Big data visualization through the lens of Peirce's visual sign theory

BY: **Alon Friedman** and **Martin Thellefsen**

ABSTRACT

Data from social media platforms, such as Twitter and Facebook, are generated by people who produce, spread, share, or exchange multimedia content. Such content may include text, images, sounds, or videos. To derive insight into the behavior of social media users, researchers often use open-source technologies to visualize data and generate models for data analytics. One of the most popular open-source applications for managing and analyzing social media data is the open-source R programming language. Friedman and Feichtinger (2017) created an R package termed 'Peirce's sign theory R package' to analyze data using Peirce's principles of discovery. Though Peirce semiotics have been introduced in the context of computer programming languages, so far, no previous work has applied Peirce's sign theory to data modelling of social media data. In this paper, we use Peirce's sign theory R package as an overall framework to gain insight into data collected from Twitter. We assembled the data using Twitter's Analytics algorithm, examined the relationships between variables, and visualized the results. Subsequently, we assessed the feasibility of analyzing those graphics using the triadic model set out by Jappy (2013) and Peirtarinen (2012) for the interpretation of visual signs. The study results showed that Peirce's sign theory R package effectively analyzes and visualizes Big Data from social media feeds. However, due to complexities in both the social media data feeds and Peirce's interpretation of meaning, as outlined by Jappy (2013) and Peirtarinen (2012), we were unable to develop algorithms that generate or suggest an interpretation of visual signs.

ARTICLE INFO:

Volume: **08**

Issue: **01**

Summer **2022**

ISSN: **2459-2943**

DOI: **10.18680/hss.2022.0007**

Pages: **115-136**

KEYWORDS:

Semiotics

Peirce's sign theory R package

Visual Sign

lOpen-source R

Twitter Analytics

Algorithm

Punctum. International Journal of Semiotics | **08:**01:2022
ISSN 2459-2943 | DOI: 10.18680/hss.2022.0007 | **punctum.gr**

A. Introduction

Every day, multiple platforms tracking human and machine activities contribute to ever-growing digital data sets. Those data sets include a vast amount of unstructured data that are hard to sort or analyze and are commonly referred to as 'Big Data.' As the number of data-driven devices used by society has increased, the definition of Big Data has evolved. However, the core elements that characterize Big Data are its volume, variety, and velocity (e.g., De Mauro et al. 2015). Big Data is vast in volume and is generated at a far higher rate than other, traditional data. In addition, those data sets are diverse in form. For example, textual data produced by users of social media platforms differ from that produced by tracking mobile devices such as electric vehicles. Therefore, it is necessary to differentiate between data generated by human activity, e.g., human actors that post about different aspects of personal interests to an audience, and machine bots or algorithms that generate data and target users. The two leading social media platforms are Twitter and Facebook. Twitter has approximately 217 million daily users, and the platform allows its users to post and interact with messages known as "tweets." Researchers who study Twitter data often refer to it as Big Data due to the large amount of data generated by human and machine activities (Boyd and Crawford, 2011). Human-generated data is based on humans deliberately interacting on social media platforms e.g., by posting comments or by liking posts they encounter on the platform. Bots on the other hand are software that produces 'comments' or generates relationships by means of algorithmic processing. Human interaction is intentional and deliberate, while the latter is governed by machine automation, also known as bot algorithms, without much human interaction. In terms of machine-generated Twitter data, Kwak et al. (2010) concluded that the quality of machine-generated Twitter data is subjected to the quality of input. As such, low quality input suggests a low-quality output i.e., garbage in and garbage out, where it is difficult to analyze or identify the meaning of those comments. Nevertheless, machine-generated data have become increasingly powerful and influential on social media platforms and are perceived and retweeted by human actors (Lokot and Diakopoulos, 2016).

The developments in AI (Artificial Intelligence) and sophisticated algorithmic information processing pose new challenges for users to distinguish between the human tweet and bot tweet and call for new methodologies and countermeasures to hinder the spread of fake tweets (Fagni et al., 2021). To date, researchers have employed a variety of theoretical frameworks and methodologies to analyze the text format and to identify the emotional state and personality of users found on Twitter. In today's technology environments, semiotics scholars often discuss C. S. Peirce's contributions to the field of visualization (e.g., Farias and Queiroz 2017). While the semiotic framework continues to contribute to our understanding of social media, several studies require the development of software to match the semiotic regimes (e.g., Poulsen et al. 2018; Djonov and van Leeuwen 2018).

Whereas Onursoy (2015), Moschini (2018), and Poulsen and Kvåle (2018) have used a semiotic framework to study social media users' content, thus far, no previous research has analyzed and visualized social media data using Peirce's sign theory and its triadic model.

The Peirce sign theory is known as the action of signs, a process in which a sign is both affected by the object and affects one's mind, thereby producing the meaning of the sign (Strand, 2005). In this paper, we used the semiotic approach to analyze how tweets function as signs, and we are especially focused on how data generated by tweets form visual structures when processed by an algorithm. Visual semiotics is a subfield of semiotics with a particular focus on how visual expressions (e.g., pictures, models, or other kinds of visual representations) function as signs of meaning. Jappy (2013), who investigated Peirce contribution to visual communication, suggests that visual semiotics in principle is concerned with "visual culture “, a culture that is experiencing a tremendous growth and use in the development and deployment of technology. Our culture is increasingly shaped by social media today and has a direct impact on the way we live (e.g., Van Dijck 2013). Social media, in addition to influencing the way we perceive and consume information, contributes to a sense of participation in society because of the way we interact with likes, shares, and comments. In this information environment Twitter plays an important role and in this context data visualization becomes paramount. It enables researchers to investigate how popular events and topics are related, which trends emerge, the frequency of numbers of tweets and re-tweets, etc. In modern data analytics, visualization is an important tool for communicating complex data structures in a human perceptive fashion. Several approaches to visual semiotics have been developed that emphasize the modalities of visual expressions, and how it communicated a certain meaning (e.g., Floch and Pinson, 2001) and (e.g., Barthes, 1977). However, according to Aiello (2020), these studies have typically analyzed visual expressions from the perspective of advertising or art.

In this study we employ Peirce's semiotic framework both as an algorithmic tool and as a theoretical approach to data visualization. The latter is based on Peirce's own formulations of visual signs:

> “Hypoicons may be roughly divided according to the mode of Firstness of which they partake. Those which partake of simple qualities, or First Firstnesses, are ***images***; those which represent the relations, mainly dyadic, or so regarded, of the parts of one thing by analogous relations in their own parts, are ***diagrams***; those which represent the representative character of a representamen by representing a parallelism in something else, are ***metaphors***.” (CP 2.276, 1903)[1]

[1] We are referring to Charles S. Peirce. Collected Papers (Harthshorne, Weiss and Burks (eds.) (1932-1958), according to the scholarly tradition, CP, volume and paragraph number, year.

