
[4]	*COM5, SP-20, Silt, K, Ca, Mg, Co, Ni, Zn, CCO2 and COV*

Source: the author.

Note: *FR:* ferruginous rupestrian grassland with thin vegetation; *FR-D:* ferruginous rupestrian grassland with dense vegetation; *QR*: quartzite rupestrian grassland; *AF*: Atlantic Forest; *COM5*: compensation area, 5 years under reclamation; *SP15*: sterile pile, 15 years under reclamation; *Pit15*: pit, 15 years under reclamation; *SP20:* sterile pile, 20 years under reclamation.

When two (or more) areas belong to the same community, this means that they are more similar to each other than to the others areas. This happens even when there is a

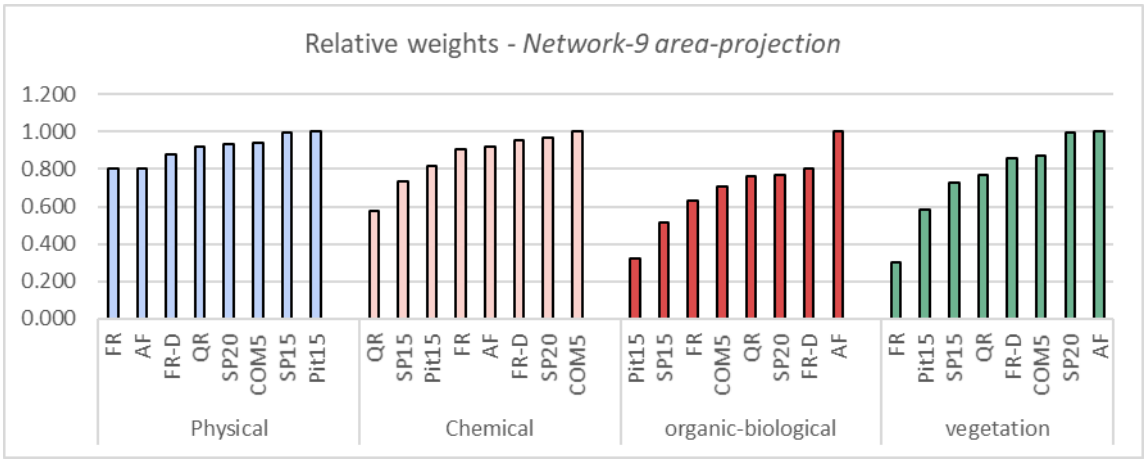
statiscally significant difference between the areas, as in the case of *SP-15* and *Pit-15* which nevertheless appeared grouped together in all comparisons with the references (see item 4.4.1, below). The same result was obtained by Alves (2019) with the PCA.

The structure of *Network-9* also coincides with other information about these areas: *SP-15* and *Pit-15* are adjacent areas, the former receiving the sterile waste from the latter; *COM5* is also adjacent to *AF*, and both belong to the same geological group as *SP-20*; there was no significant difference between *FR* and *FR-D*; *QR* is geologically different from all the other areas and the only one with grass vegetation.

It is known that the great geological diversity of the Quadrilátero Ferrífero province is one of the causes of the chemical, physical and biological heterogeneity that determines the floristic and environmental differences of these areas (BENITES *et al.*, 2003; PEREIRA, 2010). This is because the soils of this region are poorly developed and still reflects the characteristics of source rocks. Moreover, we notice in the Figure 34 the opposition and greater distance between *AF* and the community of *SP-15* and *Pit-15*, due to the low similarity between these areas and *AF*, a result also found by Alves (2019) in RQI.

The organic-biological attribute class was the most important for distinguishing the areas, showing the greatest differences between the weights of each area (Figure 35).

Figure 35 – Relative weights of *Network-9 area-projection,* according to attribute classes.



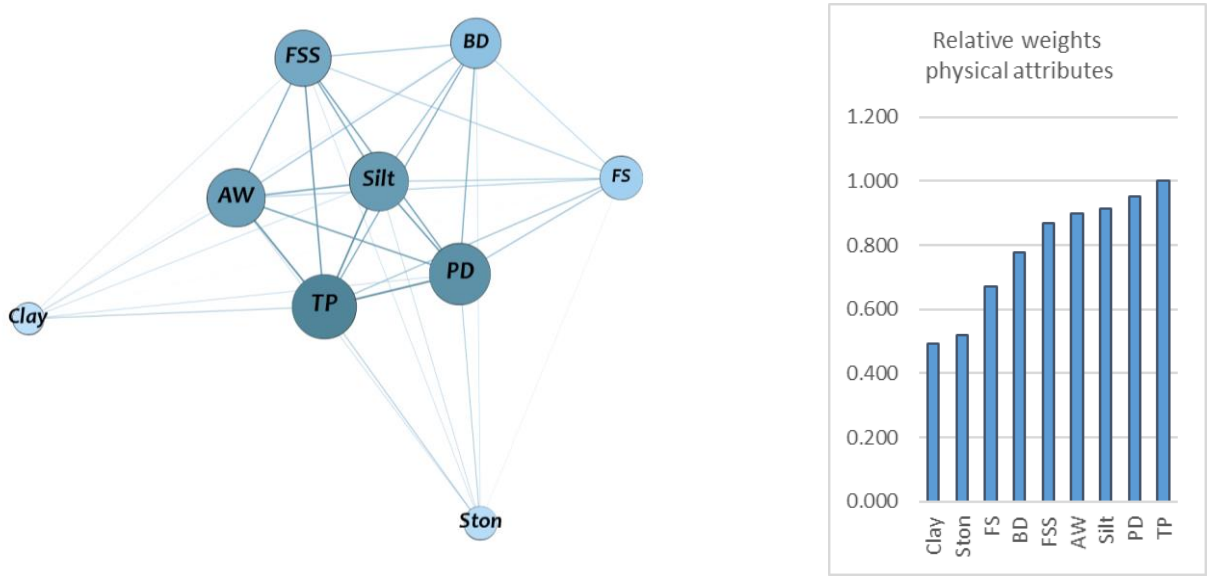
Source: the author.

Note: *FR:* ferruginous rupestrian grassland with thin vegetation; *FR-D:* ferruginous rupestrian grassland with dense vegetation; *QR*: quartzite rupestrian grassland; *AF*: Atlantic Forest; *COM5*: compensation area, 5 years under reclamation; *SP15*: sterile pile, 15 years under reclamation; *Pit15*: pit, 15 years under reclamation; *SP20:* sterile pile, 20 years under reclamation.

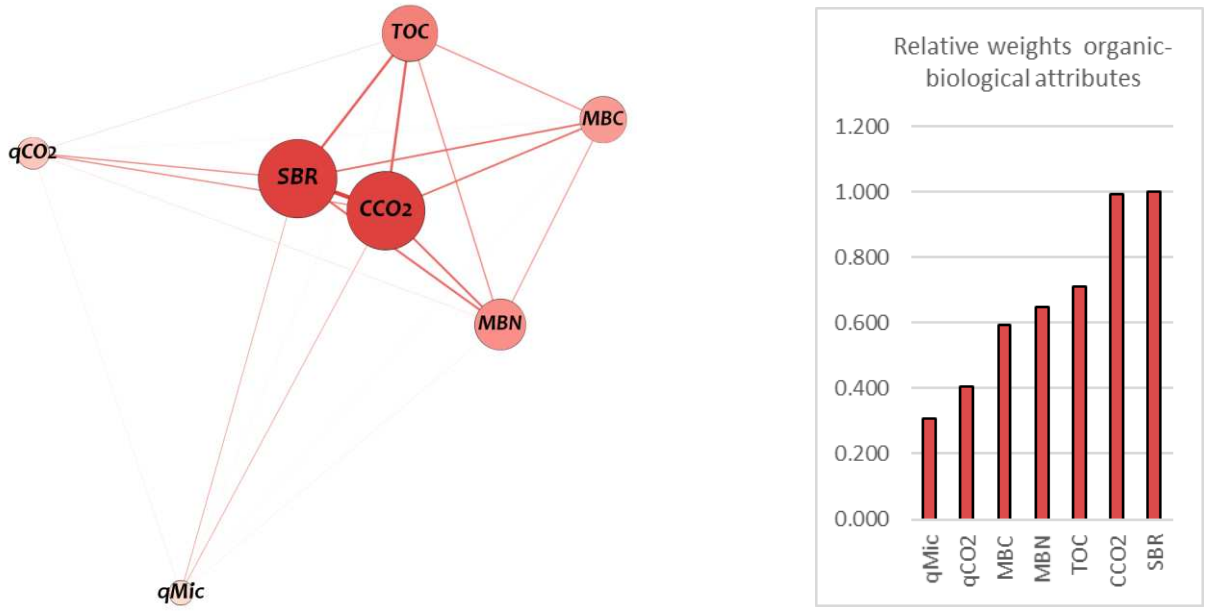
The physical attribute class contributed the least to this distinction, despite the differences in weight among the attributes (Figure 36a). However, as seen in Figure 36b, the difference between the organic-biological attributes is greater, evidenced in both the value of the weights and the layout of the projection.

Figure 36 – Projections of the physical and organic-biological attributes of *Network-9*.

(a)



(b)



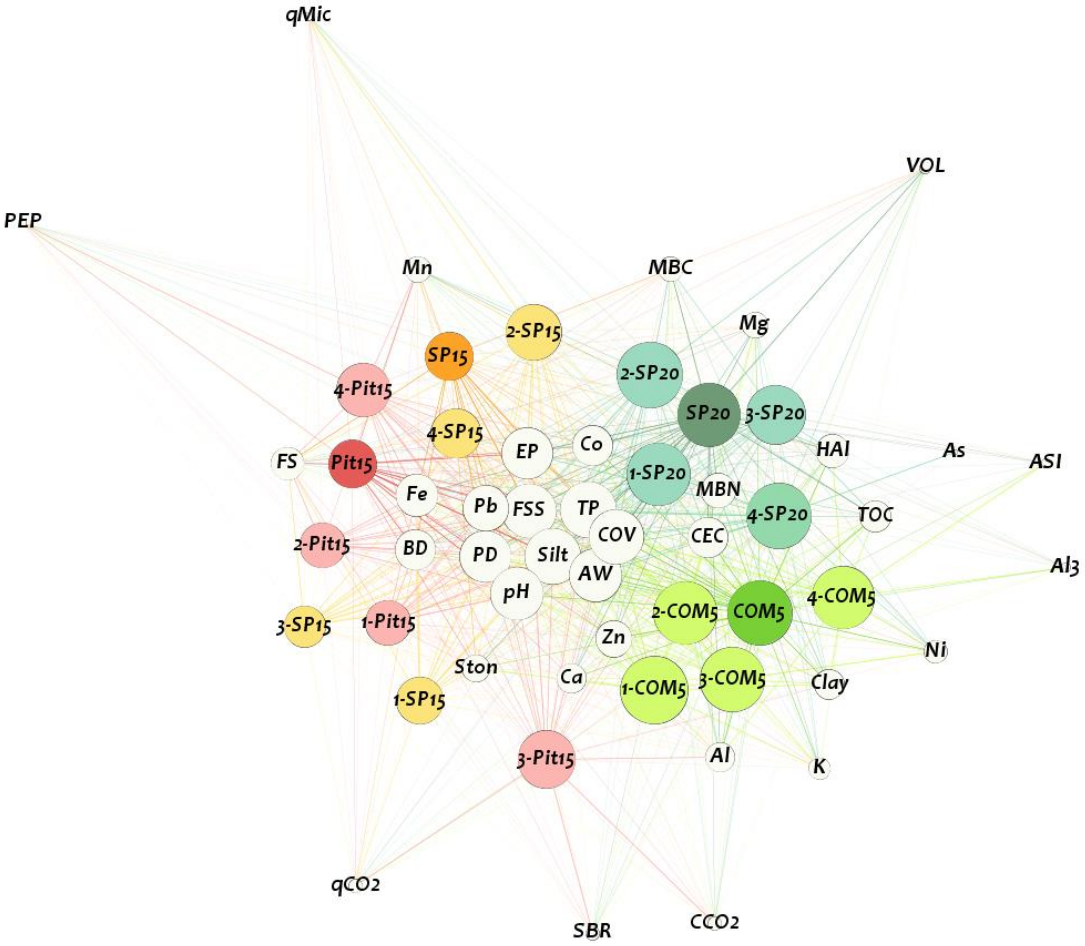
Source: the author.

