Mining, Shaping, Visualizing, and Interpreting Instagram Hipertextual Networks of Freight Train Graffiti Communalities in North America Using Machine Learning Custom Models and Graphology

Angel R. Abundis [1,*]
[1] Departamento de Estudios de la Comunicación Social, CUCSH - UDG, 44260 Guadalajara, México;
E-Mail: abundiscomunicacion@gmail.com
* Corresponding author

Abstract

The practice of benching in graffiti has evolved over time, transitioning from a gathering point for graffiti writers in New York City subway stations, where they admired and valued the artwork on passenger vehicles, to becoming an integral part of graffiti on freight trains in North America. Nowadays, interventions decorating rolling stock that circulates transnationally are documented and shared in benching communities. Although the dynamics and geographical reach have shifted from hyperlocal to international through online platforms, the underlying principle remains the same: benching serves as a meeting place where writers appreciate each other's work and gain recognition.

This methodological-practical study explores the possibilities of analyzing communalities among graffiti writers on freight trains through their online publications. Communalities can be derived from data such as the types of documented graffiti, the number of likes, the quantity of comments, the communal glossary used in hypertextual tags, and the volume of posts published inside those hashtags

This text revolves around the exposure of three hypertextual conversations with different mining scales and analyzing scopes. It showcases the hashtags of a graffiti writer in freight trains (#kosm), a communal meeting point hashtag (#freightgraffiti), and a geographically focused hashtag (#portlandbench). By selecting the seed node in the mining iterator, different types of symbolic exchanges, participants, and content within Instagram metadata and those generated through training and inference of machine learning models can be analyzed.

While the interpretation of these three examples is central, the text also presents the encoded computational techniques for data extraction, construction, and visualization of user-generated conversations on Instagram. Parameters such as depth, the number of mined posts, and the concept of seed node in data mining are discussed. The text addresses the limitations and capabilities of the machine learning models used, including object detection in images and categorization of hypertextual tags in posts. Additionally, it highlights data cleansing and parameters such as gravity, scale-ratio, and centrality measures used for real-time visualization achieved through Graphology.

Keywords
computational social science, freight graffiti, datafication, digital methodology, network visualization, communication studies

1. Introduction

Benching is a central practice in the contemporary graffiti culture, with its roots tracing back to the subway stations of New York City during the golden era of the subway graffiti. At these iconic benching spots such as GrandCourse and 149st, graffiti writers would gather to observe, appreciate, and peer-review the graffiti. These spots workd not only as observation points but also as social hubs where writers could engage with one another, exchange techniques, and earn recognition within their specific community of practice.
With the emergence and consolidation of user-generated content platforms like Instagram, the tradition of benching undergoes a transformation, adapting to new mediums. Hashtags such as #FreightBenching facilitate this transition, turning local benching spots into digital communities of practice. Here, graffiti writers and benchers gather share evidence of active and live geographically dispersed graffiti scenes. This shift not only expands the reach of different graffiti writers promotion but also creates a transnational peer-to-peer community, while still preserving the core principle of graffiti culture: the competitive getting up.
This paper aims to demonstrate the application of computational social science methods in visualizing and analyzing three hypertextual conversations with different mining scales and analytical scopes. Specifically, it explores the *autopromotion* of a freight graffiti writer using the hashtag #kosm, as well as the *communal retransmission* of graffiti interventions in two distinct spaces: the general meeting point hashtag #freightgraffiti and a geographically focused one, #portlandbench. Through these examples, the paper discusses the computational techniques employed in studying digital graffiti communities on Instagram.

2. Mine and Inference #freightgraffiti

Graffiti on freight trains in the North American region can be studied from various perspectives: from the expansion of transnational circulation circuits and the consolidation of a long-distance messaging system to the curating processes through the analysis of local and international specialized magazines, and the symbolic production of graffiti writers envolved in this community of practice.
Physically, graffiti on freight trains has a particular spread dynamic. Writers mark the sides of freight vehicles in “yards,” railway tracks used as garages, located on the outskirts of the cities or rural areas where the rail vehicles may wait for days or months before embarking on their journeys, which can be local, national, or transnational. It is in distant latitudes where other *writers* or *benchers* watch, evaluate, and document these interventions, forming a transnational circuit of New York tradition graffiti, both physical and digital.
However, in the socio-digital dimension, particularly in *Instagram* posts as materiality, several analyzable elements come together. *Instagram* is a user-generated content platform where content is organized using hypertextual tags (or hashtags). A *graffiti writer* may tag the post of a photograph of their recently completed piece with their name #kosm and a community hashtag like #freightgraffiti with the intention of other writers or graffiti enthusiasts watch this one. *Benchers*, on the other hand, primarily document interventions and may tag the writer #mecrograffiti, the location #portlandbenching where they documented the railway vehicle, and a communal tag like #fr8porn. Together, these practices generate networks of symbolic exchange, which through *self-promotion* by writers and *retransmission* by benchers, allow us to approach this phenomenon with symbolic elements, geographical references, writer and crews entities, that participants in this community of practices share and value collectively, driven by a core practice in contemporary graffiti: the *getting up* (Castleman, 1980).
User-generated content platforms generate vast amounts of data, a phenomenon referred to as the “dataification of social life” (Hepp, 2020; Gomez-Cruz, 2022), which is fundamental to this research approach. Writers and benchers actively contribute to this data pool by updating, categorizing, and reviewing interventions on rail vehicles using Instagram as a collaborative database. The platform aggregates an unprecedented volume of user-contributed data. This study utilizes computational tools to systematically collect and analyze this data with a specific theoretical-methodological perspective.
Information mining processes are employed to mimic human behavior in backing up thousands of publications, as detailed in section 2.1 Data Collection Processes. The process involves capturing images and captions of publications tagged with specific hashtags and attributing them to their respective authors. Subsequently, computational processes infer various aspects of the data, including the type of graffiti,

portrayal of daily life elements, usage of geographical or communal tags, and identification of graffiti writer tags and their crews. These inference processes are elucidated in section 2.2 Application of Machine Learning Techniques.

2.1. Data Collection Processes

The Data Collection Processes detailed herein delineate the systematic approach undertaken by the *Instagram Data Mining Bot* (idmb) to extract and process data from Instagram's really extensive and user-generated data repository. This segment provides a comprehensive list of idmb more relevant components, such as the hashtag iterator and seed node.

2.1.1 *idmb* a harmless mining bot

idmb operates within a structured sequence of specific processes. It comprises a series of functions responsible for tasks ranging from user authentication to data storage. Central to idmb functionality is the interaction with Instagram's API facilitated by the open source Instagrapi library (AdW0rd, 2022). Through this interface, we can access Instagram's database of user-generated content.

idmb backs up a really specific layout of data associated with each Instagram post. This includes essential elements such as user information (user ID, username, profile picture), media content (images or videos), captions, and engagement metrics (likes, comments). By saving these components, the mining bot ensures a working and constant repository to subsequent analysis and interpretation.

2.1.2 Seed node, is the exploration starting point

The seed node serves as a focal point for initiating the data collection process within idmb. It represents a strategically chosen hashtag from which data extraction begins. For instance, consider #freightgraffiti, which encapsulates the whole freight graffiti community of practice, is a general communal tag, in this case serves as a seed node within idmb, guiding the initial phase of data collection.