Punctum. International Journal of Semiotics | **08:**01:2022
ISSN 2459-2943 | DOI: 10.18680/hss.2022.0007 | **punctum.gr**

The hypoicons cannot be understood apart from Peirce's general semiotics and sign classes, see (Farias & Queiros, 2003, 2006). Also, the diagram is an essential part of Peirce's diagrammatic reasoning, or moving pictures of thought, see (Stjernfelt, 2007). However, our adaptation of visual semiotics and the triadic data model is mainly based on Jappy (2013) and Peirtarien (2012) interpretations.

B. Peirce's sign theory R package through open-source R

Charles Sanders Peirce was an American philosopher and considered one of the founding fathers of contemporary semiotics. Peirce's theory relies on a key tenet of semiotics: that a sign can have a meaning other than its own. In Peirce's theory, any object can be a sign, as long as it is understood as referring to, representing, or representing something other than itself (CP 2.302, 1895). Jappy (2013) investigated in detail how the sign functions in terms of the sign's constituent parts and used this triadic structure to analyze the function of the hypoicon. While Peirce did not provide a particular visual model of the sign, and there are different ways of representing the triadic structure of the sign, the most dominant ones are the triangle that put the sign, the object, and the interpretant at the corners of the model. However, the problem with this model is that it in principle is dyadic, only allowing a connection between two sides of the triangle. Another model preferred by many Peirce scholars is the fork model showing how the sign is genuine a triadic structure determined by its constituent and irreducible parts. Jappy presented a similar triangle model and re-introduced a horizontal line similar to Saussure (1959) separating the sign from the object (or signifier from signified). This was criticized by Nöth and Jungk (2015) that found several flaws in Jappy's visual interpretation. Among these flaws, they argued that the triangular model presented by Jappy lacks the connections between the three constituents of the sign. However, in this study, we find these flaws of minor importance in relation to our algorithm's application of the triadic principles.

With the growth of social media, we are witnessing a new understanding of the content we create, share, and exchange on various platforms. Despite ongoing debate about how to define the term 'social media,' most authors agree that social media technologies on smartphones and tablets create interactive platforms through which individuals, communities, and organizations may share, co-create, discuss, and modify user-generated content (Correia et al., 2009). Another aspect of social media is the technology behind these platforms. These technologies often rely on open-source applications and cloud computing to deliver the social media experience. An ongoing challenge for researchers and developers is how to handle the data generated by social media applications. Tonidandel et al. (2018) argued that traditional statistical methodologies fail to provide meaningful insight into the data and acknowledge that new methodologies are needed to analyze those unique data sets. Hochman (2014) examined the nature of social media images by examining the structure and implementation of those images that are being formed. He reported that images generated by data communities are automated and tagged using steam processes that shaped the layout of the social media images. The growing use of the automation of data analysis that is generated by advanced statistical calculations and adopted in social media research has led to an 'explosion' of software tools for scraping and analyzing social media content (e.g., Manovich, 2005). R is one of the leading open-source software environments, where its members of the community can develop specific packages that collect code, data, and instructions to solve problems beyond the reach of traditional spreadsheet software, and much of this work focuses on social media. Today, more than 32000 packages are available on the Comprehensive R Archive Network (CRAN)— those packages containing prewritten code, designed to accomplish specific tasks or a collection of tasks.

R as a computer programming language is based on an object-oriented programming (OOP) structure that relies on the concept of classes and its objects. Like any computer language, variables provide a means of accessing the data stored in memory. R does not provide direct access to the computer's memory, but rather provides specialized data structures, called objects, that are referenced via symbols or variables. In R, one can have objects based on expressions where an expression contains one or more statements. A statement is a syntactically correct collection of tokens. Expression objects are also special language objects that contain parsed, but unevaluated, R statements. All objects can have one or more attributes attached to them. The user can run multiple packages under the same code.

The first known attempt to simplify programming development under Peirce's theory was made by Kumiko Tanaka-Ishii (2010) in her book, *Semiotics of Programming*. The aim of the book was to provide a semiotic analysis of computer programs along three axes: 1) the models of signs, 2) kinds of signs, and 3) systems of signs. According to Tanaka-Ishii (2010), the transformation of Peirce's triadic theory to object-oriented programming (OOP) needs to apply a name, data, and functionalities to generate a working code. She identified two major problems regarding the conversion of Peirce's theory to a working programming code. The first issue is the difficulty in implementing the thirdness principle, and the second is in obtaining descriptions for the sign systems. To solve these problems, Friedman and Feichtinger (2017) created an R package called Peirce's sign theory R package that was used in this study. This package applies Peirce sign triangulation attributes to the objects found in the data and their variables. The R platform allows any user to run this code using their local machine and display the results graphically using different graphic packages. The package provides tools to classify and identify relationships between different components/variables in data sets by applying Peirce's sign theory of triangulation to qualitative and quantitative data.

Punctum. International Journal of Semiotics | **08**:01:2022
ISSN 2459-2943 | DOI: 10.18680/hss.2022.0007 | **punctum.gr**

Under Peirce's sign theory R package gathers the data based on the three components: the Representamen, the Object, and the Sign. As part of this study, Peirce's sign package was designed to follow Jappy's definitions of visual triadic signs, and to easily convert data into Peirce's triangulation algorithm model. Studying "visual meaning" in this case means studying Twitter's data on what the signs refer to, based on the researcher's observations of these terms, using the machine as a tool to rank and sort these classifications to produce the visualization results. The machine cannot decide for itself, the nature of the relationships among the three components of Peirce's model; for example, it does not recognize human concepts of words or numbers. Therefore, the main feature of this package is the functions in the package that can evaluate hypotheses about relationships that are found to be meaningful in the text. The package uses semiosis algorithms to find meaning in the data and gain insight into the causal relationships that are made possible by the underlying logic in Peirce's theory. In the simplest sense, the functionality in the package allows the user to study the interaction between the three variables directly and indirectly.

c. Peirce semiotics, visual signs, and how semiotics is useful for analyzing the meaning of Big Data

Semiotics, the study of signs, is derived from philosophical speculations on signification and language (e.g., Chandler 2007). The first known reference to the term 'sign' can be found in ancient Greek, where it is connected to the word semeion, meaning 'mark' or 'sign.' Elaboration on the meaning of the term continued through the 19th century, with two schools of thought providing different interpretations. This study adopts a conceptualization of 'sign' proposed by Peirce, who described the study of sign as "semeiosis," which aims to study all sign-related phenomena. Peirce offered a triadic foundation of the term, in which anything that may be interpreted as signifying, referring to, or standing for something other than itself may be considered a sign, see (CP 2.228, 1897), however in terms of semiosis, Peirce argued that symbols grow, and that the meaning of concepts tends to motivate more developed conceptions (CP 2.303, 1902). While Peirce in his extensive writings did not use the concept of visual signs or visual semiotics, he did use the photograph as an example of an icon, see (CP 2.281, 1894) where the photograph is considered as an example of representation by likeness or resemblance. And, in (CP 2.320, 1903), the photograph is mentioned as a hypoicon and a decisign. Leja (2000), also cited Peirce's own words on the importance of visual diagrams to examine the meaning in language.