4.3.3 Dispersion of data in recovering areas

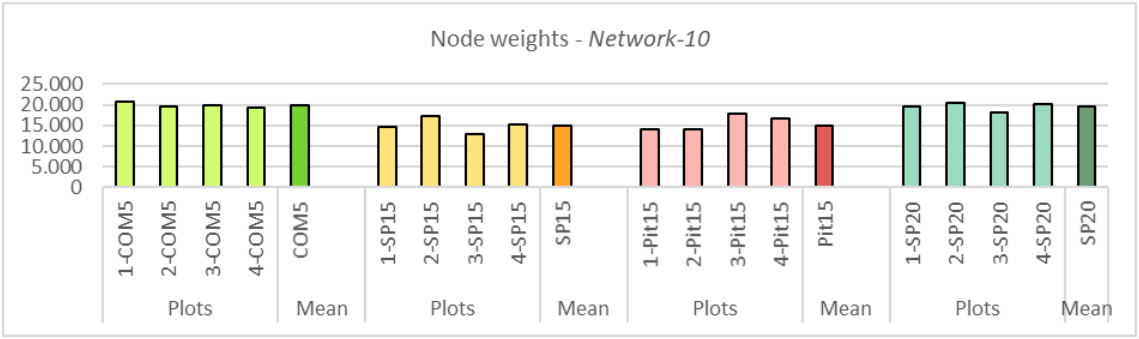
Network-10 shows the relations between all plots of recovering areas – named as 1-, 2-, 3- and 4-, and their respective averages: *COM5, SP-15, Pit-15* and *SP-20* (Figure 37).

Figure 37 – Highlighted *area* class nodes of *Network-10,* colored according to each particular area.

(a)



(b)



Source: the author.

Note: (a) plots of the same area have the same color, averages with darker color; (b) node weights of each plot, calculated by *Gephi*. *COM5*: compensation area, 5 years under reclamation; *SP15*: sterile pile, 15 years under reclamation; *Pit15*: pit, 15 years under reclamation; *SP20:* sterile pile, 20 years under reclamation.

The plots in areas *COM5* and *SP-20* grouped well together and with their averages, unlike plots *SP-15* and *Pit-15* (Figure 37a). This visual result is consistent with the configuration of the area weights: the weights of *COM5* and *SP-20* are higher and more homogeneous than the weights of *SP-15* and *Pit-15* (Figure 37b).

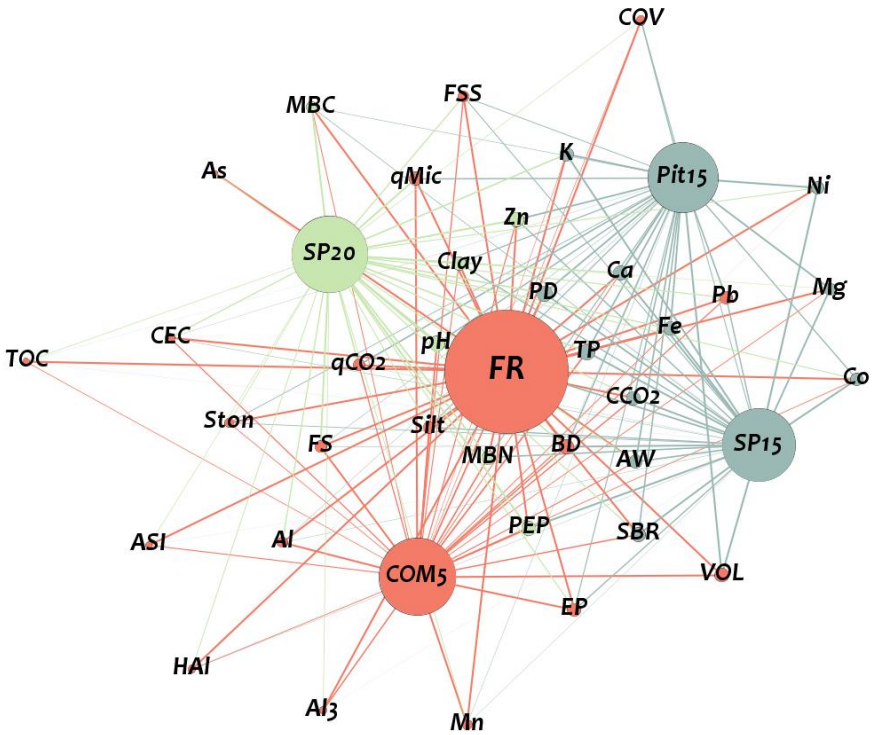
Plot *3-Pit-15* stands out from the other plots because of the discrepant values of *CCO2, SBR* and *qCO2*. So much so that it was even considered to have its own community (see Appendix I). But despite all the heterogeneity of the plots in these two area, the averages of *SP-15* and *Pit-15* belong to the same community, with weights very close to each other (14.952 and 14.879, respectively). And this is why these two areas appear grouped together in all the other networks below.

4.4 Iron mined area – *Group 2* networks

4.4.1 Differences between the four groups of reference areas

The first group of reference areas is that of *FR* (Figure 38).

Figure 38 – The three communities of *Network-11* (*FR*).



Modularity class

[0]	*FR, COM5, Ston, FS, FSS, Silt, BD, Al3, Hal, CEC, ASI, EP, TOC, As, Pb, Mn, Al, qMic, qCO2, COV, VOL.*
[1]	*SP15, Pit15, PD, TP, AW, K, Ca, Mg, Co, Ni, Fe, CCO2, SBR.*
[2]	*SP20, Clay, pH, PEP, Zn, MBC, MBN.*

Source: the author.

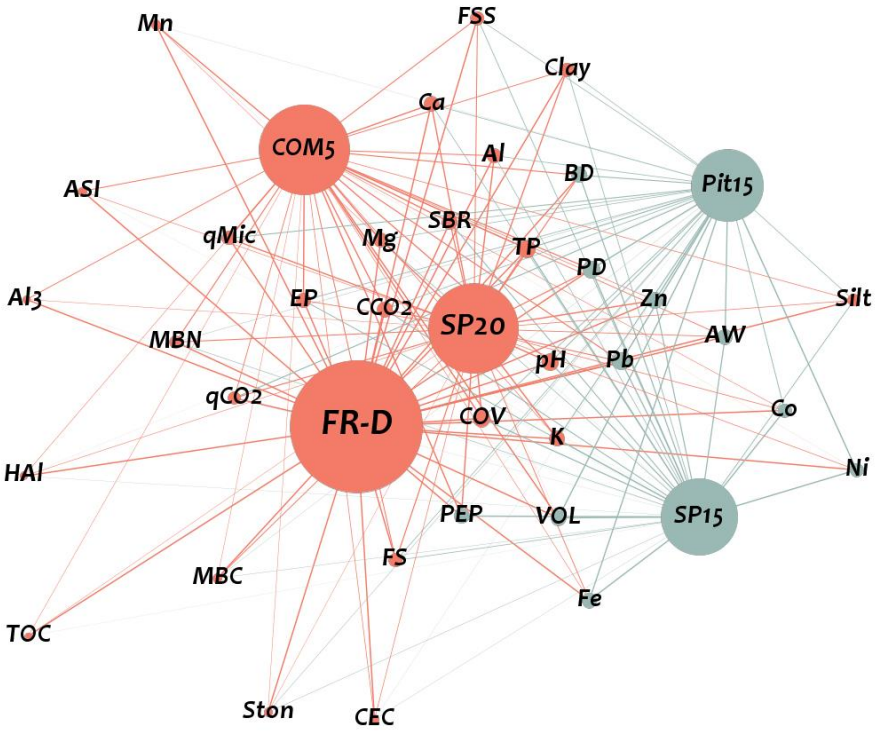
Note: *FR*: ferruginous rupestrian grassland with thin vegetation; *COM5*: compensation area, 5 years under reclamation; *SP15*: sterile pile, 15 years under reclamation; *Pit15*: pit, 15 years under reclamation; *SP20:* sterile pile, 20 years under reclamation.

The most important attributes to distinguish *FR* from the areas in recovery were: *As, TOC, ASI, AL3* and *HAl.* In fact, three of the areas in recovery do not even have *As*. The reference area FR also has much higher *TOC, ASI, Al3* and *HAl* values than the other areas.

The area most similar to *FR* was *COM5*, according to the modularity class, despite the small difference. In each community, the areas share among themselves attributes for which they have the highest weights, or for which the weights are most similar to each other. The same is true for Networks 12, 13 and 14 below.

The attributes that best differentiated *FR-D* from areas in recovery were: *TOC, HAl, Al3, Ston, ASI, CEC, MBC* and *Mn* (Figure 39). *FR* has deeper soil and denser vegetation than *FR*. This made the layout of this network slightly different from the previous one. *COM5* and *SP20* turned out to be more similar to it than the other two, more degraded areas, which have denser soils and less vegetation.

Figure 39 – The two communities of *Network-12* (*FR-D*).



Modularity class

[0] *FR-D, COM5, SP20, Ston, FS, FSS, Silt, Clay, TP, pH, K, Ca, Mg, Al3, Hal, CEC, ASI, EP, TOC, Mn, Al, CCO2, SBR, MBC, MBN, qMic, qCO2, COV.*

[1] *SP15, Pit15, BD, PD, AW, PEP, Co, PB, Ni, Zn, Fe, VOL.*

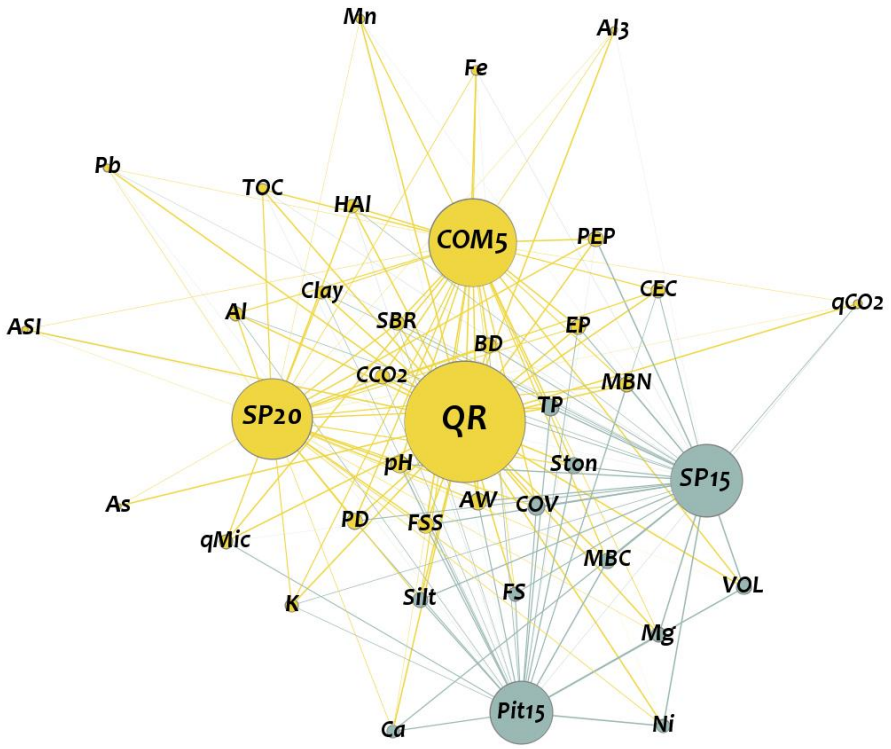
Source: the author.

Note: *FR-D:* ferruginous rupestrian grassland with dense vegetation; *COM5*: compensation area, 5 years under reclamation; *SP15*: sterile pile, 15 years under reclamation; *Pit15*: pit, 15 years under reclamation; *SP20:* sterile pile, 20 years under reclamation.

Regarding *QR*, it can be seen that there was more distinction between the areas, especially *COM5* and *Pit15*: one more and the other less similar than the reference (Figure 40). The two areas showed opposite weights for most of the attributes that distinguishing *QR*: *As, ASI, Al3, Pb, qCO2, Mn* and *Fe.*

However, the two adjacent areas, *SP15* and *Pit15*, still belong to the same community because they are actually more similar to each other than the other areas. *SP20* was less similar to *QR* than *COM5* mainly due to the difference in vegetation volume (*VOL*) and the lower weights of some chemical attributes such as *Al3, Mn* and *Fe*.

Figure 40 – The two communities of *Network-13* (*QR*).