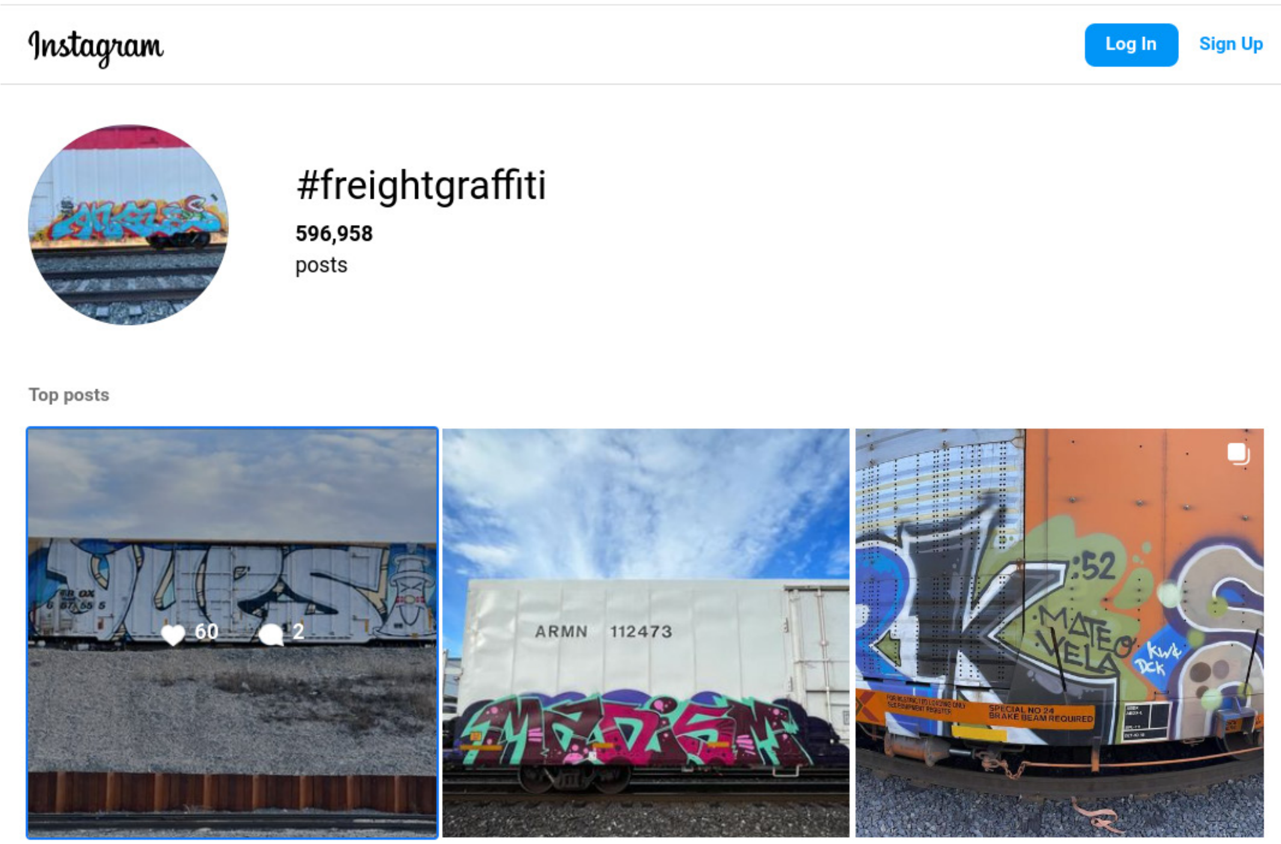

Figure 1: Instagram #FreightGraffiti screenshot.

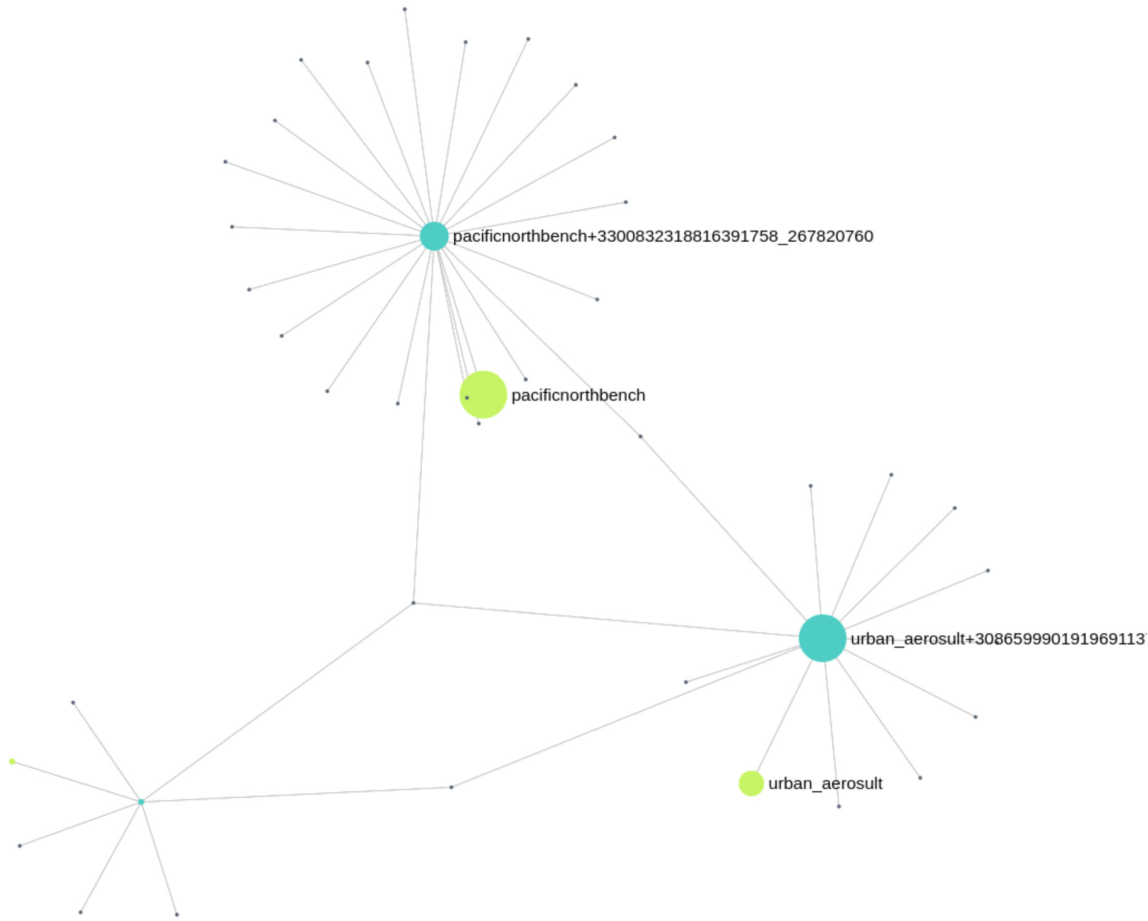

Figure 2: Graph of #freightgraffiti. Example: 0 mining depth and downloading 3 post's for each hashtag.

As an illustration, let's consider the seed node #freightgraffiti. The mining bot initiates the data collection process by extracting information from three mosted liked posts tagged with this hashtag (Figure 1). Each of these publications serves as a gateway to further exploration, because the post has several hashtags, each of them is a task that should be resolved by the mining bot with one condition: the task is should be inside the mining depth scope.

2.1.3 Hashtag Iterator

The hashtag iterator constitutes a fundamental function of idmb's data acquisition strategy, facilitating the systematic exploration of Instagram's content landscape. In the first run, this module iterates through all the post's tagged with the seed-node systematically, retrieving data pertaining to posts tagged with that specific hashtag. Then it will repeat the process for each hashtag founded in the captions, until the iterator mining depth is reached.

Continuing with our example of #freightgraffiti as the seed node, mining bot proceeds to iterate through each of the hashtag posts and downloading the three most liked publications, growing the network exponentially. The seed node acts as more than just a starting point. Carefully chosen, the seed node dictates the trajectory of exploration, guiding the analysis towards specific themes, topics, or communities of interest. Through the meticulous execution of the hashtag_iterator() function, idmb systematically get across the hypertextual landscape, downloading the dynamic conversations and interactions within this graffiti subculture

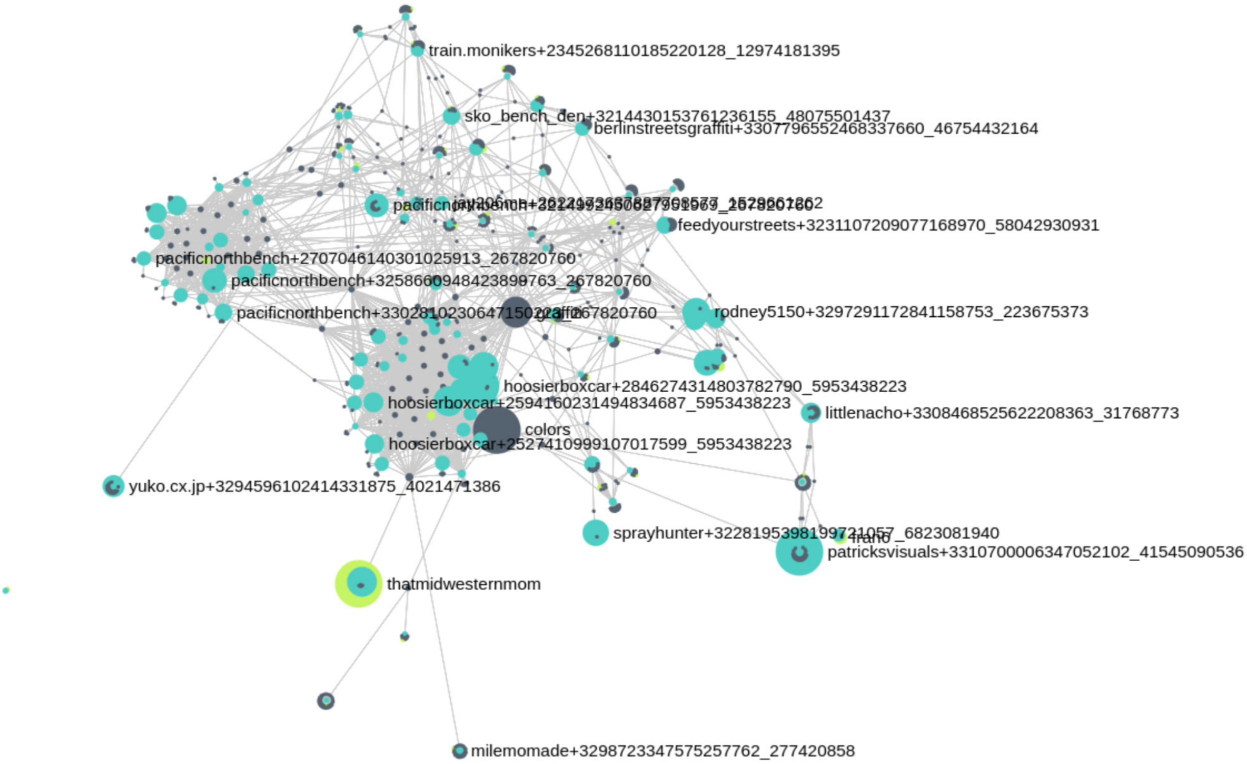

Figure 3: Graph of Freightgraffiti_1_hashtagTop_3_27419da5. Example: #freightgraffiti with 1 mining depth and downloading 3 post's for each hashtag.