> I do not think I ever reflect in words. I employ visual diagrams, firstly, because this way of thinking is my natural language of self-communion, and

> secondly, because I am convinced that it is the best system for the purpose. (MS 619, 190).[2]

Peirce definition of the sign is triadic, consisting of a Representamen, an Object and an Interpretant. Peirce formulated in his extensive writings several definitions of the sign, (Marty, 1997 [2012]), and his most famous definition is from 1897, which considers how the sign is divided.

> a sign, or a representamen, is something which stands to somebody for something in some respect or capacity. It addresses somebody, that is, creates in the mind of that person an equivalent sign, or perhaps a more developed sign. That sign which it creates I call the interpretant of the first sign. The sign stands for something, its object (or referent). It stands for that object, not in all respects, but in reference to a sort of idea, which I have sometimes called the ground of the representamen (CP 2.228, 1897).

Furthermore, Peirce's triad are qualified by additional categories where the Representamen is a first characterized by the qualisign, the sinsign, and the legisign. The Object is a second characterized by its relation to the Representamen by the icon, the index, and the symbol. The Interpretant or the meaning of the sign relationship is a third, characterized by the rheme (rhematic sign), the dicent sign (or proposition), and the argument (CP 2.243-252, 1903). This triadic nature of the sign can thus be combined into ten sign categories (CP 2.227-273, 1887). According to Queiroz and Merrell (2006), Peirce's believed that the core three elements, and the respective classifications they imposed upon signs, could be combined to give a complete list of sign types. Therefore, the sign is a triadic relation, and it involves three core elements for analysis, namely, the elements concerning the representamen, the relation between the representamen and the object, and the relation between the representamen and the interpretant respectively. Queiroz (2012) later argues that thus Peirce's classification should be considered an important advancement with respect to the task of empirically modeling the morphological variety of signs, and it constitutes one of the most important topics of Peirce's mature semiotics.

Peirce's classification of ten possible sign types provides a road map to analyzing images, diagrams, and metaphors according to Farias and Queiroz (2006). They extended Merrell's (1997) discussion of the ten classifications into figures that illustrates this transformation of Peirce's sign into visual semiotics. Figure 1 represents Peirce's 10 classifications of signs, with special attention to the hypoicons as discussed by Merrell (1997) and Farias and Queiroz (2006).

[2] We are referring to the *Annotated Catalogue of the Papers of Charles S. Peirce* (Robin (ed.) 1967), according to the scholarly tradition, MS, manuscript number and page number.

Punctum. International Journal of Semiotics | **08:**01:2022
ISSN 2459-2943 | DOI: 10.18680/hss.2022.0007 | **punctum.gr**

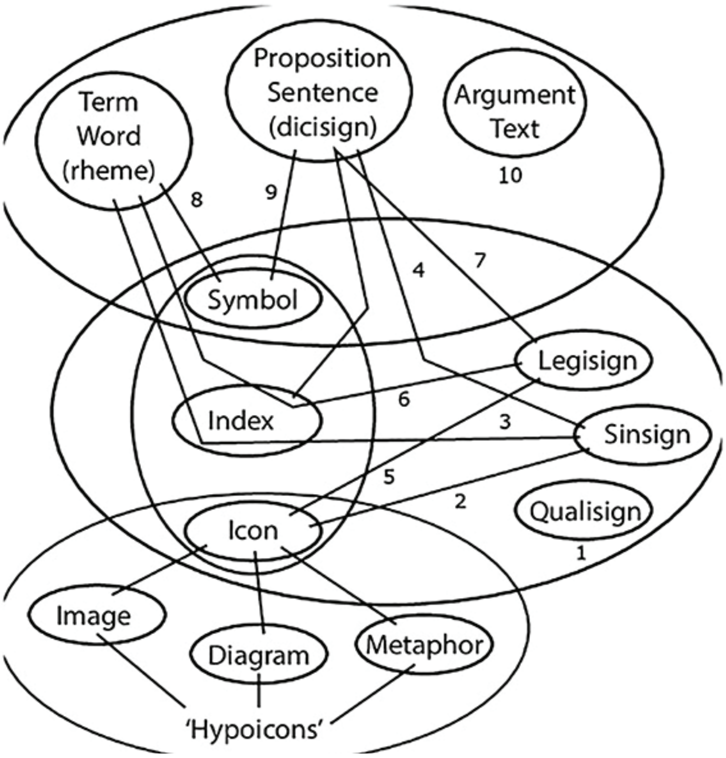

Figure 1. Peirce's classification of ten sign types supplied with the hypoicon as proposed by Merrell (1997).

With this Peirce classification system, the concept of the 'hypoicon' subdivides the icon sign into three different modes of iconicity, namely the image, the diagram, and the metaphor.

> Hypoicons may be roughly divided according to the mode of Firstness of which they partake. Those which partake of simple qualities, or First Firstness, are images; those which represent the relations, mainly triadic, or so regarded, of the parts on one thing by analogous relations in their own parts, are diagrams; those which represent the representative character of a representamen by representing a parallelism in something else, are metaphors. (CP 2.277, 1902).

As a part of their discussion of Peirce's visual semiotics, Farias and Queiroz (2006), Jappy (2013), and Peirtarinen (2012) employ the term 'hypoicons' as a core part in their discussion. They emphasized the importance of the subcategories of image, diagram, and metaphor to their overall study and interpretation of Peirce's types and symbols classification.

The image

The image represents by virtue of qualitative similarity–e.g., color on canvas, the canvas itself, etc. Of course, the very attempt to describe a sense of firstness related to an image transcends into other signs. As is the case with images e.g., paintings, they are interpreted by placement, framing, situation, and description as such, and thus involves the interpreter's ability to connect collateral experience with the image.

The diagram

The diagram represents its object by means of structural similarities to geometrical figures. A Venn diagram is an example of diagrammatic reasoning, as a map e.g., GPS (Global Positioning System) such as a navigation map or subway station maps–though different in visual expression and abstraction enables navigation that also includes pictograms and diagrammatic signs. For example, a common illustration of signs is the universal toilet gender signs. As is the case with the image, the diagram, besides its visual expression, involves indexical as well as symbolic aspects. A GPS navigation map, besides being visual and diagrammatic, is also indexical in its reference to the actual placement of the vehicle, the distance between point of departure and arrival and correspondence to the actual road being driven. The distance is measured in kilometers or miles and time, which of course are symbolic signs.

The metaphor

The metaphor that represents its object by means of a parallelism, or similarity to something else. The metaphor is considered an iconic sign; however, the metaphor involves a creative element. The metaphor may be regarded as an analogy or isomorphic, but also can be used as a metaphor for works of art (Anderson, 1984).

Peirce's theory places the metaphor under icons together with images and diagrams. Anderson (1984 pp. 463) reported on the difficulty of fitting his brief remarks about visual metaphor into a system that is not explicit. He addressed the issue by stating that there are many strands in Peirce's thought that it is easy to begin any study with conflicting fundamental views of his intentions. Yet, with all the attempts to define Peirce's line of thoughts, Anderson (1984, pp. 463) recommends us to read carefully and comprehensively, as he finds both growth and coherent direction in Peirce's work.