Modularity class

[0]	*QR, COM5, SP20, FSS, Clay, BD, PD, AW, pH, K, Al3, HAl, CEC, ASI, EP, PEP, TOC, As, Pb, Mn, Fe, Al, CCO2, SBR, MBN, qMic, qCO2.*
[1]	*SP15, Pit15, Ston, FS, Silt, TP, Ca, Mg, Ni, MBC, COV, VOL.*

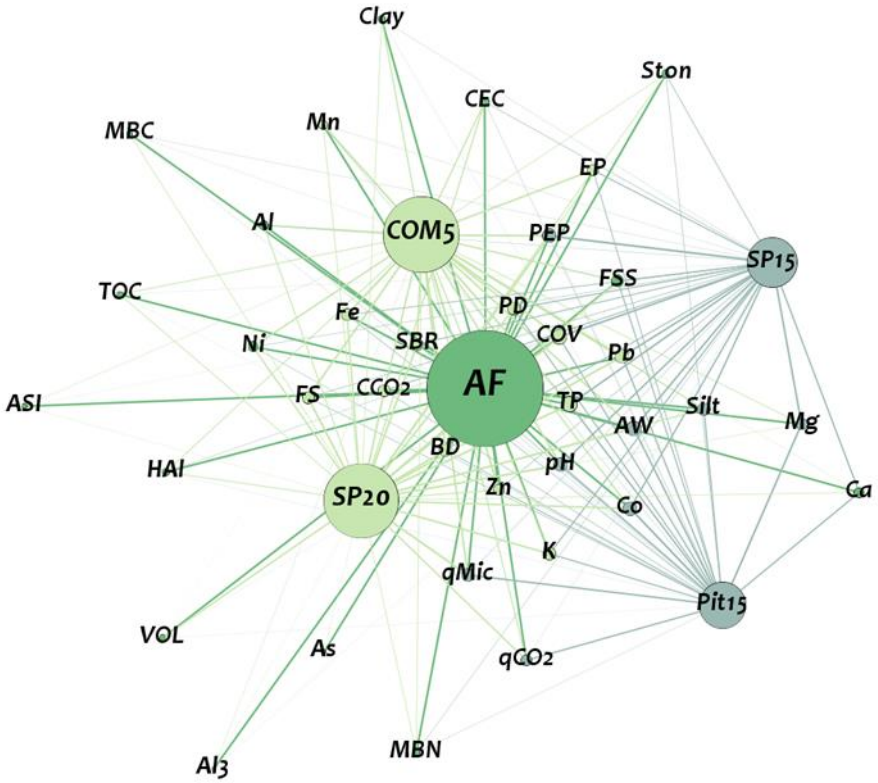
Source: the author.

Note: *QR*: quartzite rupestrian grassland; *COM5*: compensation area, 5 years under reclamation; *SP15*: sterile pile, 15 years under reclamation; *Pit15*: pit, 15 years under reclamation; *SP20:* sterile pile, 20 years under reclamation.

The forth group of referential areas showed even greater differences than the previous one, just because *AF* has very different characteristics from *SP15* and *Pit15* (Figure 41). The most relevant attributes to distinguishing *AF* were: *As, Al3, ASI, MBC, VOL, MBN, TOC, HAl, Clay* and *Ston*.

It is interesting to note that *AF* has important organic-biological and vegetation attributes which differentiate it from the others, besides the amount of clay, which is also higher. However, *Ston* is also a differential, because AF has it in smaller quantities than the other areas. Here we see a subtle difference in the interpretation of the results: when an area is set as a reference, the most relevant attributes are both the ones with the highest and lowest values in the original data. But they are certainly the most different, and the direction and thickness of the edges tell us which areas these differences relate the most.

Figure 41 – The three communities of *Network-14* (*AF*).



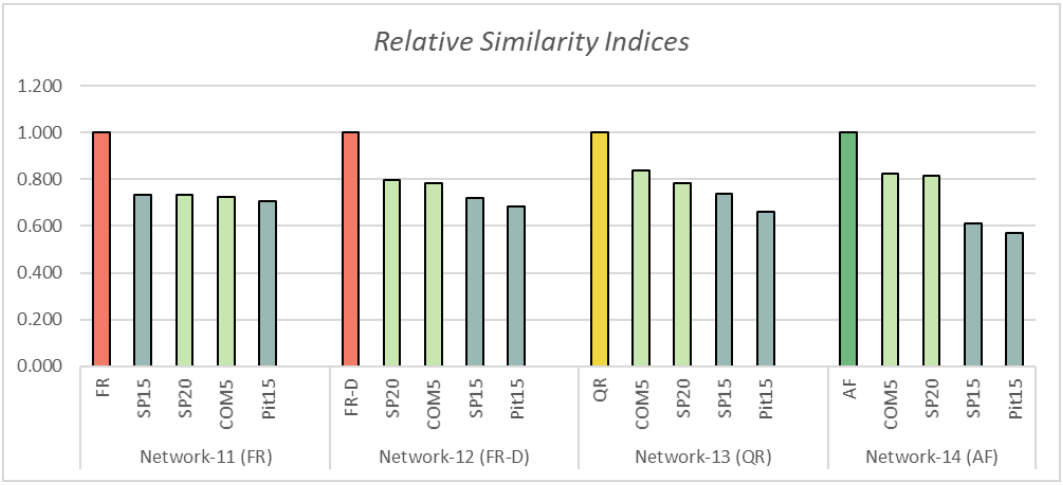
Modularity class	
[0]	*AF, Ston, FSS, Clay, Ca, Al3, HAl, CEC, ASI, TOC, As, Ni, Al, MBC, MBN, VOL.*
[1]	*COM5, SP20, FS, BD, PD, TP, K, EP, Pb, Zn, Mn, Fe, CCO2, SBR, COV.*
[2]	*SP15, Pit,15, Silt, AW, pH, Mg, PEP.*

Source: the author.

Note: *AF*: Atlantic Forest; *COM5*: compensation area, 5 years under reclamation; *SP15*: sterile pile, 15 years under reclamation; *Pit15*: pit, 15 years under reclamation; *SP20:* sterile pile, 20 years under reclamation.

As commented earlier, areas *SP-15* and *Pit-15* appear clustered in all networks (Figures 38 to 41). However, as can be seen in Figure 42, the ranking of the weights does not always reproduce the results of modularity, especially when the differences are very small (as in *Network-11*). This is because the division into communities involves more complex calculations (FORTUNATO, 2010; KHAN; NIAZI, 2017) than simply adding up the weight of the edges, performed to obtain the weight of the nodes, as shown in Figure 16. That is, it is possible to recognize patterns of similarity (in the form of communities) embedded in the relationships between areas, even when the numerical differences are very small.

Figure 42 – The *Relative Similarity Indices* (*RSI*) of *Networks 11, 12, 13* and *14.*



Source: the author.

Note: *FR:* ferruginous rupestrian grassland with thin vegetation; *FR-D:* ferruginous rupestrian grassland with dense vegetation; *QR*: quartzite rupestrian grassland; *AF*: Atlantic Forest; *COM5*: compensation area, 5 years under reclamation; *SP15*: sterile pile, 15 years under reclamation; *Pit15*: pit, 15 years under reclamation; *SP20:* sterile pile, 20 years under reclamation.

However, the *RSI* allows us to compare the pattern of similarity of the networks: for example, the most striking differences appear in *Network-14*, since *SP-15* and *Pit-15* are areas drastically affected by mining, difficult to recover, constituting very different environments from the native forest (*AF*). The smallest differences are found in *Network-11*. A simple visual inspection of these networks also denotes this. Therefore, there is always a close relationship between the similarity indices and the relationship structure of the networks.

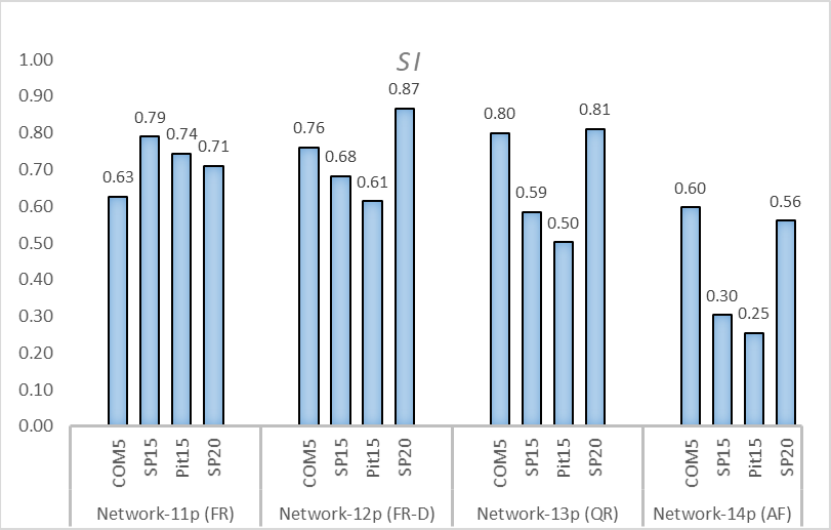
4.5 Iron mined area – *Group 3* networks

4.5.1 Method calibration

The attributes selected by Alves (2019) with the PCA for each reference area were: for *FR* group*: BD, Fe, TP, PD, Clay, COV, Silt* and *FSS;* for *FR-D* group*: TP, FS, BD, Al3, Clay, COV* and *Silt;* for *QR* group*: BD, TOC, CEC, Al3, TP, HAl, Clay, PD, AW, Fe, FS, FSS, K, COV, Zn, Co, Silt* and *Ston;* for *AF* group*: HAl, TOC, CEC, Clay, FSS, Al, PD, ASI, Al3, Fe, BD, MBN, MBC, pH, FS, EP, Ni, VOL, Ca* and *Mg*. As can be seen, there is almost no correspondence between them and the attributes considered to be the most relevant for distinguishing the areas in Networks 11 to 14 (see item 4.4.1). This is because we only highlighted the attributes that showed the greatest differences between the areas. We did not selected *indicators*, because this is a work that involves conceptual and practical decisions which may considered in future studies.

However, the comparison between the *Similarity Indices* (*SI*) and the Recovery Quality Indices (RQI) for the reference areas showed a great similarity in the trends of each group with the exception of the first group, referring to Network-11p (Figure 43).

Figure 43 – Similarity indices (*SI*) of *Networks 11p, 12p, 13p* and *14p.*



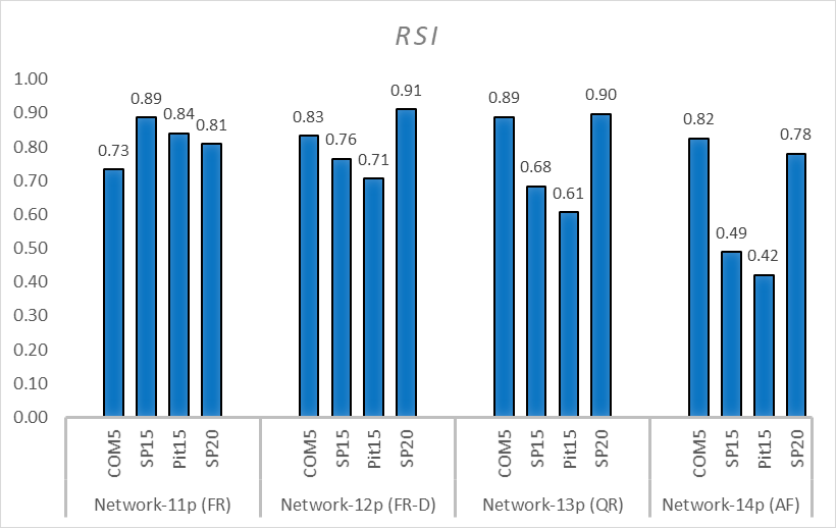
Source: the author.

Note: *FR:* ferruginous rupestrian grassland with thin vegetation; *FR-D:* ferruginous rupestrian grassland with dense vegetation; *QR*: quartzite rupestrian grassland; *AF*: Atlantic Forest; *COM5*: compensation area, 5 years under reclamation; *SP15*: sterile pile, 15 years under reclamation; *Pit15*: pit, 15 years under reclamation; *SP20:* sterile pile, 20 years under reclamation.

The numerical differences between our results and those found by Alves (2019) are due to the way of calculating the indicator scores. In the case of *Network-11p*, the difference between *Clay* and *COV* scores for Pit-15 in the two methods was so large that it caused a reversal in trends of *Pit-15* and *SP-20* (see Appendix J).