2.2 Application of Machine Learning Techniques

This work employs *TensorFlow* (Abadi, 2015) and *spaCy* (Honnibal, 2020), two open-source libraries, to process and infer the existence of significant symbolic content in Instagram posts that have been indiscriminately downloaded by the mining bot. As will be seen later in section *3. Modeling & Interpretation*, evidence is generated using algorithms from network theory such as *ForceAtlas2* (Jacomy, 2014) or *Louvain Communities* to support the existence of user clusters that are grouped by the collective use of some specific hashtags. However, by employing machine learning models or computational algorithms, it is possible to detect the type of graffiti photographed or the significant words used to tag their posts, and link these inferences to the publications themselves.

This strategy aims to complex the network previously obtained between users, posts, and hypertextual tags by adding the symbolic elements they have used in their images and captions. This section discusses the computational processes executed and their outcomes.

2.2.1 TensorFlow Object Detection

A ResNet object detection model, based on a convolutional neural network architecture, was trained to identify visual elements, including different types of graffiti and railway company logos. For the model training, a dataset consisting of a total of 1592 images was utilized. These images were automatically collected from the Instagram platform using tags related to each type of graffiti, such as #wildstylegraffiti, #tagsandthrows, #monikers or #rollergraffiti.

Figure 4: Image processed with TensorFlow custom model, that identify wildstyle graffiti and "Ferromex" train identifier.

These images were labeled manually using the LabelImg tool. Each image was associated with one or more labels corresponding to the different types of graffiti. For example, 660 images were labeled with the "Tag" label, 320 images with "Character", 309 images each with "Bomb" and "Wildstyle" labels, 116 images with "Train Identifiers" (s_tren), 115 images with "Roller", 67 images with "3D", and 64 images with "Moniker". While the initial pilot based on this model yielded promising results by enabling the identification of graffiti and its variants in the analyzed posts, it is essential to acknowledge the need for continuous improvement and expansion of the training dataset. A larger and more diverse dataset could enhance the model's generalization capability and accuracy.

2.3 Natural Language Processing (NLP) Analysis

Natural Language Processing (NLP) is a crucial tool to approach the textual content of Instagram posts, particularly when the data set reaches thousands of freight graffiti posts. In this context, spaCy, a widely utilized natural language processing (NLP) library within the field of artificial intelligence, proves indispensable for deconstructing the hashtags, thereby facilitating the analysis of these character strings.

Processing natural language confronts challenges in complex hashtags like #FreightTrainGraffiti. For instance, spaCy treats them as whole word, a singular token, complicating to understand their content. To tackle this issue, the *hashtag_splitter()* function assumes a central role in fragmenting hashtags into manageable terms such as “Freight,” “Train,” and “Graffiti”.

2.3.1 Split a hashtag and find the words

Within the framework of the *hashtagsplitter()* function, the *parsetag()* and *findword()* components assume essential roles in processing hashtags. The *parsetag()* function is tasked with scrutinizing and segmenting the hashtags, removing the initial “#” character, and segregating the hypertext labels with whitespace, using a massive list of almost all the known words in spanish and english (word_list). This process effectively breaks down hashtags like #BoxcarArtGraffiti into separate terms such as “Boxcar,” “Art,” and “Graffiti”.

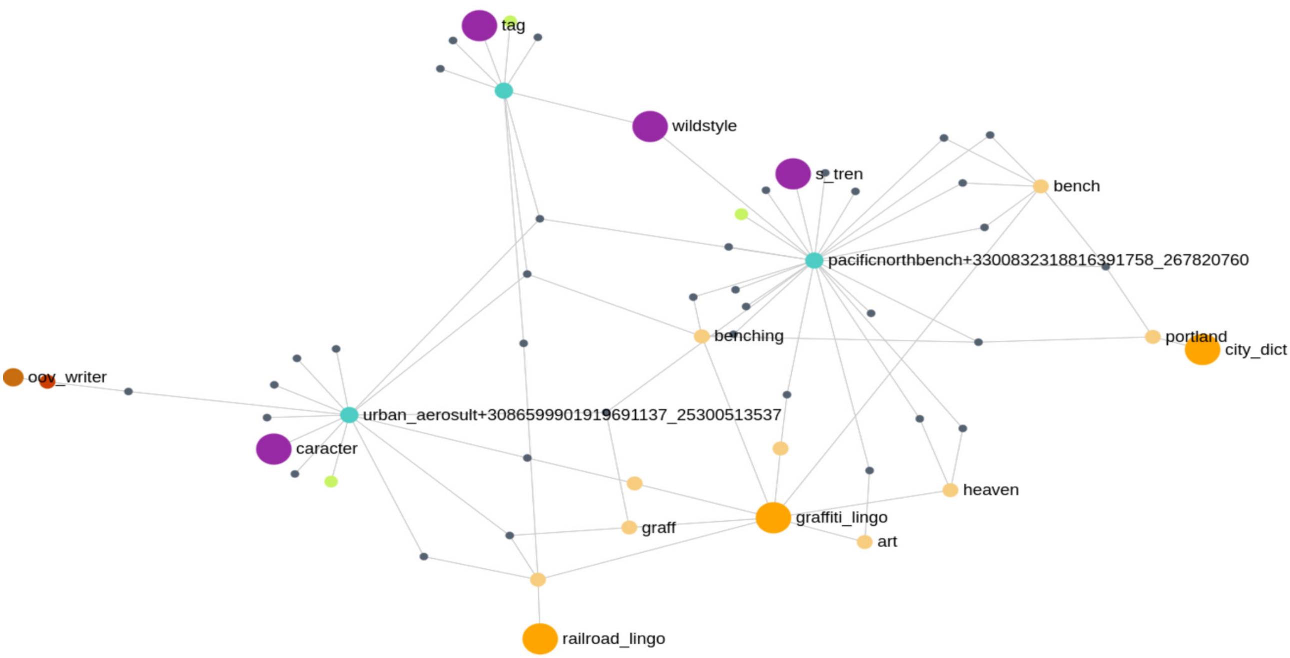

Figure 5: Example: #freightgraffiti with 0 mining depth and downloading 3 post’s for each hashtag with image and text inferences linked to hashtag or post’s nodes.

Following the segmentation, the find_word() function identifies relevant terms within the hashtags. This function is limited to specialized dictionaries, including a graffiti glossary and a dictionary of railway worker terms. By consulting these specific word lists, the function enhances the precision of tagging graffiti and railway-related terminology. For example, it identifies "Boxcar" as a term associated with railways and "Graffiti" as a term specific to the graffiti domain.

2.3.2 *Looking for graffiti entities*

The *graffitientitieslookup()* function is instrumental in identifying and categorizing graffiti entities within hashtags. During execution, the function goes through each token in the document, verifying if the hashtag culminates with keywords like "graffiti" or "crew". Then try to extract potential writer names or crew acronyms.

In cases where potential names or acronyms are not found in the list of recognized words (wordlist), they are designated as custom entities (iscustom_entity) a technique commonly known as Out of Vocabulary (OoV). Moreover, the function conducts assessments based on the token's length to discern if it could represent a writer's name or a team acronym. If the word tagged as graffiti entity is between 2 and 4 characters it will be marked as crew, otherwise if the length of the word is between 5 and 8, is under the writer category.

Using this approach, the function contributes to a more nuanced understanding of the graffiti culture and its associated entities within the context of social media platforms like Instagram. The inferences from both image analysis and NLP techniques are stored in the database. By organizing the data systematically and using the network analysis library Graphology (Plique, 2021) is possible to visualize correlations within freight train graffiti communities users and how they use significative hashtags.

3. Modeling and Interpretation of Hypertextual Conversations

In the mining section of this article (2) network graphs were presented (4.4.1, 4.3.2) that depict the outcomes of hypertextual conversation mining, and should be noted that these specific processes are originally stored in relational text database. And the visualization of these network graphs requires the construction of an object using Graphology syntax (Plique, 2021). Graphology is a JavaScript library that processes and calculates the spatial arrangement of nodes, their size based on centrality metrics or their Louvain community membership, among other graph theory algorithms.

This segment will discuss the process of network modeling, focusing on the types of nodes and the direction of their edges. Additionally, relevant centrality algorithms will be used to identify significant nodes within the conversation. The Louvain community algorithm, along with the forceAtlas2 spatial layout algorithm, will be employed to visualize node clusters.

It is noteworthy that these "nodes" correspond to users and their posts tagged with hashtags. Hashtag and post nodes are subjected to the inference of significant textual terms and graffiti types through a custom object detection model and natural language processing algorithms. Consequently, users are predominantly grouped around hypertextual tags frequently employing significant terms within the community and, to a lesser extent, around graffiti types found in their posts.

3.1 Network structure layout

Sociograms, proposed by Moreno (1934), have benefited greatly from mathematics and computer science in graph theory, but keep the core principle of linking individuals through common practices or interests. In hypertextual conversations, the most relevant node type is the hashtag, as both posts and text analysis inferences are linked to it.

It is essential to detail the syntax of nodes and links that structure this network, as this form repeats for each post and heavily influences graph morphology. The direction of edges significantly impacts the functioning of layout algorithms such as forceAtlas2 and the calculation of centrality measures.