A different approach to Peirce's sign theory focused on diagram as visual formation. Shin (2002) reviewed Peirce's logic calculations found in a single sentence. She noted that Peirce's diagram holds his theoretical foundation, with a focus on the construction of a single sentence, and its visual representation captured by the diagram. This visual structure contains a single continuous line of visual display. However, given the sheer volume of Big Data, it would be difficult to merge a single sentence into

Punctum. International Journal of Semiotics | **08:**01:2022
ISSN 2459-2943 | DOI: 10.18680/hss.2022.0007 | **punctum.gr**

millions of records to find the general meaning of those messages. Many critical thinkers of Big Data consider social media data to reflect 'garbage in and garbage out' scenarios, since tweets frequently lack standard structure and, in many cases, analogical structured tweets. (e.g., Geiger and Kubin, 2020). As a result, social media data researchers do not often use diagrams to graphically display their findings. (e.g., Tsou and Leitner, 2013). Big Data researchers identified two related problems associated with the use of diagrams to visualize Big Data. The first is based on a single sentence structure, where today's social media analysis consists of the frequency of the terms found in social media feeds without any grammatical structure (e.g., Geiger and Kubin, 2020, Tsou and Leitner, 2013). Manovich (2005, 2013), Hochman (2014) discussed a different approach to social media images, focusing on algorithms and data procedures that occurred in real-time analysis that generate specific types of images. This type of analysis is based on social media feeds into visualization where the most frequent types of visualization are displayed with visual networks, geospatial visualizations (like heatmaps), and word clouds, rather than with the deeper theoretical or graphical analysis (Miller, 2017).

Overall, Peirce's work makes substantial contributions to the field of visual communication, and many researchers have incorporated Peirce's theory of visual signs into their overview of the meaning, representation, and reference found in visual language (e.g., Van Leeuwen 2005, Nöth 2011, Dunleavy 2005). According to Jappy (2013), Peirce logically defined the importance of methodological investigation through a general theory of signs for representation and their functionalities. He outlined two primary factors that distinguish Peirce from any other visual semiotics philosophers. The first attribute of Peirce's theory is the activity and the interpretation of signs as a process. The second attribute is the use of multiple approaches to extend our understanding of nature and rhetoric to both verbal and visual representations of the world we live in (Jappy 2013, pp. xi). A different interpretation of Peirce's visual sign theory was outlined by Pietarinen (2012), who reported that image languages are constituted by logical diagrams even though they relate to non-logical vocabularies. However, none of these interpretations have ever been examined in the context of Big Data visualization.

D. Methodological notes

In this study, we ask the following question: What insight does Peirce's visual sign theory lens provide us when using algorithms and open-source technology to create and interpret Twitter data? To address this question, first, we present the results of Twitter Analytics, and the use of Peirce's sign theory R package to generate and visualize the data. Second, we investigate whether we can use digital image processes and algorithms to evaluate the results considering Jappy's (2013) and Peirtarinen's (2012) discussion of Peirce's visual interpretation.

E. The study methodology

We collected data from Twitter by following the popular TV show Squid Game. The show is a South Korean survival drama television series that gained immense global popularity. We selected this show for this study because it boasts an international audience and the show's Twitter account has more than 300,000 followers. Our first data categorization was based on Twitter Analytics data, including the numbers of retweets and impressions, and engagement. A recent report by the New York Times showed that the number of new users joining Twitter continued to rise in 2022, with its users speaking over 40 languages. (e.g., Isaac et al. 2022). According to Gligorić et al., (2020), Tweets are textual messages with a 280-character limit that users can share, retweet, and post on other social media platforms. The popularity of Twitter data as a source for Big Data produced a shift in scale, scope, and depth of analysis (e.g., Kshetri, 2014). To provide more transparency in Twitter data analysis, Twitter employed the terms impressions and engagement, taken from the advertising industry to measure the user activity on their site (Siyam et al., 2020). Twitter data may include user URLs, references to other tweets, hashtags, abbreviations, and emoticons in user content, and/or metrics of user behavior associated with platform content.

Recent research involving the Twitter platform has highlighted new metrics for measuring customer engagement (Muñoz-Expósito, 2017). This change is based on how Twitter sorts its data, with a focus on the timeline as a key element. Previously, tweets were presented chronologically, as a time-stamped record. However, a newest sorting algorithm gives preference to tweets that are more relevant to the topic, as determined by Twitter (Wang and Fikis, 2019). Associated with this change are three key algorithmic elements: tweet activity, number of impressions, and engagement. Tweet activity is a measure of one's own activity on the Twitter dashboard and allows the user to review various metrics related to a tweet's performance (e.g., Siyam et al., 2020). The number of impressions is the number of times that a tweet appears to users, either in the timeline or search results. Engagement represents the total number of times that a user interacts with the tweet and includes retweets, follows, likes, and the addition of hashtags to the tweet.

In this study, instead of using the three general components of Peirce's sign theory: the Representamen, the Object, and the Interpretant, we focus our attention on the triadic division of the Object, i.e., the icon, the index, and the symbol. According to Peirce, an iconic sign shares resemblance with the object it represents. Photographs are common examples of iconic signs because they resemble the object represented (thus the division between immediate object (the photograph) and dynamical object (the live model depicted), however, as described above, the icon can also be subdivided further into the different subordinate types known as hypoicons: the image, the diagram, and the metaphor.

Punctum. International Journal of Semiotics | **08**:01:2022
ISSN 2459-2943 | DOI: 10.18680/hss.2022.0007 | **punctum.gr**

While the symbol is conventional in nature, the index is related to causality and thus shows evidence of what is being represented (Moriarty, 2002). We matched the sign-typology of the object with the categories of Twitter Analytics: Engagement, Twitter activity, and Impression. Our first idea was to consider a resemblance between the iconic sign and Twitter activity; in this study, both terms (Icon and Twitter activity) represent the actual account of all user activity on Twitter, as such Twitter activity resembles or mirror user interaction on the Twitter platform. We then paralleled the term Symbol with the term Impression, where both terms refer to the conventionalized meaning and connections between components within their own configuration. Finally, we associated the term Engagement with the Index, which stands for the causal relationship between user interaction and a specific tweet. Figure 2 demonstrates the match between the Peirce Object typology and Twitter's analytical terminology.