The results obtained with the area-projections were also similar with regard to the trends of each group, but not numerically, since the weighs are always higher (Figure 44).

Figure 44 – Relative Similarity indices (*SI*) of *Networks 11p, 12p, 13p* and *14p area-projection.*



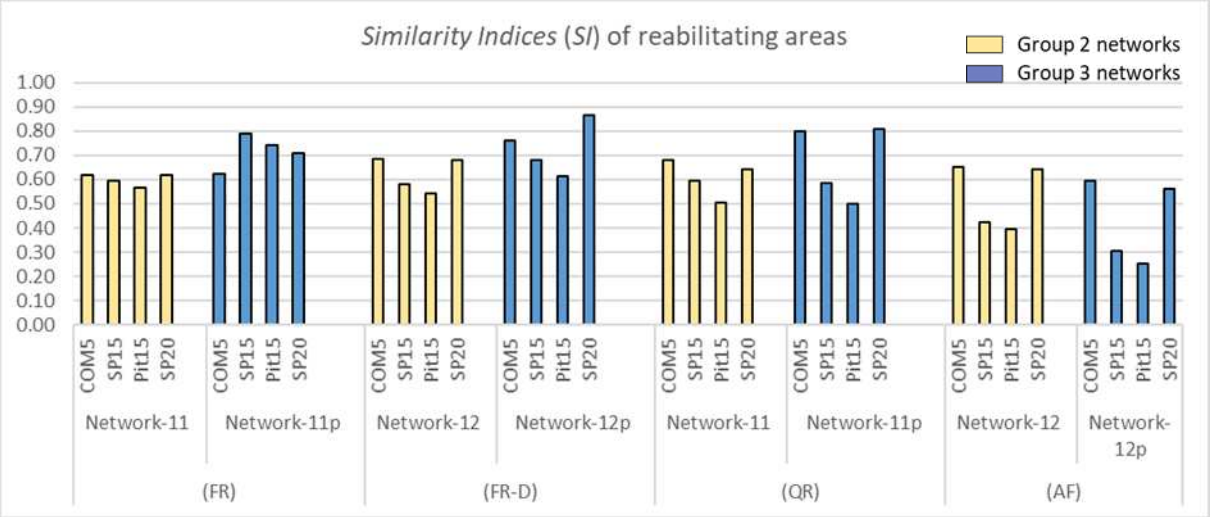
Source: the author.

Note: *FR:* ferruginous rupestrian grassland with thin vegetation; *FR-D:* ferruginous rupestrian grassland with dense vegetation; *QR*: quartzite rupestrian grassland; *AF*: Atlantic Forest; *COM5*: compensation area, 5 years under reclamation; *SP15*: sterile pile, 15 years under reclamation; *Pit15*: pit, 15 years under reclamation; *SP20:* sterile pile, 20 years under reclamation.

For this reason, we conclude that *RSI* is better suited for qualitative exploration of data, while *SI* can be used for quantitative comparisons.

Figure 45 shows the impact of the choice of indicators on the trend for each group of areas. The only group that kept the same trend was the AF group (Networks 12 and 12p), possibly because the number of indicators selected was larger (20 attributes), with greater chances of representing the original data (36 attributes). Conversely, networks 11 and 11p seem to go in opposite directions. This possibly means that the selection may not have been successful and needs further investigated. Future research ca test this hypothesis, pointing the way to a reliable method of checking the suitability of an indicators selection.

Figure 45 – Comparison between networks of Group 2 and Group 3.



Source: the author.

Note: *FR:* ferruginous rupestrian grassland with thin vegetation; *FR-D:* ferruginous rupestrian grassland with dense vegetation; *QR*: quartzite rupestrian grassland; *AF*: Atlantic Forest; *COM5*: compensation area, 5 years under reclamation; *SP15*: sterile pile, 15 years under reclamation; *Pit15*: pit, 15 years under reclamation; *SP20:* sterile pile, 20 years under reclamation.

4.6 Limitations of the method and future research

1. When an area has zero value for some attribute in the original data, there will be no link between this area and that attribute. However, when this attribute was selected in the PCA, we had to include it in the comparison, so that our results could be compared to those of other authors. Thus, the value of this attribute (equal to zero in the original data) was approximated to 0.001, so that this attribute could be included in the network. This is a good approximation, but does not represent reality, if this attribute not really exist in the area in question.

2. More accurate results could be obtained if adjustments were made to equation (1), according to the scale amplitude of each attribute or the presence of nonlinearity in the attribute functions. For example, in the case of pH, which has a logarithmic scale, a difference of 1 point on the scale means a difference of 10 points in practice.

3. The use of variable standardization, taking into account its standard deviation, and mean comparison tests would certainly give more precision to the results and better support for its interpretation.

4. We believe that it is possible to select quality indicators using Network Theory, but this requires further study, both conceptually and practically.

5. Network Science has been expanding very rapidly and the possibilities for future research are very broad. Within mined reclamation it also be possible to use this networks, in addition to dynamical systems theory, to better understand the dynamics of mined soils and make predictions about the state of the reclamation areas, aiding in their monitoring.

5 CONCLUSIONS

Network Theory allows both visual and quantitative comparisons between independent variables (areas) and dependent variables (soil attributes/vegetation parameters). In this way, it is possible to group areas that are most similar to each other, compare areas under rehabilitation with their references, and identify the most relevant attributes for distinguish the areas. It is also possible to visualize the dispersion of the data (area plots) in relation to their averages and identify the variables (attributes) responsible to this.

The above results can be obtained by using weighted bipartite networks, having as links (edges) the original data values, transformed into weights from 0 to 1. The method developed in this work to calculate the edge weights proved to be very accurate, because the weights maintained the same proportion as in the original data.

The use of weighted bipartite networks to discriminate the areas and grouping them into communities, according to their attributes, allowed consistent results to be obtained, comparable to those of PCA.

Two different numerical results were obtained for the nodes in each network: in the full graph and in the *area-projection*. The results obtained with network projections (of both the *area* and *attribute* classes) seems to be more refined than those obtained with the complete graphs, better reflecting the complexity of the data.

REFERENCES

ALVES, M. Q. *Ambientes referenciais e monitoramento da recuperação de áreas mineradas no Quadrilátero Ferrífero.* 2019. 140 f. Universidade Federal de Viçosa, 2019.

BARABÁSI, A. L. *Network Science*. [S.l.]: Cambridge University Press, 2016.

BARBERÁN, A. *et al.* Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME Journal*, v. 6, n. 2, p. 343–351, 2012.

BASTIAN, M.; HEYMANN, S.; JACOMY, M. Gephi: An open source software for exploring and manipulating networks. BT - International AAAI Conference on Weblogs and Social. *International AAAI Conference on Weblogs and Social Media*, p. 361–362, 2009.

BECKETT, S. J. Improved community detection in weighted bipartite networks. *Royal Society Open Science*, v. 3, n. 1, 2016.

BEN-ELI, Michael. Understanding Systems. In *Systems thinking and systems modeling*: a course of understand systems and creating systems models. Switzerland, 2019. module one, p.1-28. Available in: < https://systemsinnovation.io/system-dynamics-book/> Access in 01.04.2020.

BENITES, V. DE M. *et al.* Caracterização dos solos em duas toposseqüências sobre diferentes litologias em áreas altimontanas na Serra da Mantiqueira. p. 31, 2003.

BHARDWAJ, A. K. *et al.* Ecological management of intensively cropped agro-ecosystems improves soil quality with sustained productivity. *Agriculture, Ecosystems and Environment*, v. 140, n. 3–4, p. 419–429, 2011. Disponível em: <http://dx.doi.org/10.1016/j.agee.2011.01.005>.

BLONDEL, V. D. *et al.* Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, v. 2008, n. 10, 2008.

BORGES, SILVANO R. *et al.* Practices for rehabilitating bauxite-mined areas and an integrative approach to monitor soil quality. *Land Degradation and Development*, v. 30, n. 7, p. 866–877, 2019.

BORGES, SILVANO RODRIGUES. *Qualidade do solo em áreas em recuperação com forrageiras e cafeeiro pós-mineração de bauxita*. . [S.l.]: Universidade Federal de Viçosa, 28 fev. 2013.

CÁRDENAS, J. P. *et al.* Soil porous system as heterogeneous complex network. *Geoderma*, v. 160, n. 1, p. 13–21, 2010. Disponível em: <http://dx.doi.org/10.1016/j.geoderma.2010.04.024>.

COTA, W. *Métodos de simulação de processos epidêmicos Markovianos em redes complexas*. 2016. 54 f. Universidade Federal de Viçosa, MG, 2016.

FANG, L. *et al.* A long-term study on the soil reconstruction process of reclaimed land by coal gangue filling. *Catena*, v. 195, p. 104874, 1 dez. 2020.

FILOTAS, E. *et al.* Viewing forests through the lens of complex systems science. *Ecosphere*, v. 5, n. 1, p. 1–23, 2014.

FORTUNATO, S. Community detection in graphs. *Physics Reports*, v. 486, n. 3–5, p. 75–174, 2010. Disponível em: <http://dx.doi.org/10.1016/j.physrep.2009.11.002>.

FORTUNATO, S.; CASTELLANO, C. *Community Structure in Graphs*. . [S.l: s.n.], 2007.

GUIMARÃES, P. T. G.; ALVAREZ, V. H.; RIBEIRO, A. C. 5 Aproximacao. *Recomendações Para o Uso de Corretivos e Fertilizantes em Minas Gerais - 5º Aproximação*, p. 13–20, 1999. Disponível em: <https://www.google.com.br/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=recomendações para o uso de corretivos e fertilizantes em minas gerais>.

HAIR, J. F. *Análise Multivariada de Dados*. Porto Alegre: Bookman, 2009.

HOLOVATCH, Y.; KENNA, R.; THURNER, S. Complex systems: physics beyond physics. *European Journal of Physics*, v. 38, n. 2, p. 1–22, 2017.

JACOMY, M. *et al.* ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE*, v. 9, n. 6, p. 1–12, 2014.

JOHN, B. *et al.* Storage of organic carbon in aggregate and density fractions of silty soils under different types of land use. *Geoderma*, v. 128, n. 1–2, p. 63–79, 2005.

KHAN, B. S.; NIAZI, M. A. Network Community Detection: A Review and Visual Survey. 2017. Disponível em: <http://arxiv.org/abs/1708.00977>.

KWAPIEŃ, J.; DROZDZ, S. Physical approach to complex systems. *Physics Reports*, v. 515, n. 3–4, p. 115–226, 2012. Disponível em: <http://dx.doi.org/10.1016/j.physrep.2012.01.007>.

LIEBIG, M.A; VARVEL, G; DORAN, J. A simple performance-based index for assessing multiple agroecosystem functions. *Agronomy Journal*. Mar;93(2):313-8. 2001.

LLOYD, S. Measures of Complexity: A Nonexhaustive List. *IEEE Control Systems*, v. 21, n. 4, p. 7–8, 2001.

LORENZ, K.; LAL, R.; SHIPITALO, M. J. Stabilization of organic carbon in chemically separated pools in no-till and meadow soils in Northern Appalachia. *Geoderma*, v. 137, n. 1–2, p. 205–211, 2006.

MA, B. *et al.* Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. *ISME Journal*, v. 10, n. 8, p. 1891–1901, 2016. Disponível em: <http://dx.doi.org/10.1038/ismej.2015.261>.

MOONEY, S. J.; KOROŠAK, D. Using Complex Networks to Model Two- and Three-Dimensional Soil Porous Architecture. *Soil Science Society of America Journal*, v. 73, n. 4, p. 1094–1100, 1 jul. 2009. Disponível em: <http://doi.wiley.com/10.2136/sssaj2008.0222>. Acesso em: 4 jul. 2020.

NEWMAN, M. E. J. Analysis of weighted networks. *Physical Review E - Statistical Physics,*

Plasmas, Fluids, and Related Interdisciplinary Topics, v. 70, n. 5, p. 9, 2004.

NEWMAN, M. E. J. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America*, v. 103, n. 23, p. 8577–8582, 2006.

OPSAHL, T.; PANZARASA, P. Clustering in weighted networks. *Social Networks*, v. 31, n. 2, p. 155–163, 2009.
PAVLOPOULOS, G. A. *et al. Bipartite graphs in systems biology and medicine: A survey of methods and applications*. [S.l: s.n.], 2018. v. 7.