3.1.1 Basic Structure: Users, Posts, Hashtags

The basic network structure consists of three types of nodes with specific attributes: 1) The user node comprises as attributes the username, post count, number of followings, number of followers, and profile privacy status. 2) The post node includes the Instagram post identifier, number of likes, number of comments, and the hashtag where it was found and backed up. 3) The hashtag node contains the count of posts tagged with that label. This information is relevant data

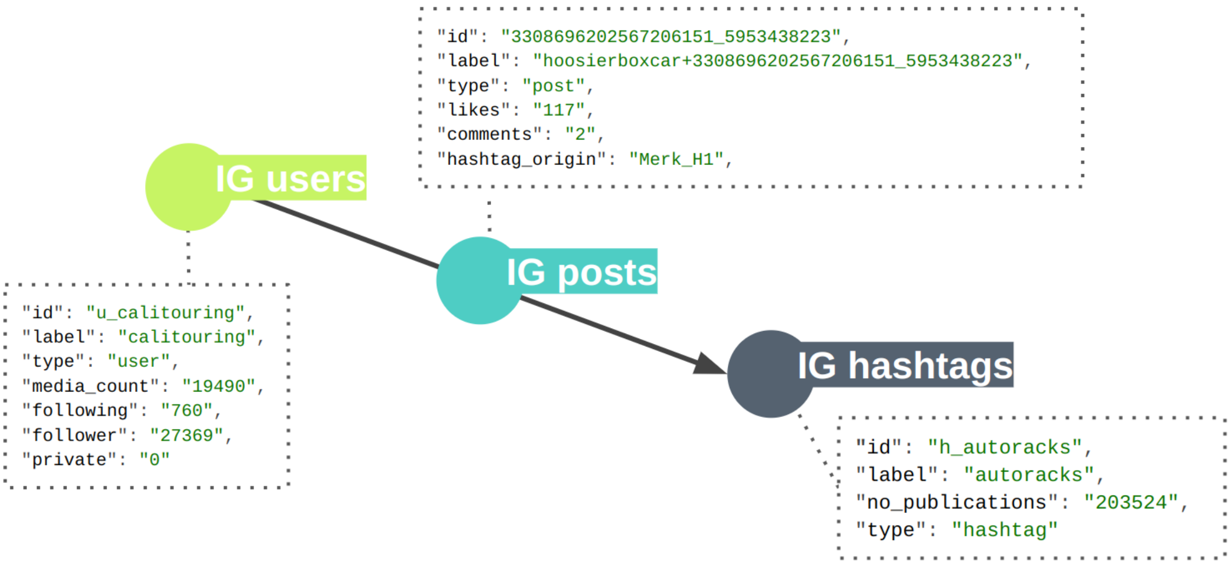

Figure 6: Basic graph morphology with mined attributes.

to make insights about how popular are users (followers), posts (likes and comments) and hashtags (community use). This metadata used as attributes is used to change the node size and may be understood as community driven data visualization

To construct this network, a list of all posts identified with the MUID is generated. Updated information for each author and previously mined hypertextual tags is added. Links are then established, first from the user to the post (user -> post), followed by extraction of all hashtags from the post caption. A link between the post and each hashtag is created (post -> hashtag).

3.1.2 Morphology with Inferences: Graffiti Types, Entities, and Dictionaries

The graph structure may include elements inferred from computational processes. The first step in the network construction with Graphology syntax consists in the creation of inferential nodes: 1) The graffiti type nodes corresponding to objects with which the convolutional object detection model was trained: tags, throw-ups (or bombs), roller, wildstyles, 3D, character and train identifier (s_tren). Then links are generated directly connecting the post to the graffiti type inference (post -> graffiti_ type). 2) Text analysis yields two major categories: entities identified through the out-of-vocabulary technique and terms found in dictionaries. Entities are linked first to the hashtag and then to subcategories (hashtag -> entity -> entity_category). Terms are linked from the hashtag to the identified term and then to the relevant dictionary (hashtag -> term -> dictionary).

3.2 Clustering is (almost) wrapping everything

The basic structure, augmented with inference nodes, is repeated for each mined publication, generating densely interconnected networks. For instance, the hypertextual conversation around the #Kosm hashtag, belonging to a mexican freight graffiti writer, consists of a staggering 7,687 nodes, including 748 'post' nodes, 510 'user' nodes, 6,167 'hashtag' nodes, 173 'ai_text_word' nodes, and 76 'entity_individual' nodes, intertwining impressively into a total of 30,102 links, where each node, irrespective of type, averages 3.9 links.

To identify node clusters within this hyperconnected graph, two strategies are applied. Firstly, applying a spatial layout algorithm like *ForceAtlas2*, which calculates gravitational force to determine node positions. Secondly, utilizing *Louvain communities*, this method clusters nodes based on edge density, optimizing network modularity over multiple passes,

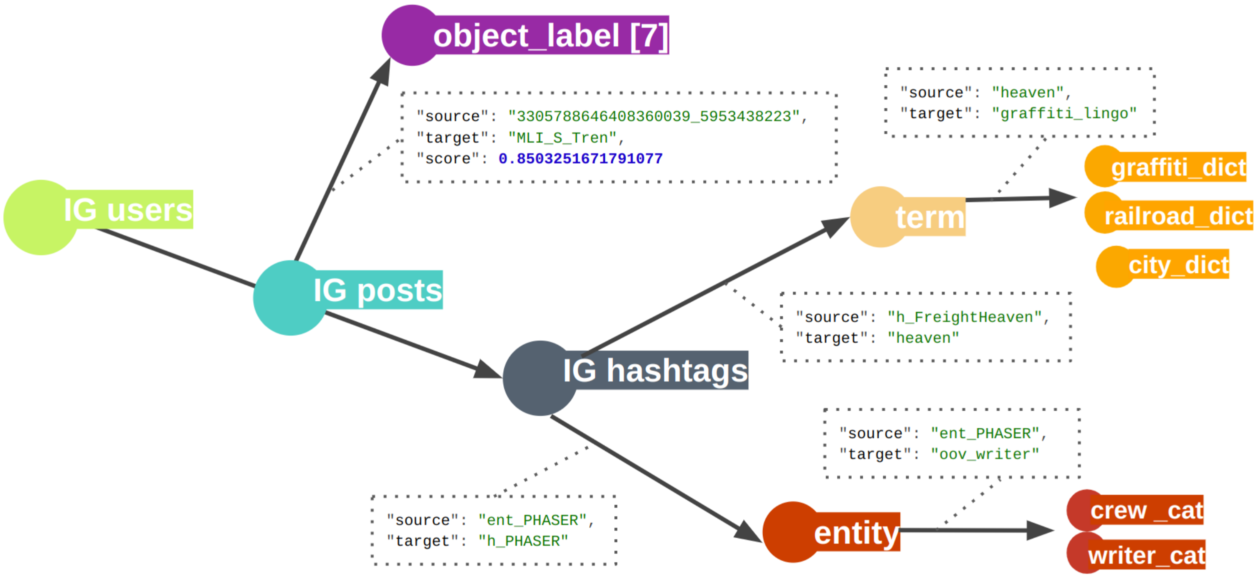

Figure 7: Inference graph morphology with edge direction.

with each pass refining the granularity of grouping. Through these techniques, it becomes feasible to visualize intensively linked node clusters, revealing intricate interconnection patterns not readily discernible otherwise. These clusters not only identify nodes (regardless of being users, hashtags, graffiti types, or terms) of heightened significance and symbolic value within the network but also aid in recognizing accompanying lateral terms helping the understanding of the social and cultural dynamics manifesting on this user-generated content platform.

In order to showcase these clustering methods, we are going to analyze the hypertextual conversation from the seed-node Kosm (#Kosm_1_hashtagTop_9_62568b82), using these algorithms the graph progresses from an initial unordered visualization to a clustered representation. Initially, the CirclePack algorithm is employed to generate a baseline visualization, arranging data elements in a compact, circular layout. This arrangement, while visually cohesive, does not prioritize or distinguish between the relationships, authorship or even hierarchies inherent to the data, thus serving primarily as a graphical representation of the dataset's volume and density.

Advancing our analytical precision, we employ the *ForceAtlas2* algorithm, a sophisticated network visualization technique based on physical simulations. This algorithm treats nodes as charged particles, exerting attractive or repulsive forces upon each other. The *ForceAtlas2* layout dynamically adjusts node positions based on the strength and density of their connections. This spatial redistribution allows for the emergent properties of the dataset—such as clustering of closely related nodes and dispersion of loosely connected ones—to be visually and analytically discerned.

To refine our analysis, we integrate the Louvain method for community detection. This algorithm identifies densely connected subgroups within the network, optimizing modularity by iteratively grouping nodes into communities based on the density of edges between them. Applying this method enables us to delineate the network into distinct communities or clusters, which we then visually encode using a color scheme. This step not only enhances the visualization's informational depth but also provides a clear, immediate insight into the social and thematic subdivisions within the hashtag network.