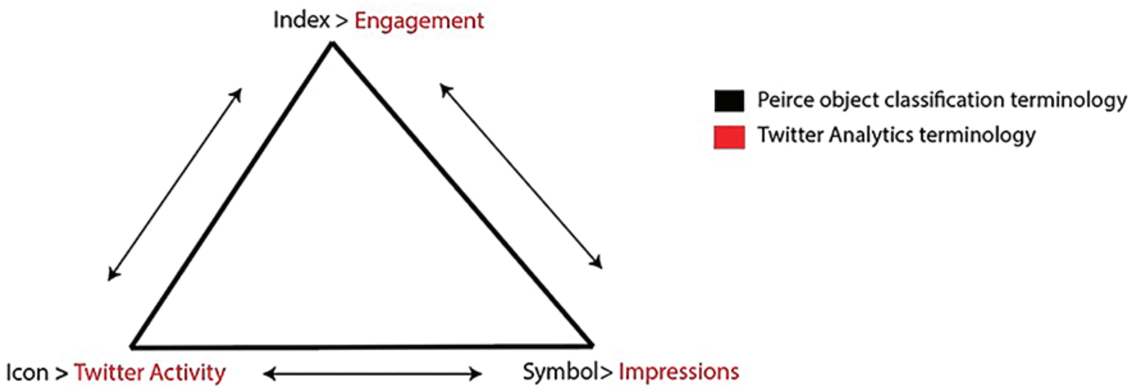

Figure 2. Shows the coordination match conducted in this study between Peirce object and Twitter's analytical terms.

Our next step was applying the Twitter data analytics data through the Peirce's sign theory R package. This package is based on open-source R which provides a tool to classify and identify and visualize the relationships among the components of data sets by applying Peirce's sign theory R package to qualitative and quantitative data. By implementing Peirce's model from a semiotic perspective, it allows the user to find meaning among the components represented by the data. In this study, the meaning of tweets for the show Squid Game posted on Twitter was examined. Studying "meaning" in this case means studying the interpretations of Twitter's Analytics classification and their relationships. The machine cannot decide for itself the nature of the relationships among the three components of Peirce's model; for example, it does not recognize human conceptions of words or numbers. Therefore, the main feature of this package is its use of user-defined input, where the functions in the package can evaluate hypotheses about relationships that are found to be meaningful in the text.

The Peirce's sign theory R package contains four main algorithms that were utilized to analyze the data. Function 1 is for numeric data; it calculates distributions by placing each numeric value in a percentile rank based on all values in that column of a data set, where columns represent each of the three components and each row represents a triad. This procedure is carried out for all three columns and for each variable. The average percentage rank is then computed for all three columns in each row. By quantifying where each data point is within a distribution, we can assess whether numbers in one column are associated with numbers in another. The average percentage values for all three components indicate whether a relationship exists, based on semiosis. Values that depart from 50% in one direction (either greater than or lesser than) indicate a significant association among the three variables. The researcher then determines whether these relationships are meaningful in the context of the hypotheses and goals of the study. Function 2 applies Peirce's classification to the output from Function 1 to look for further relationships within the data, as defined by the user's concept of data meaning. Function 3 first evaluates whether two or more rows are exactly equal, meaning that they represent two instances of the same triad. A logical value (true or false) is returned. A value of 'true' indicates that two rows have exactly equal values for index, symbol, and icon. A value of 'false' indicates that the rows do not have exact matches for each component. Function 3 then evaluates the column positions (corresponding to the index, symbol, and icon) within the rows to look for matches in any of the three components of a triad represented by a row in the data frame. It returns another set of logical values that indicates whether any components have an exact match within the rows. For example, it would identify a match between a triad with good/retweet/quote and another with bad/retweet/quote. To accomplish this task, the package makes use of R's loop function, which allows the user to iterate a procedure over the entire data set. The loop function provides instructions that allow automation of the code that is to be repeated. Figure 3 summarizes the workflow of Peirce's sign theory R package.

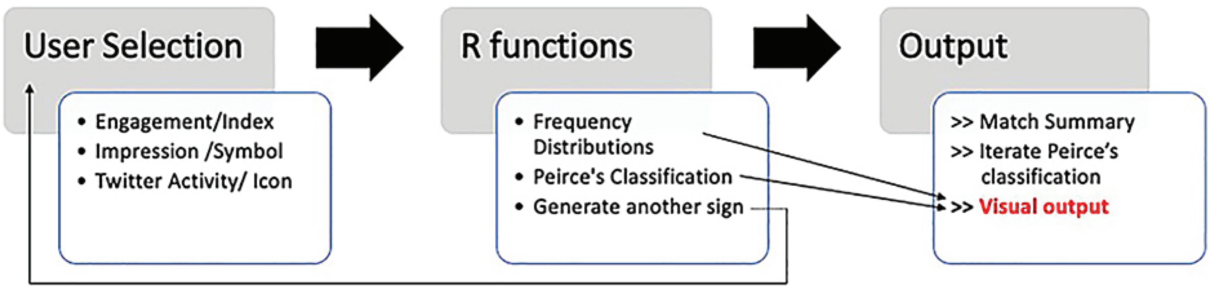

Figure 3. Outlines the workflow of Peirce's R package and its implementation in this study.

Punctum. International Journal of Semiotics | **08:**01:2022
ISSN 2459-2943 | DOI: 10.18680/hss.2022.0007 | **punctum.gr**

F. Results

We collected more than 350,000 individual messages (tweets) posted on the official Squid Game Twitter account in 2020 and 2021. First, we utilized the Twitter Analytic algorithms and then sorted the data using Peirce's sign theory R package to execute the triangulation algorithms and visualize the results. The Twitter Analytics function sorted the data based on Twitter activity, number of impressions, and engagement. The Twitter activity algorithm represents the sum of all account activity; in this case, it referred to all activities on the Squid Game account. Impressions represented the number of users who saw the Squid Game tweets. The engagement was calculated as a total rating, based on the total number of user engagements divided by the total number of impressions, and multiplied by 100. We then applied Peirce's sign theory R package to these data to sort them and visualize the result. The first visualization was based on the most frequent words that appeared in Squid Game tweets. The top five most frequently used terms were love, game, squid, fan, and life. User comments were tokenized, cleaned, counted, and reordered so that the most mentioned terms appeared first. Figure 4 shows the 16 most common terms appearing in user tweets.

Next, we examined interconnections between variables obtained from Twitter Analytics (engagement, impression, and Twitter activity) using Peirce's sign theory R package. Friedman and Smiraglia (2013) examined Peirce's visual signs in academic products and reported that line graphs, bar graphs, and histograms were the most common types of graphs used by researchers in conference presentations. Accordingly, Figures 5, 6, and 7 show the results of the current analyses of variable interconnections using a bar

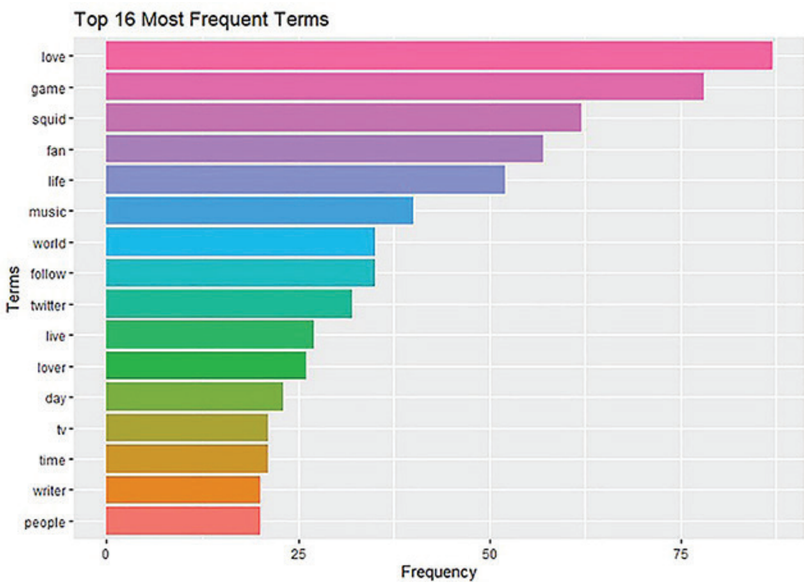

Figure 4. The top 16 most frequent terms found in the 2020 and 2021 Squid Game Twitter feeds.

chart, line graph, and bubble graph format, respectively. Figure 5 details the term engagement represented on the Y-axis and the number of impressions shown on the X-axis. According to Chae (2015), impressions and user engagement represent the core metrics of Twitter Analytics. However, to our knowledge, no previous study has examined the interactions between Twitter Analytic core variables.