PEREIRA, Aianã Francisco Santos. *Florística, fitossociologia e relação solo-vegetação em campo rupestre ferruginoso do quadrilátero ferrífero, MG.* Viçosa, MG, 2010. xi, 97 f.

PÉREZ-RECHE, F. J. *et al.* Prominent effect of soil network heterogeneity on microbial invasion. *Physical Review Letters*, v. 109, n. 9, p. 1–5, 2012.

PINTO, F. S.; MARIA, R.; CASTILHOS, V. Biomassa E Atividade Microbiana Em Solo Construído Após Mineração De Carvão E Submetido a Diferentes Coberturas Vegetais. *Revista Brasileira de Agrociência*, v. 14, n. 3, p. 515–525, 2008.

SAMEC, M. *et al.* Quantifying soil complexity using network models of soil porous structure. *Nonlinear Processes in Geophysics*, v. 20, n. 1, p. 41–45, 2013.

SER. The SER International Primer o n Ecological Restoration. *British journal of pharmacology*, v. 55, n. 2, p. 282P-283P, 2002. Disponível em: <http://www.ncbi.nlm.nih.gov/pubmed/116%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1666813>.

SYSTEMS INNOVATION. What is a complex system? Available in: <https://www.youtube.com/watch?v=vp8v2Udd_PM>. Access in 01.04.2020.

SHRESTHA, R. K.; LAL, R. Land use impacts on physical properties of 28 years old reclaimed mine soils in Ohio. *Plant and Soil*, v. 306, n. 1–2, p. 249–260, 2008.

SIEGENFELD, A. F.; BAR-YAM, Y. An Introduction to Complex Systems Science and its Applications. p. 1–14, 2019. Disponível em: <http://arxiv.org/abs/1912.05088>.

SIMARD, S. W. Mycorrhizal networks and complex systems: Contributions of soil ecology science to managing climate change effects in forested ecosystems. *Canadian Journal of Soil Science*, v. 89, n. 4, p. 369–382, 2009.

STEENBOCK, W.; VEZZANI, F. M. Agrofloresta: aprendendo a produzir com a natureza. Curitiba: A autora, 2013. 139, [8] p.

ZHANG, Z.; WANG, J.; LI, B. Determining the influence factors of soil organic carbon stock in opencast coal-mine dumps based on complex network theory. *Catena*, v. 173, n. September 2018, p. 433–444, 2019.

Appendix A – Matrices M_1 and M_{1r}.

(a)

$M_1 \doteq$

	Forest	PreMin	PosMin
TOC	1.000	0.626	0.262
LOC	1.000	0.916	0.190
TN	1.000	0.676	0.243
P	0.260	1.000	0.133
Mn	0.562	1.000	0.352
Fe	1.000	0.418	0.758
Zn	0.982	1.000	0.115
pH	0.819	1.000	0.902
Al3	1.000	0.064	0.090
HAl	1.000	0.491	0.266
EP	0.741	0.905	1.000
TEB	0.140	1.000	0.216
BD	0.467	0.566	1.000
Mic	0.893	1.000	0.857
Mac	1.000	0.865	0.635
TP	1.000	0.961	0.750
AciP	1.000	0.589	0.209
AlkP	1.000	0.683	0.000
Gluc	1.000	0.828	0.278
CCO2	0.783	0.783	1.000

(b)

$M_{1r} \doteq$

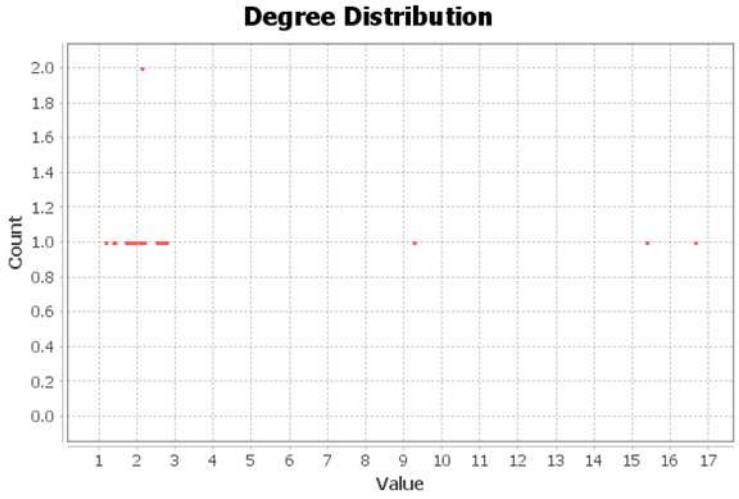
	Forest	PreMin	PosMin
TOC	1.000	0.626	0.262
LOC	1.000	0.916	0.190
TN	1.000	0.676	0.243
P	1.000	0.260	0.873
Mn	1.000	0.562	0.790
Fe	1.000	0.418	0.758
Zn	1.000	0.982	0.133
pH	1.000	0.819	0.917
Al3	1.000	0.064	0.090
HAl	1.000	0.491	0.266
EP	1.000	0.836	0.741
TEB	1.000	0.140	0.924
BD	1.000	0.902	0.467
Mic	1.000	0.893	0.964
Mac	1.000	0.865	0.635
TP	1.000	0.961	0.750
AciP	1.000	0.589	0.209
AlkP	1.000	0.683	0.000
Gluc	1.000	0.828	0.278
CCO2	1.000	1.000	0.783

Source: the author.

Note: (a) M_1 shows the weights (w_{ia}) of the edges of *Network-1*; (b) M_{1r}shows the weights (w_{ir}) of the edges of *Network-1r. Forest:* Atlantic forest; *PreMin*: 10-year-old coffee crop, pre-mining; *PosMin*: reconstructed soil, six months after mining; *TOC*; total organic carbon; *LOC*: labile organic carbon; *TN*: total nitrogen; *P*: available phosphorous; *Mn*: manganese content; *Fe*: iron content; *Zn*: zinc content; *pH*: pH in water; *Al3*: exchangeable aluminum, Al^{3+}; *HAl*: potential acidity, $H + Al^{3+}$; *EP*: soil equilibrium phosphorus; *TEB*: total exchangeable basis, K^{+}+ Ca^{2+}+ Mg^{2+}; *BD*: bulk density; *Mic*: microporosity; *Mac*: macroporosity; *TP*: total porosity; *AciP*: acid phospomonoesterase enzyme activity; *AlkP*: alkaline phospomonoesterase enzyme activity; *Gluc*: ß-glucosidase enzyme activity; *CCO2*: microbial respiration.

Appendix B – Weighted degree distributions of *Networks 1* and *Network-1r.*

(a)

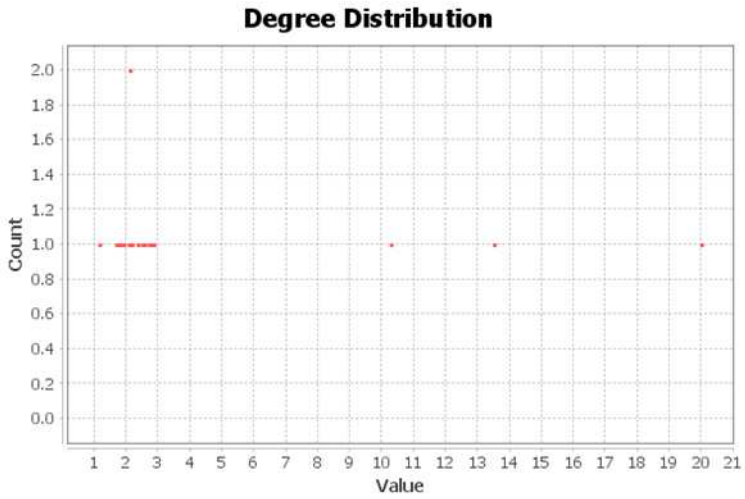


(b)

Forest	16.647
PreMin	15.371
PosMin	9.256
Mic	2.750
pH	2.721
TP	2.711
EP	2.646
CCO2	2.566
Mac	2.500
Fe	2.176
LOC	2.106
Gluc	2.106

Zn	2.097
BD	2.033
TN	1.919
Mn	1.914
TOC	1.888
AciP	1.798
HAl	1.757
AlkP	1.683
P	1.393
TEB	1.356
Al3	1.154

(c)



(d)

Forest	20.000
PreMin	13.511
PosMin	10.273
Mic	2.857
CCO2	2.783
pH	2.736
TP	2.711
EP	2.577
Mac	2.500
BD	2.369
Mn	2.352
Fe	2.176

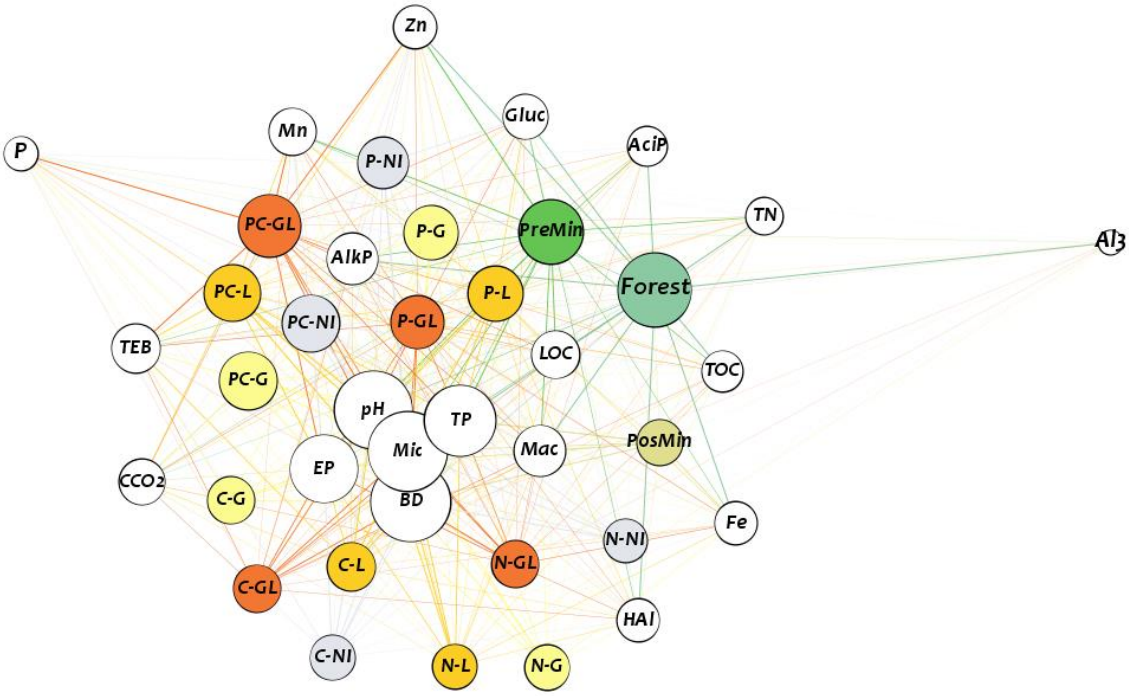
P	2.133
Zn	2.115
LOC	2.106
Gluc	2.106
TEB	2.064
TN	1.919
TOC	1.888
AciP	1.798
HAl	1.757
AlkP	1.683
Al3	1.154

Source: the author.

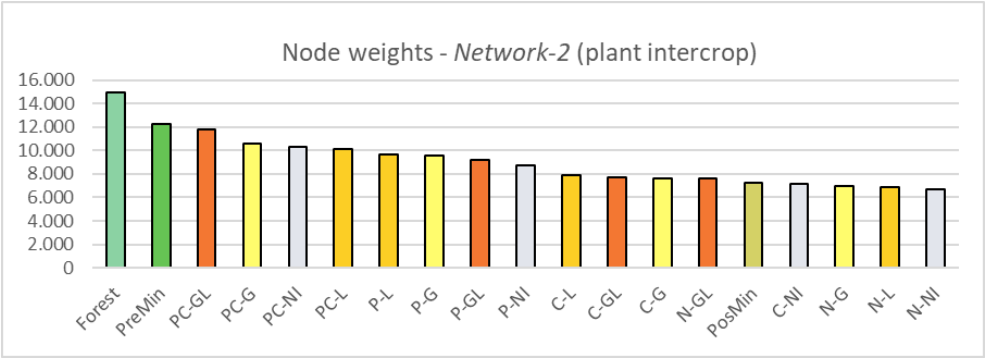
Note: (a) and (c) weighted degree distributions of *Network-1 and Network-1r,* reported by *Gephi*; (b) and (d) node weights of *Network-1* and *Network-1r*, in descending order.