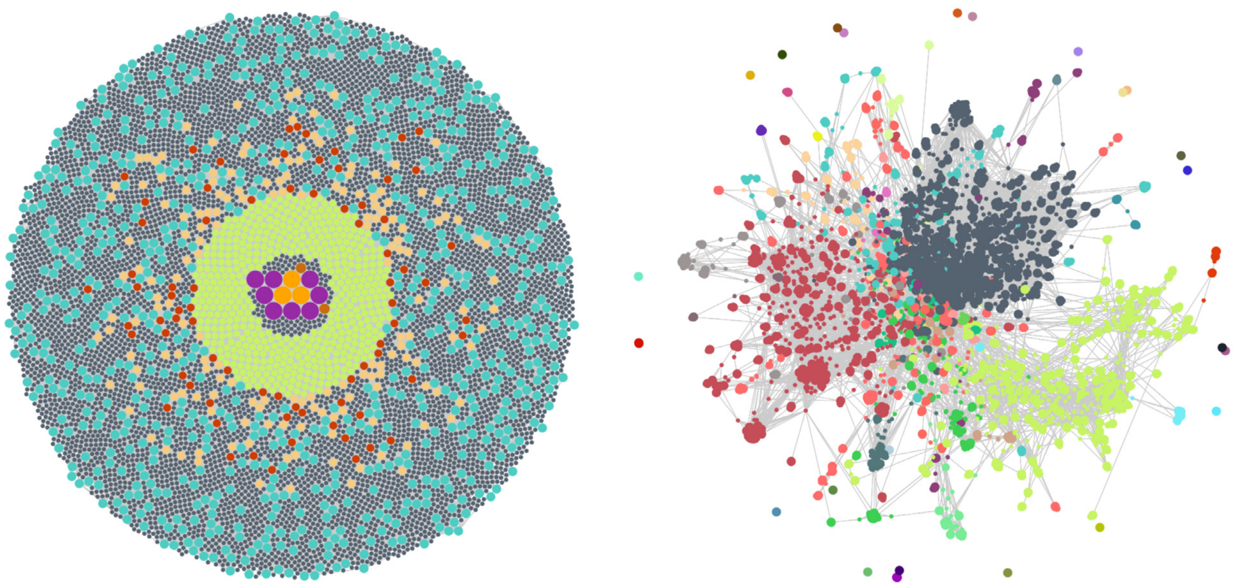

Figure 8: On the left, the initial circlePack layout is shown, while on the right, the spatial force atlas 2 layout is applied, and the colors of the nodes are changed to represent specific communities identified by the Louvain community detection algorithm.

In Figure 8 the network graph on the right, where *ForceAtlas2* has been applied and nodes are colored using the Louvain algorithm, three distinct clusters are evident. A blue cluster occupies the upper part, a crimson cluster is situated on the left, and finally, there is a lime green group. Each cluster represents a different semantic use of the term "kosm." The blue cluster comprises conversations related to freight train graffiti, to which the graffiti writer Kosm belongs. The red segment consists of nodes tagging their posts with spiritual and religious terms, while the green cluster includes a mix of Polish language posts, vegetarian groups, and makeup enthusiasts.

Polysemy is common in these hypertextual conversations, particularly when the mining depth is 1 or higher (indicating the mining of more hashtags and their associated posts). #Kosm exemplifies this phenomenon, but it's not an isolated case. For instance, the hashtag #HeavenSpot may incorporate religious or spiritual posts, while #Bench might feature gym workout-related content. It's important to note that thematic divergence in mining is not an error of idmb, as the function of this mining robot is to indiscriminately back up tagged posts to build a systematic dataset, sort of a big temporal snapshot of the user-generated content. Although it's not possible to avoid mining these posts, the network graph can be refined to retain only those containing elements relevant to the study object of interest through inference reduction, as detailed in segment 4 of this text.

Before describing the functions of node reduction, it is pertinent to explain the phenomenon of polysemic divergence using the illustrative example of the hashtag #Kosm. This hashtag is associated with a Mexican freight train graffiti artist renowned for his freight end-to-end rollers. To clear up the semantic variations of #Kosm, a NeighborsNeighbors filter is employed. This filter entails compiling a list of first-order neighbor nodes connected to #Kosm, along with the second-order neighbors of these directly linked nodes. The left graph in Figure 9 depicts this NeighborsNeighbors filter on the whole network graph shown previously, enabling the identification of specific users, hashtags, and posts linked to #Kosm.

The analysis reveals distinct semantic associations attributed to #Kosm within sociodigital realms. Firstly, the religious/spiritual community, depicted in red, interprets #Kosm in relation to "Cosmos" and employs terms like "spirit," "believe," and "faith." Notably, @Kosm_World emerges as the one and only user within this second-order neighbor list. Conversely, the lime green cluster signifies the use of #Kosm as an abbreviation for the Polish word "kosmetyce".

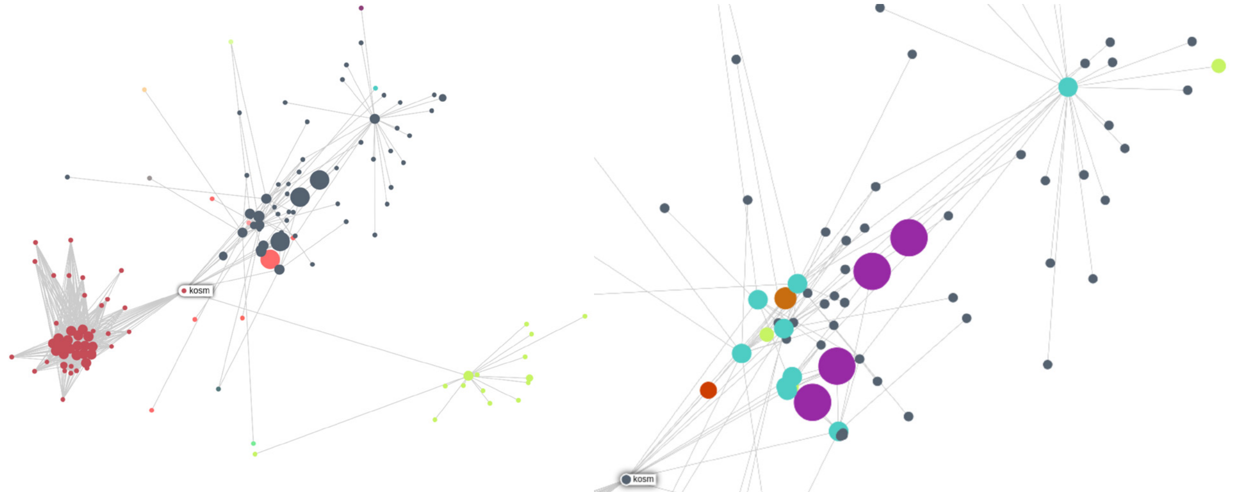

Figure 9: On the left, NeighborsNeighbors of the hashtag node "Kosm" with forceAtlas2 layout applied and Louvain community color scheme.While on the righ, the same NeighborsNeighbors node zoomed to graffiti cluster.

Here, divergence is evident, driven by @happyrabbit_blog, who tags a single post with #Kosmeytic, #CrueltyFree, #WeganskieKosmetyk, or #Weganskie. Remarkably, posts within the second ring of depth (1) are preserved, with #Kosm serving as seed node, thereby fostering polysemic divergence and engendering thematic subnetworks within the hypertextual conversation.

Moreover, the blue cluster denotes the network associated with the graffiti writer (Figure 9, right graph), offering insights into the utilization of #Kosm within this context. A closer examination of the graph, filtered by NeighborsNeighbors and color-coded by node type, reveals the involvement of three users such as @__kosm, @lineofsights__, and @jrb1067, that tag publications with #Kosm as a hashtag. Also, in some posts tagged the object detection machine learning model identified train identifiers, wildstyle, roller, and character. Finally, #Kosm was also identified as graffiti writer's name by the sPacy custom algorithm.

Through the NeighborsNeighbors filter, computational methods effectively discern symbolic content within tagged node clusters. While these methods may not be perfect, they provide evidence that a cluster of nodes associated with specific hashtags holds symbolic content pertinent to conclude its affiliation with the freight graffiti community.

3.3 Node size by centrality metrics

Finally, with the Graphology library, researchers and analysts can compute essential centrality measures such as degree centrality, betweenness centrality, and PageRank.

Degree centrality, a fundamental metric in network analysis, is calculated by counting the number of connections or edges each node has within the network. This measure provides insights into the importance of nodes based solely on the quantity of ledges they possess (Smith & Jones, 2010). Moving on to betweenness centrality, this measure quantifies the extent to which a node serves as a bridge or intermediary along the shortest paths between other nodes within the network. Nodes with high betweenness centrality play a crucial role in facilitating efficient communication and information flow between disparate parts of the network (Johnson et al., 2015)

Normalization involves adjusting the centrality values to ensure consistency and comparability across different networks and scales. This step is crucial because centrality values can vary significantly depending on factors such as the size and density of the network. By normalizing centrality values, researchers can effectively compare the relative importance of nodes across different networks and derive