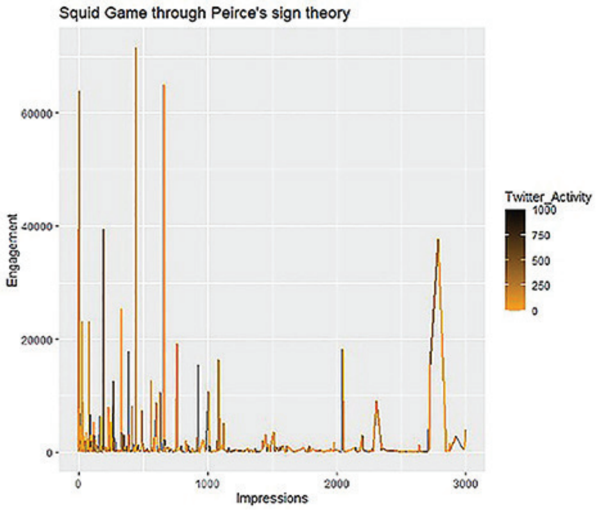

Figure 5. The relationship between the number of impressions and user engagement, shown using a line graph display.

To examine interconnections among all three user analytics variables, we counted the number of activities for each variable in our data set and summarized the result as a bar graph. Figure 6 summarizes user activity ranked highest in overall value.

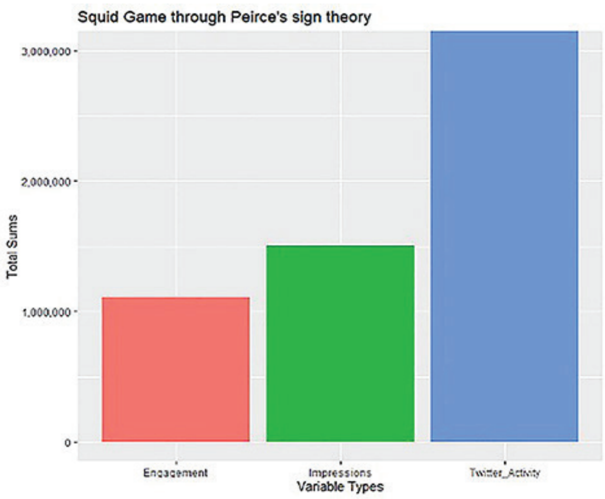

Figure 6. Summary of users' Squid Game Twitter activities shown using a bar graph.

Punctum. International Journal of Semiotics | **08:**01:2022
ISSN 2459-2943 | DOI: 10.18680/hss.2022.0007 | **punctum.gr**

Next, we counted the correlation analysis between the three terms and displayed the results as a correlogram. Twitter activity had a stronger correlation measurement of the strength of the relationship between the two other variables. However, extreme outliers were present in each category with a negative score observed under the impressions and engagement categories. Figure 7 summarizes the correlation found between the three categories.

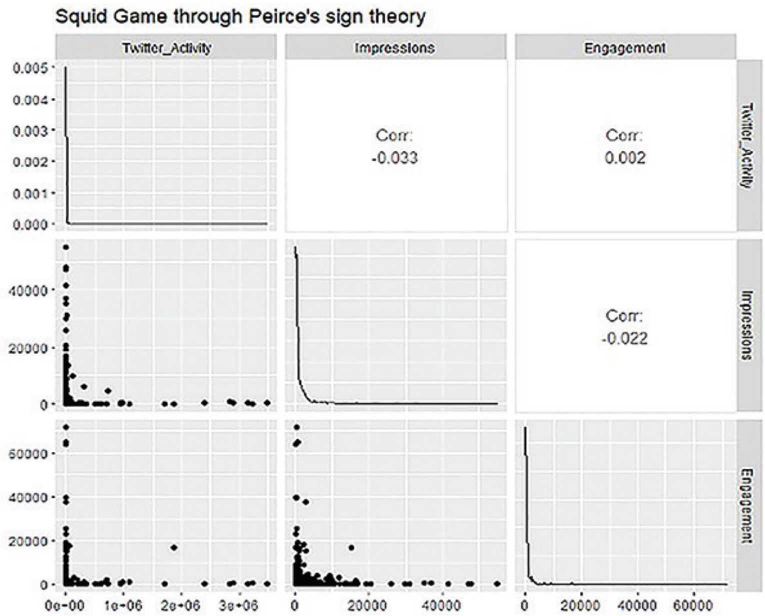

Figure 7. The correlations between the three variables in Twitter Analytics, shown as a correlogram.

The leading form of user activity on the Squid Game Twitter account was retweeting. Yang et al. (2010) reported that retweeting behavior can be captured by a statistical power-law distribution 35% of the time while other retweeting behavior does not match any statistical distributions with 31%. Our findings showed similar patterns, but we were unable to capture retweeting behavior under our data collection distribution. Approximately 69% of retweets did not refer to Squid Game, but rather to other activities.

Last, we tried to develop algorithms consistent with the criteria of Jappy (2013) and Peirtarinen (2012) for the evaluation of images according to Peirce's principles of visualization. Many open-source processes exist that allow for the analysis of image content. Those processes include techniques that allow the machine to analyze the physical attributes of the image, such as the pattern of the image and the point differential found on the image, among others. This process is based on the ability to convert images to data that are conducive to analysis. According to Pavlidis (2012), it is difficult to develop dual processes for the analysis of images that also incorporate the human interpretation of image meaning. We attempted to convert Jappy's (2013) and Peirtarinen's (2012) criteria to algorithms to analyze the images we created. However, we found many of Peirce's principles of logic as outlined by Jappy (2013) and Peirtarinen (2012) are not easily applied to the analysis of Big Data visualizations. For example, the graphs we created were based on more than 350,0000 data points, using a machine that could measure and assess the interactions between the three variables. However, according to Jappy (2013, p 167), interpretation of the graphs according to Peirce's principles would require an independent lens of psychology, ethnology, and cultural influences. We were unable to generate an independent algorithm or data set to match this demand. Likewise, one of Peirtarinen's core principles is reminiscence. This principle recommends the user to collect histological data and images to better evaluate the image. However, no historical Twitter data is currently available. These barriers prevented us from developing algorithms that faithfully captured Peirce's principles, as outlined by Jappy (2013) and Peirtarinen (2012).