Appendix C – *Network-2:* treatments colored by plant intercrop.

(a)



(b)



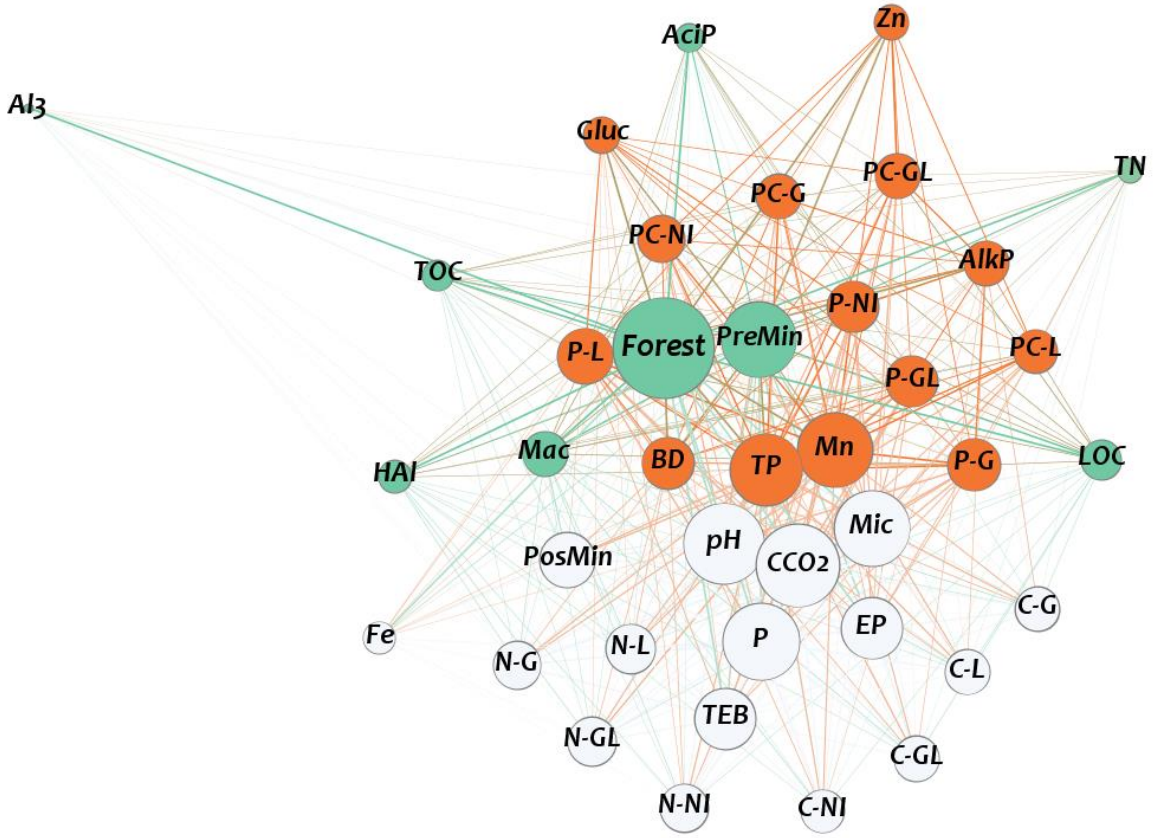
Plant intercrop

NI	no intercrop	G	Grass	L	Legume	GL	Grass + Legume

Source: the author.

Note: (a) Network-2 with treatments colored by plant intercrop; (b) node weights calculated by *Gephi*, in descending order. *Forest:* Atlantic forest; *PreMin*: 10-year-old coffee crop, pre-mining; *PosMin*: reconstructed soil, six months after mining; *N-*: no fertilization; *P-*: Poultry litter fertilization; *C-*: chemical fertilization; *PC-*: poultry litter and chemical fertilizations combined.

Appendix D – The three communities of *Network-2*r.

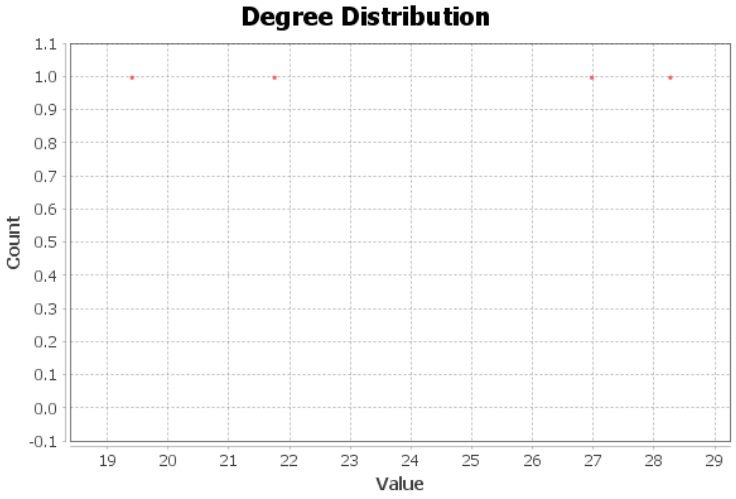


Source: the author.

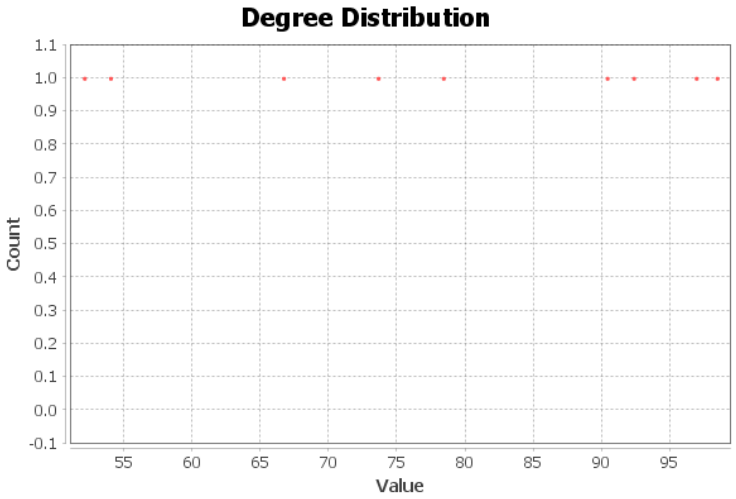
Note: the three communities (or modularity classes) of *Network-2r* and their respective nodes: [0] *Forest, PreMin, TOC, LOC, TN, Al3, HAl, Mac* and *AciP;* [1] *P-G, P-GL, P-L, P-NI, PC-G, PC-GL, PC-L, PC-NI, Mn, Zn, BD, TP, AlkP* and *Glic;* [2] *PosMIn, N-G, N-GL, N-L, N-NI, C-G, C-GL, C-L, C-NI, P, Fe, pH, TEB, Mic,* and *CCO2. Forest:* Atlantic forest; *PreMin*: 10-year-old coffee crop, pre-mining; *PosMin*: reconstructed soil, six months after mining; *N-*: no fertilization; *P-*: Poultry litter fertilization; *C-*: chemical fertilization; *PC-*: poultry litter and chemical fertilizations combined; *G*: grass intercrop; *L*: leguminous intercrop; *GL*: grass + leguminous mix intercrop; *NI*: no intercrop.

Appendix E – Weighted degree distribution of Networks 5, 6 and 7 *attribute-projection*.

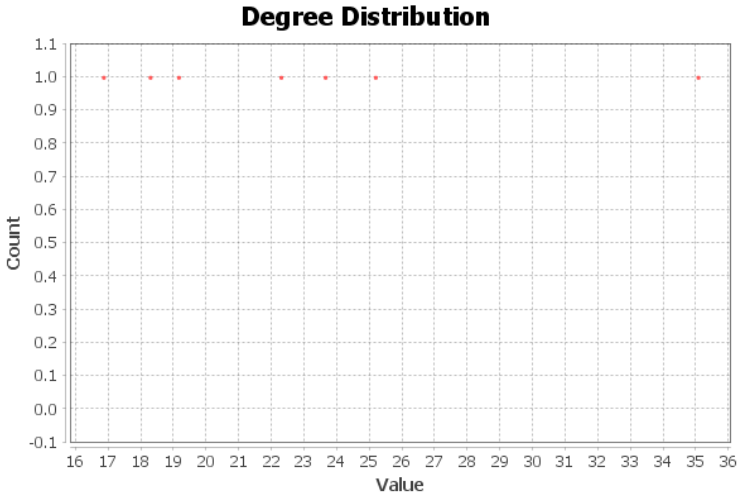
(a)



(b)



(c)

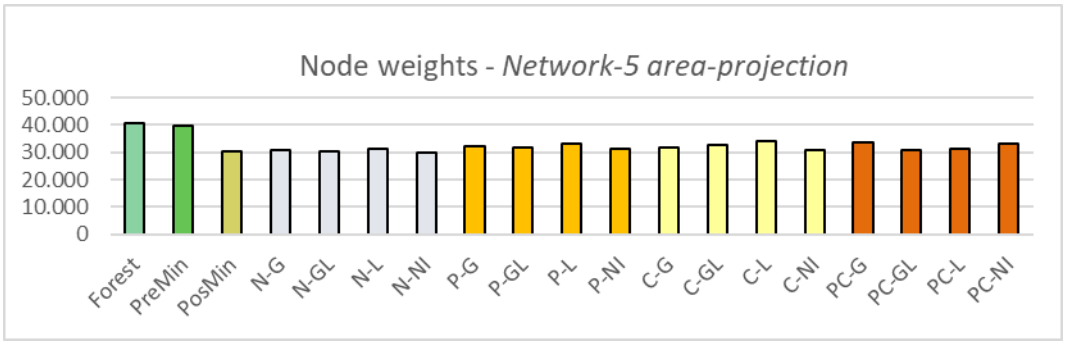


Source: the author.

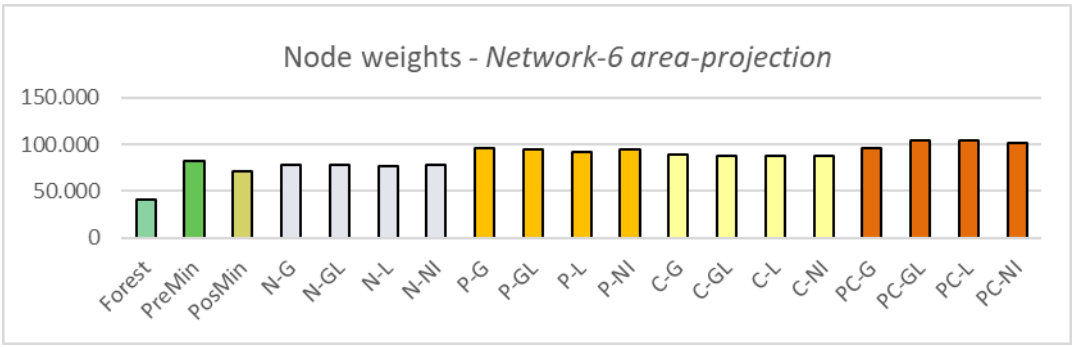
Note: (a), (b) and (c): the weighted degree distribution of *Networks 5, 6* and *7 attribute-projection*, respectively.

Appendix F – Physical, chemical and organic-biological soil quality from *Networks 5, 6* and *7 area-projections.*

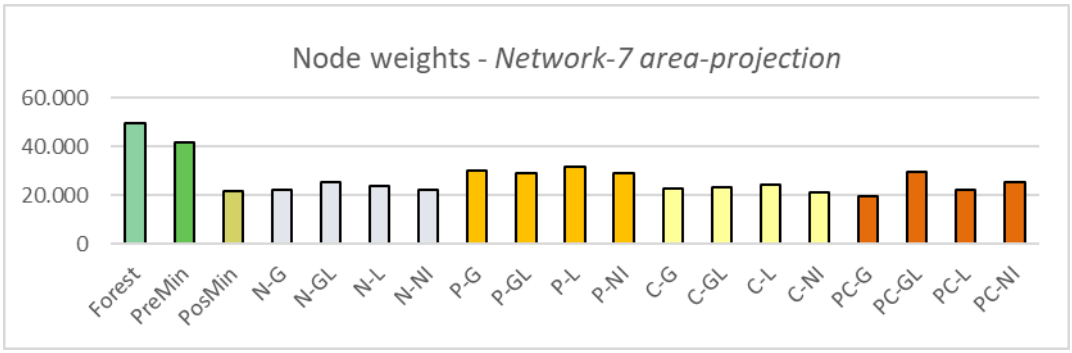
(a)



(b)



(c)



Fertilization type

N-	no fertilizer	C-	chemical	P-	poultry litter	PC-	poultry litter + chemical

Source: the author.