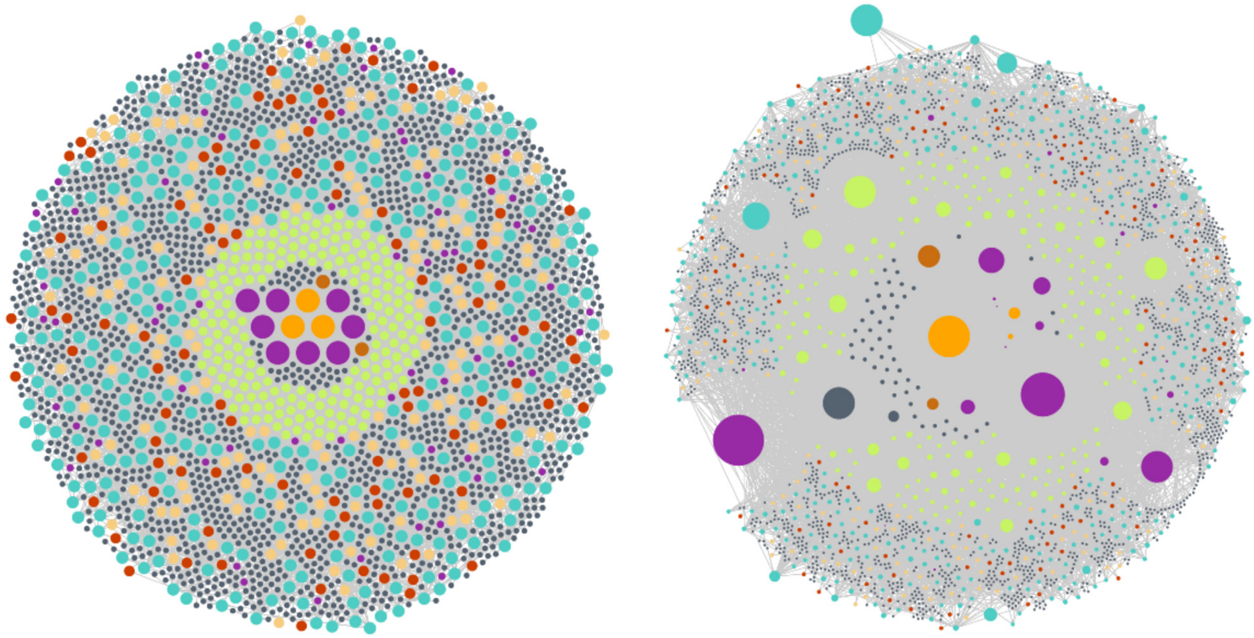

Figure 10: On the right, graph with circlePack layout and fixed node type size. On the left, the node size is defined by the degree centrality metric.

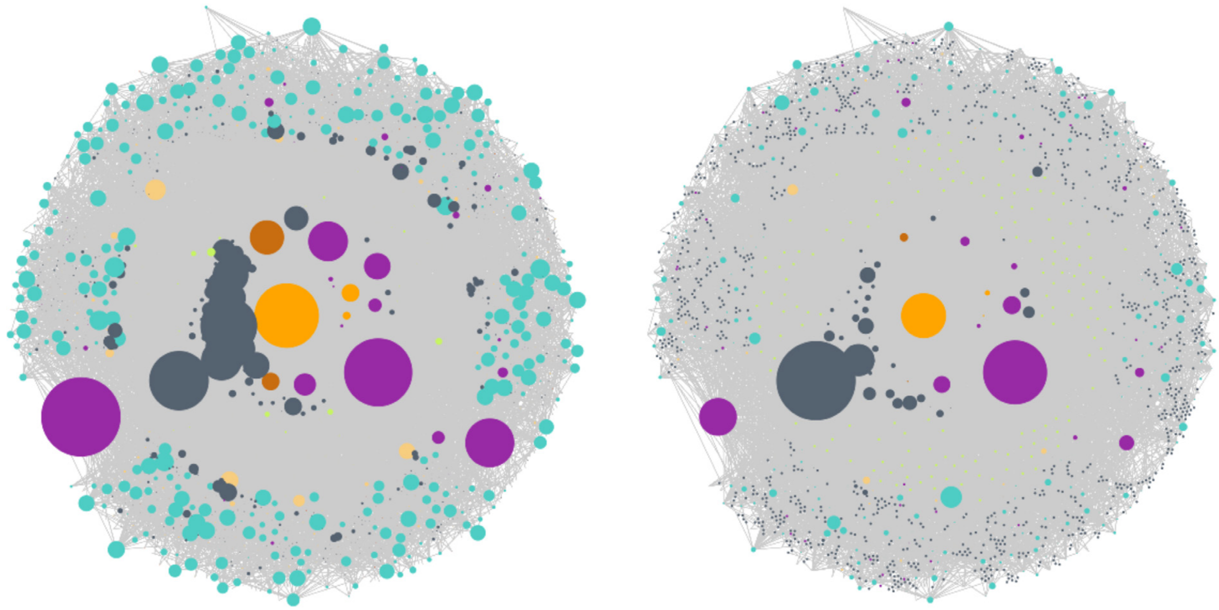

Figure 11: On the left, the node size is defined by degree centrality. On the right, the node size is based on their betweenness centrality.

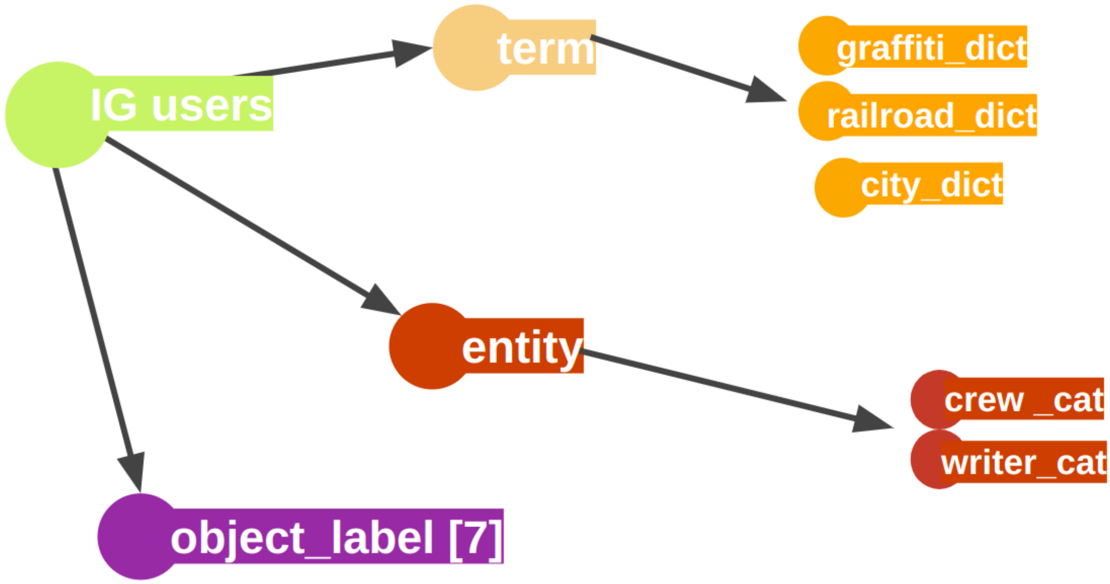

Figure 12: Node reduction graph morphology.

meaningful insights into their structural significance (Adams & Brown, 2012). In this exercise, the normalization process plays a crucial role in ensuring coherence across graphs obtained from different depths of mining. When mining at depths of 0, the resulting graphs are relatively small, but as the depth increases to 1 or 2, the mining process grows exponentially, leading to massive networks. Normalization ensures that the node sizes remain consistent across increasing depths of mining.

For example, in Figure 10, the graph is presented with node sizes representing different metrics: node type and degree centrality . On the left node sizes are fixed by node type, there is no further insight, just the uniform representation of the hypertextual conversation. On the right, node sizes are scaled according to their popularity on Instagram, offering insights into the nodes' influence within the social media platform.

In Figure 11, On the left graph, the node sizes are defined based on their degree centrality, bigger nodes means they are highly connected to first order neighbors. The right graph illustrates node sizes based on betweenness centrality, which measures the nodes' importance in connecting different part s of the network; the bigger ones can be interpreted as relevant bridge nodes.

4. Node reduction

In our research, inference reduction is essential for simplifying the network graph and focusing exclusively on nodes and connections relevant to freight graffiti.

This process extracts meaningful information related to our study's focal point. Initially, we use a custom object detection model (explained in segment 2.2) to identify posts featuring specific graffiti types or train identifiers. Then, we link each post's author to the visual content detected. Additionally, we identify terms within post hashtags, such as city names, graffiti-related words, and railway references. These terms are directly linked to the respective post authors. Furthermore, we employ Out-of-Vocabulary technique to identify entities associated with graffiti writers and crews within the hashtags (as discussed in segment 2.3). This enhances our understanding of community conformation, revealing shared symbolic terms and providing deeper insights into participants.

In conclusion, the inference reduction process involves removing less relevant posts and hashtags, refining the network's morphology. Figure 12 visually represents this focused graph morphology, shaped by pertinent content extracted from dictionaries, entity inferences, and the custom object detection model.

4.2 Example of node reduction

Before the inference application, the network (Figure 13, left) consisted of a large number of nodes, totaling 18,089. These nodes included 1,860 posts, 1,226 users, and a significant number of 14,067 hashtags. Additionally, there were nodes for other data types, such as 8 for custom inferences, 2 for entity subcategories, 353 for text words, 3 for hashtag classes, 412 for individual entities, and 158 for world inferences.

After the inference application, there was a noticeable reduction in the network size (Figure 13, right) , resulting in a total of 1,732 nodes. Interestingly, there were no nodes representing posts or hashtags remaining after the inference process. However, the majority of the remaining nodes were users, totaling 954. The counts for nodes related to custom inferences, entity subcategories, text words, and individual entities remained relatively consistent, with 8, 2, 353, and 412 nodes, respectively.