Taken together, this study exhibited that the Peirce's sign theory R package can be used to analyze data from social media feeds to produce multiple types of data visualizations, together with information gleaned from Twitter Analytics. However, we were unable to develop suitable algorithms to follow Peirce's visual interpretation language due to difficulty in providing accurate interpretation of the meaning found in visuals, as understood through the lens of Peirce's theory. Indeed, according to Ezhilraman and Srinivasan (2018) the development of suitable algorithms and processes to evaluate and support the creation of visualizations presents an ongoing challenge.

G. Discussion

This study highlighted two major gaps between Big Data visualization under Peirce's sign theory found in social media feeds. The first gap is the unmatched examination between Twitter Analytics data and Peirce's sign algorithm conducted on the open-source R platform. The subject of data authenticity found in Twitter data was discussed by many researchers who have reported that great quantities of tweets are generated by AI bots and are difficult to identify and remove from Twitter. To address this challenge, we analyzed Twitter Analytics as our first step in examining the core triangulation variables, for which no studies had been conducted. Our next step in this study employed open-source Peirce's sign theory R package and matched it to Twitter Analytics to report on interconnection and correlation between these three categories. The second gap the study found was the lack of a machine image evaluation algorithm based on Peirce's visual sign and grounded in social media data. Under this gap, we examined Jappy (2013) and Peirtarinen (2012) interpretations of Peirce's sign. However, we were unable to generate visual evaluation algorithms that matched Peirce's sign theory due to the complexity of the criteria together with the problematic nature of Twitter data authenticity.

Punctum. International Journal of Semiotics | **08:**01:2022
ISSN 2459-2943 | DOI: 10.18680/hss.2022.0007 | **punctum.gr**

To address those two gaps, this study employed the open-source Peirce's sign theory R package as its supporting platform for its social media data analysis and visualization. The open-source R programming language is known for its Big Data and visualization capabilities, especially with Twitter data. (e.g., Bello-Orgaz et al. 2016). Future studies will need to examine more advanced algorithms developed using Peirce's sign theory in terms of human perspectives to better understand observed data and data visualization models found in social media.

H. Summary and conclusions

Social media platforms generate large amounts of data that reflect the complexity of human and machine activities. Many researchers have reported new concerns about the viability of these social media channels regarding machine-generated messages. As those data sets continue to grow in quantities, researchers have also used various theoretical and algorithmic frameworks to analyze social media feeds. Semiotics is the discipline that studies the foundation of sign associated with text and images and their meaning found in language and communication. One of the founders of semiotics is the American philosopher Charles Sanders Peirce who offered a triadic foundation for the term sign. This study raises the question: what insights do Peirce's visual sign theory contribute when we try to interpret Twitter data analysis? To address this question, first, we present the results using Twitter Analytics, and Peirce's sign theory R package to generate results visually. Second, we investigated whether we could use digital image processes and algorithms to evaluate the results considering Jappy's (2013) and Peirtarinen's (2012) discussion of Peirce's visual interpretation. The study's visual result showed common data visualization-type displays of Twitter data analysis using the sign theory of Peirce's R package technology with Twitter Analytics. Those visual results present interactions among the three Peirce visual sign categories and provide deeper insight into key concepts using Twitter Analytics terminology. We then explored the feasibility of developing algorithms for analyzing these visualizations based on criteria set out by Jappy (2013) and Peirtarinen (2012), who outlined Peirce's search for meaning found in visualizations. However, such algorithms were hard to assemble due to the complexities of those criteria and the complexity of social media data feeds: especially with human vs machine messages. As a result of our study, we established a close relationship between Peirce's semiotics framework and Twitter analytics and the generation of visualization using Peirce's theory. Future studies should investigate the transformation of Peirce's sign theory into more complex algorithms, as well as the behavior and interaction of social media users with visual content.

References

Aiello, Giorgia 2020. Visual semiotics: Key concepts and new directions. In: Luc Pauwels and Dawn Mannay (eds.) *The SAGE handbook of visual research methods*. London: Sage, 367-380.

Anderson, Myrdene 1984. Synthetic potential within, beyond, and through semiotics. *Semiotics* 363-372.

Barthes, Roland 1977. *Image-music-text*. London: Macmillan.

Bello-Orgaz, G., J.J. Jung, and Camacho, D. 2016. *Information Fusion* 28: 45-59.

Boyd, Danah and Crawford, Kate 2011. Six provocations for big data. In: *A decade in internet time: Symposium on the dynamics of the internet and society.*

Chae, Bongsug K. 2015. Insights from hashtag# supplychain and Twitter Analytics: Considering Twitter and Twitter data for supply chain practice and research. *International Journal of Production Economics* 165: 247-259.

Chandler, Daniel 2007. *Semiotics: the basics*. London: Routledge.

Correia, P. A., I.G. Medina, Z.F.G. Romo, and R.S. Contreras-Espinosa 2014. The importance of Facebook as an online social networking tool for companies. *International Journal of Accounting and Information Management*.

De Mauro, A., G. Andrea and M. Grimaldi, 2015. What is big data? A consensual definition and a review of key research topics. In: *AIP conference proceedings* (Vol. 1644, No. 1, pp. 97-104). American Institute of Physics.

Djonov, Emilia, and Leeuwen Theo Van 2018. Social media as semiotic technology and social practice: the case of ResearchGate's design and its potential to transform social practice. *Social Semiotics* 28(5): 641-664.

Dunleavy, Dennis 2004 The image and the archive: A semiotic approach. In: Kenneth L. Smith, Sandra Moriarty, Keith Kenney and Gretchen Barbatsis (eds.). *Handbook of visual communication*. New York: Routledge, 279-290.

Elhamy, Hossam M. 2022. Social Semiotics for Social Media Visuals: A Framework for Analysis and Interpretation. In: Gabriella Punziano and Angela delli Paoli (eds.) *Handbook of Research on Advanced Research Methodologies for a Digital Society*. IGI Global, 548-570.

Ezhilraman, S. Vahini, and Srinivasan, Sujatha 2018. State of the art in image processing and big data analytics: issues and challenges. *International Journal of Engineering and Technology*, 7(2.33): 195-199.

Fagni, T., F. Falchi, G. Margherita, M., Antonio and Tesconi, M. 2021.TweepFake: About detecting deepfake tweets. *PLoS one*, 16(5), e0251415.

Farias, P., & Queiros, J. 2003. On diagrams for Peirce's 10, 28 and 66 classes of signs. *Semiotica 147*(1/4): 165-184.

Farias, Priscila, and Queiroz, João 2006. Images, diagrams, and metaphors: Hypoicons in the context of Peirce's sixty-six-fold classification of signs *Semiotica* 162: 287-307.