Note: *Forest:* Atlantic forest; *PreMin*: 10-year-old coffee crop, before mining; *PosMin*: reconstructed soil, six months after mining. *G*: grass intercrop; *L*: leguminous intercrop; *GL*: grass + leguminous mix intercrop; *NI*: no intercrop.

Appendix G –BD scores and BD weights of Network-2p.

(a)

	Forest	PreMin	PosMin	N-G	N-GL	N-L	N-NI	P-G	P-GL	P-L	P-NI	C-G	C-GL	C-L	C-NI	PC-G	PC-GL	PC-L	PC-NI
BD scores [1]	1.000	0.826	0.467	0.425	0.410	0.438	0.419	0.456	0.442	0.467	0.419	0.445	0.456	0.467	0.435	0.467	0.452	0.429	0.471
BD weights [2]	1.000	0.914	0.532	0.446	0.410	0.475	0.432	0.511	0.482	0.532	0.432	0.489	0.511	0.532	0.468	0.532	0.504	0.453	0.540
difference	0%	9%	7%	2%	0%	4%	1%	5%	4%	7%	1%	4%	5%	7%	3%	7%	5%	2%	7%

(b)

	Forest	PreMin	PosMin
BD [3] $(kg.dm^{-3})$	0.57	0.69	1.22
difference from Forest	0.00	0.12	0.65
		(a)	(b)
ratio of differences (b/a)			**5.42**
BD scores [1]	1.00	0.83	0.47
difference from Forest	0.00	0.17	0.53
		(c)	(d)
ratio of differences (d/c)			**3.06**
BD weights [2]	1.00	0.91	0.53
difference from Forest	0.00	0.09	0.47
		(e)	(f)
ratio of differences (f/e)			**5.42**

(c)

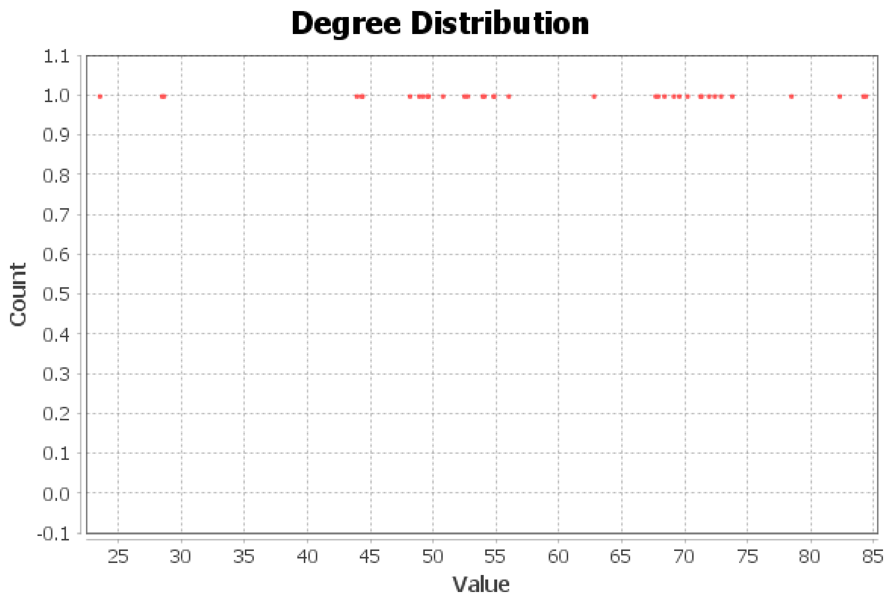
	Forest	PC-GL	PC-L
BD [3] $(kg.dm^{-3})$	0.57	1.26	1.33
difference from Forest	0.00	0.69	0.76
		(a)	(b)
ratio of differences (b/a)			**1.10**
BD scores [1]	1.00	0.45	0.43
difference from Forest	0.00	0.55	0.57
		(c)	(d)
ratio of differences (d/c)			**1.04**
BD weights [2]	1.00	0.50	0.45
difference from Forest	0.00	0.50	0.55
		(e)	(f)
ratio of differences (f/e)			**1.10**

Source: the author

Note: [1] calculated according to Liebig et al (2001) apud Borges et al. (2019, p. 869); [2] calculated according to equation (2); [3] values from Table 1.

Appendix H – Weighted degree distribution of *Network-8 attribute-projection.*

(a)



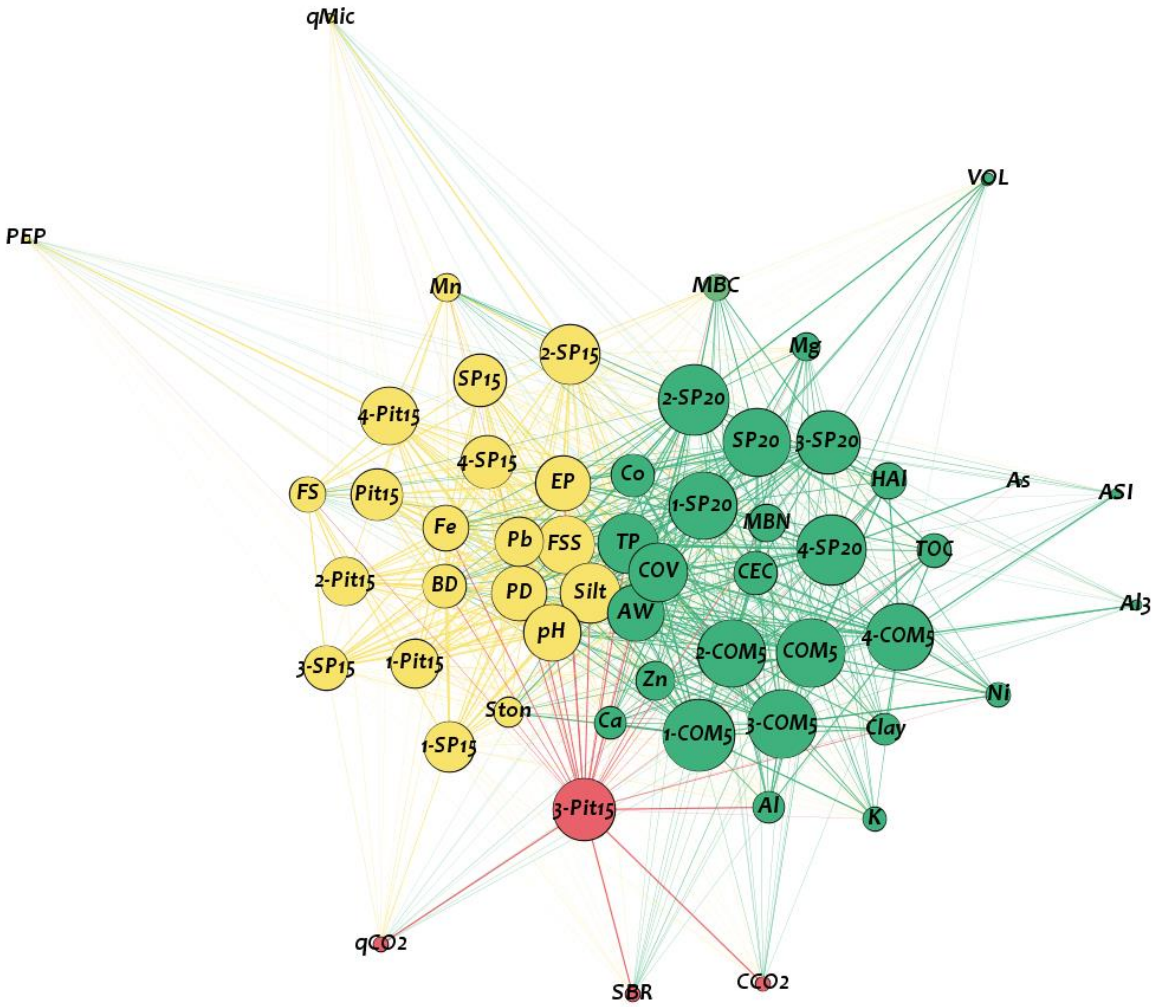
(b)

Group	Attribute	Weight
1	qCO2	23.390
1	VOL	28.336
1	Ni	28.485
2	As	43.829
2	Ca	44.190
2	Mg	44.258
2	Al3	48.060
2	Ston	48.759
2	qMic	49.041
2	MBC	49.409
2	MBN	49.494
2	Zn	50.665
2	Co	52.386
2	Clay	52.610
2	FS	53.803
2	Fe	53.941
2	ASI	54.686
2	Pb	54.732
2	Mn	55.902
3	K	62.662
4	CEC	67.560
4	FSS	67.757
4	Al	68.259
4	TOC	69.017
4	BD	69.451
4	AW	70.109
4	CCO2	71.110
4	HAI	71.196
4	SBR	71.800
4	COV	72.258
4	PD	72.761
4	Silt	73.659
5	EP	78.366
6	pH	82.197
6	PEP	84.093
6	TP	84.258

Source: the author.

Note: (a) the weighted degree distribution of *Network-8 attribute-projection,* reported by *Gephi;* (b) node weights of it attribute in this projection; the *Similarity Indices* (highlighted in gray) correspond to the two first groups of nodes in the distribution (a)

Appendix I – The three communities of *Network-10*.



Source: the author.

Note: nodes belonging to the same community [or modularity class] have the same color: [0] in green: *1-COM5, 2-COM5, 3-COM5, 4-COM5, COM5, 1-SP-20, 2-SP-20, 3-SP-20, 4-SP-20, SP-20, Clay, TP, AW, K, Ca, Mg, Al3, HAl, CEC, ASI, TOC, As, Co, Ni, Zn, Al, MBC, MBN, COV* and *VOL*; [1] in orange: *1-SP-15, 2-SP-15, 3-SP-15, 4-SP-15, 1-Pit-15, 2-Pit-15, 4-Pit-15, Pit-15, Ston, FS, FSS, Silt, BD, PD, pH, EP, PEP, Pb, Mn, Fe* and *qMic; [2]* in red: *3-Pit-15, CCO2, SBR* and *qCO2. COM5*: compensation area, 5 years under reclamation; *SP15*: sterile pile, 15 years under reclamation; *Pit15*: pit, 15 years under reclamation; *SP20:* sterile pile, 20 years under reclamation.

Appendix J –Indicators Scores and weights of *Pit-15* in *Network-11p*.

Indicator	Scores [1]	Weights [2]	difference
FSS	0.450	0.450	0%
Silt	0.540	0.559	-2%
Clay	0.240	0.600	**-36%**
BD	0.750	0.749	0%
PD	0.890	0.893	0%
TP	0.840	0.864	-2%
Fe	0.890	0.892	0%
COV	0.490	0.682	**-19%**

Source: the author.

Note: [1] calculated by Alves (2019, p.111) according to the methodology adapted from Bhardwaj (2011) and Borges (2013), proposed by Liebig et al. (2001); [2] calculated by us according to equation (5).

Annex A – Table 3: values of soil attributes and vegetation parameters of all plots in the areas under iron mining reclamation.