With the inference reduction, we can identify users and the symbolic content of their posts, allowing us to understand the links between symbolic content. This shift not only minimizes the noise within the hypertextual conversation but also, through first and second-order neighborhood filters (Neighbors & NeighborsNeighbors), enables us to discern how users are connected on the platform using symbolically relevant textual terms or identified entities. By eliminating intermediate steps (posts and hashtags), we gain a clearer insight into the connections between users using shared symbolic content.

5. Tool application in three examples

Up to now, the focus of this paper has been on delineating the processes leading to the visualization of hypertextual networks on Instagram, particularly examining the case of freight train graffiti. This segment will showcase three conversations in order to enlist capabilities and limits of visualization obtained by this tool. The first example is #FreightGraffiti conversation, with a mining depth of 2, that leads to an expansive data exploration endeavor; secondly, the #Kosm dialogue undergoes examination, especially in interpreting outcomes after node reduction filter that deletes most of the divergence due polysemy; lastly, #PortlandBench underscores a conversation anchored in a locality, aimed at discerning writers and regional symbolic content, in addition inter-locality relations are supposed to appear.

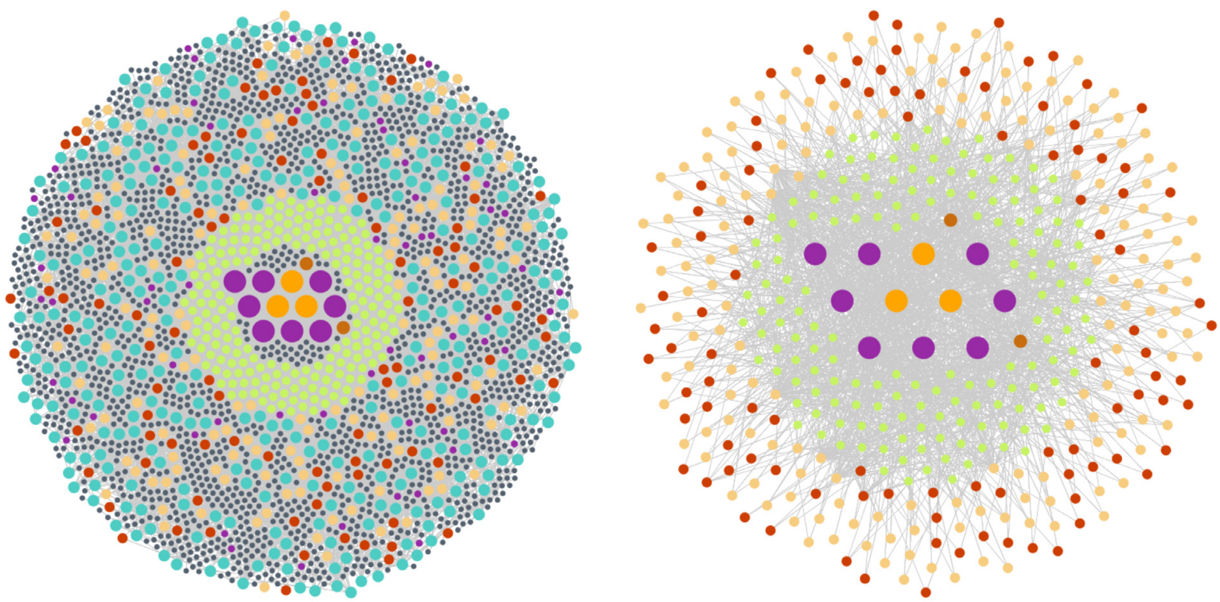

Figure 13: On the left, the initial graph has 18,089 nodes. On the right, the same graph after inference reduction was applied, resulting in 1.732 nodes.

Until now, the paper unfolds with a hard methodological-technical orientation; yet, it's imperative to acknowledge its theoretical underpinning, even though brief. The interpretation leverages Figueroa's triadic concept of self-annunciation (2014), comprising "*yo-soy* (I am), *yo estuve aquí* (I was here), and *yo existo* (I exist)." This Spanish author projects these announcements as graffiti writers' motivational catalysts to mark cities, trains, and, conceivably, any medium, to broadcast their presence. Specifically, in this context, the operational adoption of this self-annunciation triad influences the inferences, with the *yo-soy* attributing to entities—e.g., "I am Kosm" upon tagging the rolling stock and marking my post with the self-referential hashtag #Kosm. The *yo existo*, slightly modified, correlates to a symbolic community dimension, i.e., "I exist through the collective usage of specific hashtags, serving as convergence points within a particular community," interpretable as an existence within the graffiti realm through familiarity with terminologies, styles, and shared knowledge. On a physical level, *yo-existo* manifests when marking names following certain conventions, and digitally in this study, when posting materials on appropriate channels, using community hashtags, identified via graffiti and railway glossaries. Lastly, *yo-estuve-aquí* pertains to conquering intervened spaces physically; yet, within hypertextual conversations, it relates more to localities tagged of pieces, bombs, or tags that writers or benchers have published on Instagram

5.1 #FreightGraffiti

This example is the most straightforward of the three presented in this paper, focusing on a communal label that describes the activity being analyzed. With over 600,000 posts tagged on Instagram, this label serves as a mandatory meeting point, facilitating the identification of top-rated writers and Instagram profiles within the community. Beyond merely highlighting popular elements, it offers insights into lesser-known hypertextual tags and subsets of node groupings, including users, posts, and hashtags, all contributing to the overarching phenomenon under examination.

Initially, the graph in this example is substantial, consisting of 18,089 nodes and generating 89,044 edges, indicative of the extensive nature of the hypertextual network. However, after applying node reduction by inference, the network experiences a significant reduction to 1,732 nodes. Notably, despite this numerical reduction, the practical impact is most apparent in the removal of 14,067 hashtags and 1,860 posts, while only 272 user nodes are eliminated. This selective reduction underscores the primary objective of node reduction: to establish links between content authors and the symbolic content deemed relevant to freight graffiti, guided by the concept of self-annunciation. Therefore,

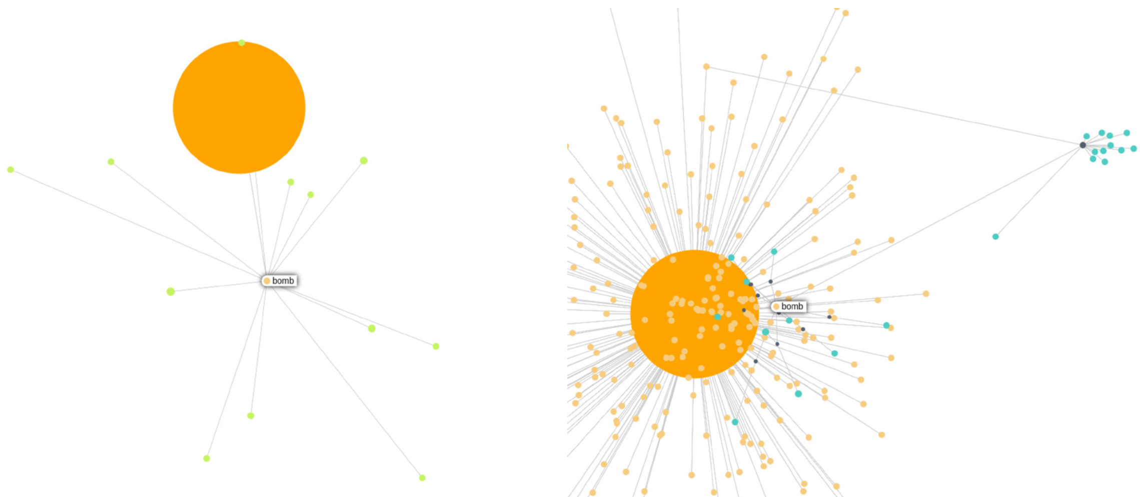

Figure 14: On the left, first neighbors of "bomb", after inference reduction was applied. On the right, the same node is highlighted but before node reduction.

the preservation of a substantial number of user nodes post-reduction suggests that the divergence within this conversation is relatively low.

In this particular case, node reduction, I believe, results in both positive and negative outcomes, though I think the negatives outweigh the positives. Positively, the reduction of nodes speeds up the calculation of centralities and the application of the *forceAtlas2* algorithm, which significantly quickens the process of analyzing a network with so many nodes. By linking users directly with the found terms and these terms with the dictionaries they belong to (graffiti and railroad), we can identify which particular users use certain terms. In this context, it's documented that 879 users use 257 words from the graffiti glossary. Applying degree centrality highlights the most used terms and their authors. And through the neighbors' filter, it's presumed we could identify groupings of users and terms. However, this presumption has failed since, in this example, the community is so intertwined that in practical terms everything relates to each other, difficulting the interpretation.

Another example of the synthesis achieved from this process can be seen in the left graph of Figure 14 (left), where the neighbors of the term "bomb" are filtered, showing through node reduction by inference that users who use this specific term in their hashtags are @berlinstreetsgraffiti, @bad_ridm, @duesselspotter, @samfactory9, @cbscrew, @lettering.kings, @__p.funk, @pointgraff_rennes, @graffiti_sachsenanhalt, @graff.hunterr, @rakugaki.dame.zettai, @handstyles.worldwide. Although we don't know the hashtag from which the term was extracted, we have the link with the users and with the writers or crews they also tagged with this term, among which can be highlighted the writers: Mecri, Trio, Skaf, Powder, Mesy, Kois, or Ernst and the crews Kog, Lts, Kgs, Swv, Puc. Finally, this is the goal of node reduction by inference: to link relevant symbolic terms in two steps with types of graffiti found, and entities. However, in this case, the connection is so dense that it appears to be a large, hyperconnected, and obscured network.