Punctum. International Journal of Semiotics | **08:**01:2022
ISSN 2459-2943 | DOI: 10.18680/hss.2022.0007 | **punctum.gr**

Farias, Priscila, and Queiroz, João 2017. Visualizing triadic relations. *Information Design Journal* 23(2): 127-147.
Floch, J.-M. 2001. *Semiotics, marketing, and communication : beneath the signs, the strategies* (R. Orr-Bodkin, Trans.). New York: Palgrave.
Friedman, Alon and Smiraglia, Richard J. 2013. Nodes and arcs: concept map, semiotics, and knowledge organization. *Journal of documentation 69*(1).
Friedman Alon and Feichtinger, Erin 2017. Peirce's sign theory as an open-source R package. *Signs* 8(1-24).
Geiger, Bernhard, and Kubin, Gernot 2020. Information bottleneck: Theory and applications in deep learning. *Entropy* 22(12): 1408.
Gligorić, Kristina, Anderson, Ashton, and West, Robert 2020. Adoption of Twitter's New Length Limit: Is 280 the New 140? *arXiv preprint arXiv:2009.07661*.
Hochman, Nadav 2014. The social media image. *Big Data and Society* 1(2): 205395171 4546645.
Isaac, Mike, Lauren, Hirsch, and Das Anupreeta 2022. Inside Elon Musk's Big Plans for Twitter. *The New York Times*, May 6, 2022.
Jappy, Tony 2013. *Introduction to Peircean visual semiotics*. London: Bloomsbury.
Kshetri, Nir 2014. The emerging role of Big Data in key development issues: Opportunities, challenges, and concerns. *Big Data and Society* 1(2): 2053951714564227.
Kwak, Haewoon et al 2010. What is Twitter, a social network, or a news media? *In Proceedings of the 19th international conference on World wide web,* 591-600.
Leja, Michael 2000. Peirce, visuality, and art. *Representations* 72: 97-122.
Onursoy, Sibel. 2015. A semiotic analysis of an activist image in social media. *Online journal of art and design* 3(2): 1-13.
Lokot, Tetyana, and Diakopoulos, Nicholas 2016. News Bots: Automating news and information dissemination on Twitter. *Digital Journalism* 4(6): 682-699
Manovich, Lev 2005. Understanding meta-media. *CTheory* 10(26): 2005.
Manovich, Lev 2013. Media after software. *Journal of Visual Culture* 12(1): 30-37.
Marty, Robert 1997. 76 Definitions of the sign by C. S. Peirce (Department of Mathematics, University of Perpignan). retrieved from https://arisbe.sitehost.iu.edu/rsources/76DEFS/76defs.HTM
Merrell, Floyd 1997. *Peirce, signs, and meaning*. Toronto: University of Toronto. DOI: 10.3138/9781442678330.
Miller, James D 2017. *Big data visualization*. Birmingham-Mumbai: Packt Publishing Ltd.
Moriarty, Sandra, E. 2002. The symbiotics of semiotics and visual communication. *Journal of Visual Literacy* 22(1): 19-28.
Moschini, IIaria 2018. Social semiotics and platform studies: an integrated perspective for the study of social media platforms. *Social Semiotics* 28(5): 623-640.

Muñoz-Expósito, M., Oviedo-García, M. Á., and M. Castellanos-Verdugo (2017). How to measure engagement in Twitter: advancing a metric. *Internet Research*, 27(5): 1122-1148.

Nöth, Winfried, and Jungk, Isabel 2015. Peircean visual semiotics: Potentials to be explored. *Semiotica* 207: 657-673.

Nöth, Winfried 2011. Visual semiotics: Key features and an application to picture ads. In: Luc Pauwels and Dawn Mannay (eds.) *The Sage handbook of visual research methods*, 298-316.

Peirce, Charles S. 1931-1958. *The Collected Papers*. Charles Hartshorne, Paul Weiss, and Paul Burks (eds.) Cambridge, MA: Harvard University Press.

Peirce, Charles S. 1967. *The Annotated Catalogue of the Papers of Charles S. Peirce.* Robin, Richard S. (ed.), Amherst Massachusetts: The University of Massachusetts Press.

Pietarinen, Ahti-Veikko 2012. Peirce and the logic of image. *Semiotica* 192: 251-261

Poulsen, Søren, Gunhild Kvåle, and Van Leeuwen, Theo (2018). Social media as semiotic technology. *Social Semiotics* 28(5): 593-600.

Poulsen, Søren and Kvåle Gunhild 2018. Studying social media as semiotic technology: a social semiotic multimodal framework. *Social Semiotics* 28(5): 700-717.

Queiroz, João, and Merrell, Floyd 2006. Semiosis and pragmatism: Toward a dynamic concept of meaning. *Sign systems studies* 34(1): 37-65.

Queiroz, João 2012. Peirce's ten classes of signs: Modeling biosemiotic processes and systems. *Semiotics in the wild—Essays in honour of Kalevi Kull on the occasion of his 60th birthday*, 55-62.

Saussure, Ferdinand De 1959. *Course in general linguistics*. New York, NY: Philosophical Library.

Serafini, Frank, and Reid Stephanie F (2022). Analyzing picturebooks: semiotic, literary, and artistic frameworks. *Visual Communication*. https://doi.org/10.1177/14703572211069623

Shin, Sun-Joo 2002. *The iconic logic of Peirce's graphs*. Cambridge (MA): MIT Press.

Siyam, Nur, Alqaryouti, Omar and Abdallah, Omar (2020). Mining government tweets to identify and predict citizens engagement. *Technology in Society* 60: 101211.

Stjernfelt, F. 2007. *Diagrammatology: An investigation on the borderlines of phenomenology, ontology, and semiotics*. Dordrecht, The Netherlands: Springer.

Strand, Torill 2005. Peirce on educational beliefs. *Studies in philosophy and education* 24(3): 255-276.

Tanaka-Ishii, Kumiko 2010. *Semiotics of programming*. Cambridge University Press.

Tsou, Ming-Hsiang, and Leitner Michael 2013. Visualization of social media: seeing a mirage or a message? *Cartography and Geographic Information Science* 40(2): 55-60.

Tonidandel, Scot, King Eden, and Cortina Jose M. 2018. Big data methods: Leveraging modern data analytic techniques to build organizational science. *Organizational Research Methods* 21(3): 525-547.

Punctum. International Journal of Semiotics | **08:**01:2022
ISSN 2459-2943 | DOI: 10.18680/hss.2022.0007 | **punctum.gr**

Wang, Yinying and Fikis, David J. 2019. Common core state standards on Twitter: Public sentiment and opinion leaders. *Educational Policy* 33(4): 650-683.

Van Dijck, José 2013 *The culture of connectivity: A critical history of social media*. Oxford: University Press.

Van Leeuwen, Theo 2005. *Introducing social semiotics*. New York: Routledge.

Yang, Z., Guo, J., Cai, K., Tang, J., Li, J., Zhang, L., & Su, Z 2010. Understanding retweeting behaviors in social networks. In: *Proceedings of the 19th ACM international conference on Information and knowledge management* (pp. 1633-1636).

AUTHOR

Alon Friedman Associate professor at the School of Information, University of South Florida.

AUTHOR

Martin Thellefsen Associate professor at the Department of Communication, University of Copenhagen.