(continues)

		Plots				Mean	Plots				Mean	Plots				Mean	Plots				Mean
Attribute	unit	1-COM5	2-COM5	3-COM5	4-COM5	COMP5	1-SP15	2-SP15	3-SP15	4-SP15	SP15	1-Pit15	2-Pit15	3-Pit15	4-Pit15	Pit15	1-SP20	2-SP20	3-SP20	4-SP20	SP20
		Soil physical attributes																			
Ston	g kg^{-1}	356.98	108.90	125.49	25.71	154.27	108.61	207.96	129.90	143.89	147.59	179.78	98.56	156.64	203.02	160.45	189.88	176.77	95.66	94.25	139.14
FS	kg kg^{-1}	0.04	0.04	0.04	0.05	0.05	0.23	0.23	0.25	0.20	0.23	0.19	0.27	0.10	0.23	0.23	0.05	0.09	0.09	0.09	0.08
FSS	kg kg^{-1}	0.64	0.66	0.62	0.63	0.64	0.74	0.67	0.82	0.79	0.75	0.76	0.85	0.71	0.80	0.80	0.71	0.68	0.61	0.69	0.67
Silt	kg kg^{-1}	0.60	0.61	0.58	0.58	0.59	0.52	0.43	0.57	0.58	0.52	0.57	0.57	0.61	0.57	0.57	0.66	0.59	0.52	0.60	0.59
Clay	kg kg^{-1}	0.28	0.28	0.32	0.31	0.30	0.07	0.09	0.03	0.06	0.06	0.04	0.04	0.17	0.04	0.04	0.15	0.17	0.15	0.16	0.16
BD	g cm^{-3}	0.94	0.90	0.92	0.87	0.91	1.64	1.50	1.92	1.70	1.69	1.87	1.77	1.12	2.09	1.91	0.91	0.97	0.83	0.82	0.88
PD	g cm^{-3}	2.83	2.59	2.80	2.48	2.68	3.51	3.40	3.76	3.52	3.55	3.42	3.40	3.02	3.85	3.56	2.69	2.81	2.52	2.40	2.61
TP	m m^{-3}	0.67	0.65	0.67	0.65	0.66	0.53	0.56	0.49	0.52	0.52	0.45	0.48	0.63	0.46	0.46	0.66	0.66	0.67	0.66	0.66
AW	kg kg^{-1}	0.22	0.24	0.26	0.24	0.24	0.19	0.17	0.18	0.19	0.18	0.17	0.19	0.23	0.15	0.17	0.24	0.24	0.25	0.25	0.25
		Soil chemical attributes																			
pH	-	5.62	5.50	5.24	4.94	5.32	5.93	5.04	5.65	5.22	5.46	6.61	6.46	5.74	5.85	6.31	4.87	4.26	4.59	5.03	4.69
K	mg dm^{-3}	179.25	121.50	115.50	78.00	123.56	27.25	42.50	13.25	35.25	29.56	29.00	21.25	28.75	30.75	27.00	56.50	45.00	74.00	66.75	60.56
Ca	$cmol_c$ dm^{-3}	3.81	2.95	2.24	1.15	2.54	0.64	1.65	0.65	1.64	1.14	1.36	0.86	0.99	1.30	1.17	3.35	2.65	1.12	1.44	2.14
Mg	$cmol_c$ dm^{-3}	0.77	0.54	0.41	0.22	0.48	0.10	0.28	0.14	0.30	0.20	0.23	0.12	0.19	0.22	0.19	0.92	0.52	0.40	0.50	0.58
Al3	$cmol_c$ dm^{-3}	0.09	0.14	0.36	1.37	0.49	0.00	0.14	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.14	0.36	0.32	0.20	0.25
HAl	$cmol_c$ dm^{-3}	5.30	5.43	6.48	8.20	6.35	2.70	4.38	1.13	2.50	2.68	0.55	0.90	2.33	1.75	1.07	4.93	7.10	7.48	6.25	6.44
CEC	$cmol_c$ dm^{-3}	10.34	9.22	9.43	9.77	9.69	3.51	6.41	1.95	4.53	4.10	2.21	1.93	3.58	3.35	2.50	9.34	10.38	9.18	8.36	9.31
ASI	%	1.78	3.83	11.53	48.35	16.37	0.00	7.50	0.00	0.00	1.88	0.00	0.00	0.00	0.00	0.00	3.23	10.68	17.10	10.13	10.28
EP	mg_c L^{-1}	29.15	28.98	25.95	22.48	26.64	35.20	38.83	48.23	40.98	40.81	44.98	43.20	36.25	48.45	45.54	41.45	40.85	49.43	46.85	44.64
PEP	-	0.08	0.07	0.08	0.09	0.08	0.06	0.10	0.16	0.21	0.13	0.16	0.08	0.05	2.61	0.95	0.12	0.14	0.15	0.12	0.13
As	mg.kg^{-1}	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.66	0.00	0.00	0.00	0.00	0.00	12.22	3.06
Co	mg.kg^{-1}	14.88	14.37	17.13	9.31	13.92	6.91	7.98	4.99	5.60	6.37	9.09	9.68	6.49	12.99	10.59	14.90	14.85	7.51	12.03	12.32
Pb	mg.kg^{-1}	28.41	33.84	19.68	14.30	24.06	26.80	23.48	18.48	35.51	26.07	25.20	29.05	31.54	36.52	30.26	21.61	26.61	17.03	23.07	22.08

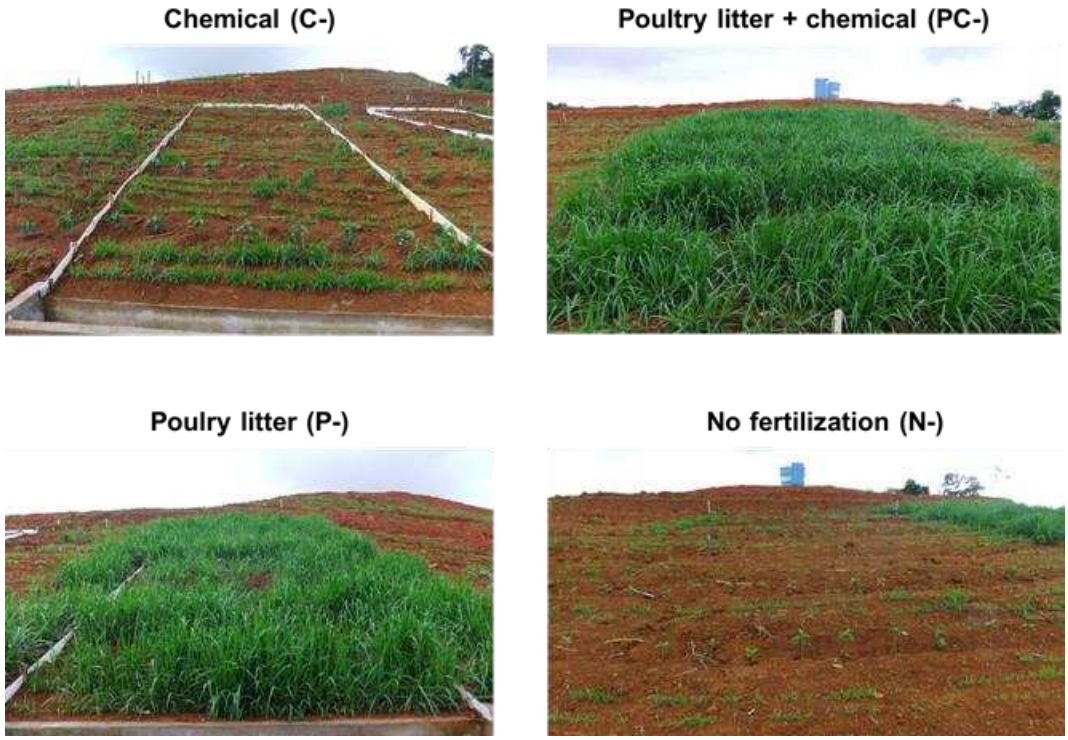
(conclusion)

		Plots				Mean	Plots				Mean	Plots				Mean	Plots				Mean
Attribute	unit	1-COM5	2-COM5	3-COM5	4-COM5	COMP5	1-SP15	2-SP15	3-SP15	4-SP15	SP15	1-Pit15	2-Pit15	3-Pit15	4-Pit15	Pit15	1-SP20	2-SP20	3-SP20	4-SP20	SP20
Ni	$mg.kg^{-1}$	96.22	101.58	130.59	130.80	114.80	1.42	7.49	0.00	0.00	2.23	4.84	1.63	23.23	0.45	2.31	46.48	43.95	58.30	91.65	60.09
Zn	$mg.kg^{-1}$	69.63	61.31	70.23	68.50	67.42	19.00	24.07	8.61	26.83	19.63	25.38	26.75	43.93	44.14	32.09	50.30	41.92	23.20	25.20	35.15
Mn	$mg.kg^{-1}$	856.84	713.54	740.96	414.11	681.37	3032.62	6480.13	1308.08	4167.08	3746.98	1471.17	3688.84	1502.51	5561.90	3573.97	2502.64	4618.01	1751.87	2257.87	2782.60
Fe	$g.kg^{-1}$	97.27	96.59	92.93	68.80	88.90	255.46	232.33	233.51	247.86	242.29	254.02	263.08	206.35	278.66	265.25	139.16	165.36	144.37	126.80	143.92
Al	$g.kg^{-1}$	20.84	19.44	25.73	23.03	22.26	7.64	7.44	1.41	2.92	4.85	4.45	4.18	28.44	6.51	5.05	16.27	15.76	11.97	15.07	14.77
		Soil organic-biological attributes																			
TOC	$dag.kg^{-1}$	3.47	3.22	3.08	3.60	3.34	0.78	0.75	0.16	0.65	0.58	0.39	0.34	1.00	0.53	0.42	2.82	2.53	3.89	2.82	3.02
CCO2	$CO_2.kg^{-1}$	177.79	171.02	156.59	142.52	161.98	138.15	113.72	38.60	94.81	96.32	48.28	50.43	784.31	84.40	61.04	157.21	125.80	150.51	295.26	182.19
SBR	$CO_2.kg^{1}.h^{1}$	0.35	0.34	0.31	0.28	0.32	0.27	0.23	0.08	0.19	0.19	0.10	0.10	1.56	0.17	0.12	0.31	0.25	0.30	0.59	0.36
MBC	$mg.kg^{-1}$	234.76	240.00	145.73	129.26	187.44	120.47	221.82	43.24	203.54	147.26	124.52	59.97	186.74	158.34	114.28	267.79	551.51	438.86	232.96	372.78
MBN	$mg.kg^{-1}$	26.00	46.04	40.87	32.72	36.41	17.19	33.32	10.54	19.79	20.21	13.88	8.94	31.25	18.81	13.88	33.27	19.46	33.00	50.70	34.11
qMic	%	0.69	0.76	0.47	0.34	0.56	1.14	25.33	1.93	1.85	7.56	3.92	1.47	1.74	3.01	2.80	0.90	2.34	1.15	0.74	1.28
qCO2	x_c10^{-3}	1.52	1.42	2.60	2.22	1.94	6.61	1.21	3.50	3.58	3.73	0.69	2.23	12.80	0.99	1.30	1.31	0.45	0.72	2.78	1.32
		Vegetation parameters																			
COV	%	99.84	99.90	99.94	99.86	99.89	83.47	81.24	88.03	73.61	81.59	85.98	82.61	37.69	19.47	62.69	100.00	100.00	100.00	100.00	100.00
VOL	m^{-3}	115.06	85.27	56.68	35.55	73.14	8.71	15.06	19.54	16.17	14.87	36.99	44.57	22.59	6.02	29.19	1230.00	2284.00	1064.94	1365.29	1486.06

Source: adapted from Alves, Maísa Q. (2019, unpublished)

Note: *COM5*: compensation area, 5 years under reclamation; *SP15*: sterile pile, 15 years under reclamation; *Pit15*: pit, 15 years under reclamation; *SP20:* sterile pile, 20 years under reclamation; *Ston*: stoniness; *FS*: fine sand; *FSS*: fine sand + silt; *BD*: bulk density; *PD*: particle density; *TP*: total porosity; *AW*: available water content; *pH*: pH in water; *K*: available potassium, K^+ ; *Ca*: exchangeable calcium, Ca^{2+}; *Mg*: exchangeable magnesium, Mg^{2+}; *Al3*: exchangeable aluminum, Al^{3+}; *HAl*: potential acidity, $H + Al^{3+}$; *CEC*: cation exchange capacity; *ASI*: aluminum saturation index; *EP*: equilibrium phosphorus; *PEP*: available phosphorous/equilibrium phosphorus; *TOC*: total organic carbon; *As*: semi-total arsenic; *Co*: semi-total cobalt; *Pb*: semi-total lead; *Ni*: semi-total nickel; *Zn*: semi-total zinc; *Mn*: semi-total manganese; *Fe*: semi-total iron; *Al*: semi-total aluminum; *CCO2*: microbial respiration; *SBR*: soil biomass respiration; *MBC*: microbial biomass carbon; *MBN:* microbial biomass nitrogen; *qMic*: microbial quotient; *qCO2*: metabolic quotient; *COV*: vegetation coverage; *VOL*: vegetation volume.

Annex B – Areas in recovery with coffee crop and *Brachiaria.*



Source: Silvano R. Borges (2010, unpublished),

Note: these areas correspond to grass intercrop treatments (-G), forty days after sowing, with four types of fertilization. Images kindly provided by professor Ivo Ribeiro da Silva.