In this particular case, the interpretation is richer in symbolic terms without the reduction by inference. This can be observed in the right side of Figure 14 (right) where the neighbors of neighbors of "bomb" are filtered, and we can identify which lateral terms are lost either because they were not included in the glossary or because they become insignificant after applying degree centrality to the node sizes, among which are listed: #trainbomb, #graffitbomb, #bombedepeinture, #citybombsquad, #cbscrewbombs, #graffitibomb, #bombesdepeinture, #droppingbombs, #streetbomb, #stickerbomb, #bomb. In this particular case, the breakdown of a not so popular term can be identified by at least one term that needs to be added to the graffiti glossary ("dropping").

5.2 #Kosm

This hypertext conversation example starts with a graffiti writer's name as the mining seed node. This instance is chosen due to the pronounced divergence arising from polysemy. This is a great representative case to validate the efficacy of node reduction through inference.

Upon applying the filter in the prior example (#FreightGraffiti with extensive mining depth), it was observed that inference-based reduction proved effective, despite the survival of a substantial number of users. In cases where polysemy is prevalent from the outset, it becomes imperative to minimize those users and posts deviating from the object study topic. This conjecture shows to be true upon the application of the filter; out of the 510 initial users constituting the graph, only 266 survived. This figure could potentially be lower if the graffiti glossary did not encompass terms such as "heaven," "devil," "fly," "inspiration," "king," "angels," and "star." This is a big challenge as these terms hold relevance within the community, facilitating semiotic analysis of graffiti writers or benchers' posts, yet they complicate interpretation by retaining users and terms not aligned with our thematic interest. Nonetheless, the number of users linked to these fuzzy terms are minimal.

Two distinct forms of self-announcement in the socio-digital dimension are proposed. On one hand, there is self-promotion, where the writer himself tags posts on Instagram where his activity is documented. This digital mark aims to organize the writer's published material under hashtags like #Kosm, and to be part of community tags with a higher level of dissemination such as #FreightGraffiti, #fr8porn, #Benching. On the other hand, there is relay-transmission, where the "I-am" is recomposed into "He-is" and "He-was-where". This broadcasting in the announcement is crucial for this phenomenon. And more importantly, this can be observed in the data.

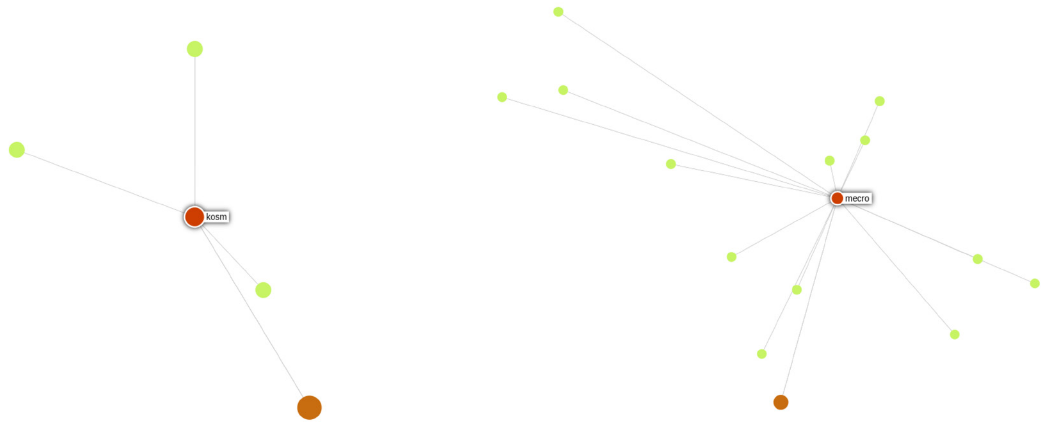

Figure 15: On the left, first neighbors of "Kosm" entity node, after inference reduction was applied. On the right, first neighbors of "Mecro" entity, one of the more benched graffiti writers in this specific conversation.

On the left side of Figure 15, it is documented, in the first-order neighbor filter, how the entity "kosm" is linked to three users who have posted and tagged this writer's interventions. On one side, the writer himself in his profile @__kosm, and two benchers @jrb1067 and @linesofsight. On the right graph of the same figure, the network of Instagram profiles that posted and tagged the pieces by Mecro can be observed. This broadcasting is also a way to value writers within the community; it can be asserted that the greater the broadcasting of graffiti writers, the higher their valuation within the community. Thus, to conclude this example, the top 5 writers with the highest number of records are listed: Ichabod (19), Mecro (13), Visah (12), Renik (11), and Aser (10).

5.2 #PortlandBench

In the final example of this article, we will revisit a geographical hashtag, #PortlandBench, a tag that has been recurrent in the explorations made in general character conversations such as #FreightGraffiti or #fr8porn. Exploring this from a geographical perspective is fascinating because it allows the identification of users and writers documented (benched) in a certain city. A practice that lends meaning and promotes participation in this community, namely the assurance that someone else knows Kosm was there, motivates the writer to continue decorating trains.

The reduction through inference in this type of node has similar results to dictionary term nodes since they share the same edge morphology. It is important to note that the lateral terms accompanying the principal word "Portland"—which, in this particular case, are bench, benching, tattoo, graffiti, murals, art, graff—are lost. This limits the ability to segment posts related to each subset of nodes. However, it benefits the connection to entities, significant terms, and the authors who have used them.

In this context, we can identify using the geographic node "Portland" at least three Instagram benchers who hold significant relevance in this geographically nuanced discussion, but who also have surfaced in other explorations: @pacificnorthbench, @oddiophoto, @micah_hawaii, and to a lesser extent, @cnwneverdies, @gooseonacruise, @paintspotting_pdx. Through a process of inferential reduction, using the geographic node "Portland" as a reference filtered by the second-order neighbors network (NeighborsNeighbors) list (left side of Figure 16), we detect: 9 users, 5 graffiti styles including wildstyle, character, tags, rollers, throwups, and train signals, in addition to 26 entities among which writers like Mecro, Visah, Enron, Kwaz, Shelto, Oaph, Dyse, and crews such as VRS, ATD, 925, CFS, ATD emerge. Moreover, we've identified symbolic terms classified as Wholecar, Boxcars, Benching, Train, Spotting, Painted, Throwies, Heaven and Fr8.

Figure 16: On the left, neighborsNeighbors of the geographic node "Portland".- On the right, the geographic category node and its second order neighbors.

This outcome allows for the identification of small geographic segments within a broader symbolic circuit underpinned by the freight train's moving nature. It is worth noting that the geographic node "Portland" represents merely one of the 33 geographic nodes conforming this hypertextual conversation, which originally was mined from the hashtag #PortlandBenching as a seed node (right side of Figure 16). This list encompasses cities across the North America region, such as Portland, Norfolk, Chicago, Seattle, Minneapolis, Winnipeg, Riverside, Atlanta, Philadelphia, Oakland, Vancouver and Oaxaca, among the most notable. Each of these geographic tags presupposes active labels that correlate with the graffiti phenomenon, such as #OaxacaBenching or #WinnipegGraffiti, within which we can identify the most significant entities and symbolic terms.

5. Conclusion

In summary, this paper has shown how computer techniques can help us understand conversations on Instagram, especially among freight train graffiti writers and benchers. We used machine learning and computational methods to back up and actually visualize how freight graffiti writers and benchers interact online. We simplified the network by reducing the number of nodes and links, making it easier to analyze. The process of node reduction has proved instrumental in simplifying the network structure, facilitating closer connections between nodes while enhancing analysis efficiency. However, it's important to acknowledge the loss of significant symbolic material during this process, highlighting the need for careful consideration of methodological decisions.

By addressing the challenge of polysemy, exemplified by the case of Kosm, our study has advanced thematic precision within the network analysis. This strategic filtering approach ensures that our focus remains on genuinely relevant users and content, contributing to a more refined understanding of community dynamics. After simplifying the network, we found it easier to see connections between users and the symbolic content they share.

6. References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., & Zheng, X. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*.

Adams, A., & Brown, B. (2012). Normalization techniques for network centrality measures. *Journal of Network Analysis*, 8(3), 123-135.

Figueroa, F. (2014). *El grafiti de firma: un recorrido histórico-social por el grafiti de ayer y hoy*. Minobitia.

Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). *spaCy: Industrial-strength Natural Language Processing in Python*. doi: 10.5281/zenodo.1212303

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLoS ONE*, 9(6), e98679. https://doi.org/10.1371/journal.pone.0098679

Johnson, R., Smith, T., & Davis, M. (2015). Exploring betweenness centrality in social networks. *Social Network Analysis*, 15(2), 45-58.

Moreno, J. (1934). *Who Shall Survive? A New Approach to the Problem of Human Interrelation*s. Nervous and Mental Disease Publishing Co.

Page, L., & Brin, S. (1998). *The PageRank citation ranking: Bringing order to the web*. Stanford InfoLab.

Plique, G. (2021). *Graphology, a robust and multipurpose Graph object for JavaScript*. Zenodo. https://doi.org/10.5281/zenodo.5681257

Smith, J., & Jones, R. (2010). Understanding degree centrality in network analysis. *Journal of Computational Sociology*, 5(1), 32-45.