Code Ethnography and the Materiality of Power in Internet Governance

Fernanda R. Rosa

Accepted: 23 May 2022 / Published online: 30 August 2022
© The Author(s) 2022

Abstract
The purpose of this article is to discuss an ethnography of code, specifically code ethnography, a method for examining code as a socio-technical actor, considering its social, political, and economic dynamics in the context of digital infrastructures. While it can be applied to any code, the article presents the results of code ethnography application in the study of internet interconnection dynamics, having the Border Gateway Protocol (BGP) as code and two of the largest internet exchange points (IXPs) in the world as points of data collection, DE-CIX Frankfurt, and IX.br São Paulo. The results show inequalities in the flows of information between the global North and the global South and concentration of power at the level of interconnection infrastructure hitherto unknown in the context of the political economy of the internet. Code ethnography is explained in terms of code assemblage, code literacy, and code materiality. It demonstrates the grammar of BGP in context, making its logical and physical dimensions visible in the analysis of the formation of giant internet nodes and infrastructural interdependencies in the circulation information infrastructure of the internet.

Keywords Code Ethnography · Digital Ethnography · Digital Inequalities · Internet Interconnection Infrastructure · Internet Exchange Points (IXP) · Internet Governance · Border Gateway Protocol (BGP)
Introduction

More than a zettabyte (a billion terabytes) of data has been circulating online since 2015 according to estimations from the industry (Cisco 2016). This growing data traverses multiple paths from origin to destination, following a packet-switching design in which decisions about the routes of data packets occur through the interaction of devices, internet protocols, and the interests of internet intermediaries at the moment that networks interconnect. The choices and arrangements at play are barely known to the public, as they happen at the discretion of private actors, but the social, political, and economic implications they generate suggest that new ways to understand these flows of information online should be considered, especially in light of the pervasive commodification of personal data and the historical inequalities rooted in colonialism that mark the divide between the global North and the global South. This paper proposes to examine the information flow of the internet by including code in the discussion.

But what is code? As a social construct, it is a contested term, and can be defined in a way that includes certain artifacts and coders and excludes others (Couture 2019). Simply defined, it is that which makes an application or routine work. Taken more broadly, it is a social, political, and economic agent and the result of all those spheres. The term, as used here, includes, but is not limited to, software code. It includes the codes in digital social media, natural computer languages, internet protocols, etc. Clearly, we can opt to exclude code as an actor worthy of study when examining digital technologies, but this is becoming increasingly difficult. Despite the fact that it is not always accessible, or possesses properties that make it hard to understand, code is an ever-present avenue for analysis of the institutions that materialize the digital environment, its political economy, and the people who constitute it. Code allows researchers to “follow the thing” in ethnography (Marcus 1995), and, to denote such orientation, this paper proposes an ethnography of code, specifically code ethnography, which adds to ethnographies that have studied code by, for instance, following the people (e.g. Jaton 2021; MacKenzie & Monk 2004) or following the controversies (Coleman 2012; Couture 2019).

I was first introduced to the Border Gateway Protocol (BGP) in a fieldwork project in 2016. BGP is the code with which routers direct data packets circulating on the internet to their destination address. It is an internet protocol by definition, which went less visible in view of the initial ethnography of infrastructure approach I was applying to study other actors, including people, servers, switches, routers, and cables, via observation and interview, following the inspiring work of Susan Star (1999). With the initial goal of studying the interplay between local and global dynamics in internet governance, I spent two months going daily on weekdays to the Brazilian Network Information Center (NIC.br), in Brooklin, São Paulo to interview employees, to do participant observation, and to attend events sponsored by the organization. Ultimately, I sought to grasp what agency would perform national internet governance in the context of a global internet. NIC.br is administrated by a pioneer multistakeholder organization, the Brazilian Internet Steering Committee (CGI.br) (for more on it see Adachi 2011; Glaser and Canabarro 2016). During one week, I did my shift as a researcher at the department called IX.br, which is responsible for building and governing more than thirty internet exchange points (IXPs) in the country.
IXPs, or IXs, are distributed physical infrastructures that facilitate data traffic from one point to another on the internet by physically mediating interconnection arrangements between internet networks and about which more will be said later. In my first day, Antonio Moreiras, the IX.br project manager, introduced me to his team and then spent more than an hour explaining the functions of IXPs. To do that, he referred to a well-cited case of a BGP hijacking in 2008 that compromised YouTube access by maliciously diverting its traffic to Pakistan Telecom (RIPE NCC 2008). That became just an ordinary note in my laptop, but the visual of seeing traffic being diverted between networks would help me later understand that BGP is the language that logically structures the internet data traffic that IXPs physically mediate. Nevertheless, I dismissed BGP and its significance for a while to follow the people and the physical actors in internet interconnection around IXPs in that fieldwork.

A year later, I was continuing my ethnographic work at DE-CIX (Deutscher Commercial Internet Exchange), a world-wide commercial IXP native from Frankfurt, Germany. Differently from NIC.br, its board did not allow me to spend time at the company, so I looked for events sponsored by them, communicated with attendants and employees, sometimes arranging full interviews, and visited data centers in the city. During a participant observation at a DE-CIX event organized in a small town accessible by train from Frankfurt, I asked one of the directors if I could have access to anonymized data circulating within the company’s facilities in a bid to answer the question about the infrastructural dependencies between the global South and the global North in internet interconnection. The response was no. The data, I was told, was the private property of each autonomous system—a term that generally refers to internet networks. Actually, autonomous systems are operated by organizations (e.g., Google, Meta, internet service providers—ISPs—etc.), identified with an autonomous system number (ASN), and responsible for making decisions about where internet data traffic goes next after traversing their networks. I would need to talk to each of them. However, he suggested that I access the Looking Glass app, saying that its public BGP data might help. When I accessed the app on my cellphone, the BGP became an unavoidable actor, one I could interact with if I had a level of code literacy to understand the answers to my questions just like any other participant at the event. I did not have such literacy, and so code ethnography started.

It took a considerable amount of time for me to notice that my focus on the physical artifacts and the people building, running, and governing internet infrastructure facilities, specifically IXPs, had to be broadened to capture a scenario that was always there, constantly neglected by the ethnography I was conducting. When enrolled in the research, BGP amalgamated the design of a multi-sited study where the global North and global South could be examined symmetrically. As George Marcus wrote:

> in multi-sited ethnography, comparison emerges from putting questions to an emergent object of study whose contours, sites, and relationships are not known beforehand, but are themselves a contribution of making an account that has different, complexly connected real-world sites of investigation. (1995, 102)

Moreover, as Wendy Chun states, “Software cannot be separated from hardware, except ideologically” (2006, 19). As we shall see, the physical and the logical are
inseparable—IXPs, their physical artifacts, and BGP are part of an assemblage that requires them to be understood altogether. In this way, the relevance of any code does not mean that it and its attributes are superior to other actors, including programmers, users, companies, and other technologies needed for its functioning and existence. It means that code is part of a chain of relations and if understood as an actor, it can be inquired in its own terms and situated in its environment.

In this article, I present an application of code ethnography, a method for examining code as a socio-technical actor, considering its social, political, and economic aspects within the context of digital infrastructures. I apply code ethnography to BGP, showing how code potentially provides a vast, and as-yet unexplored field of investigation when included in the research.

Code ethnography draws from internet governance, infrastructure studies, and science and technology studies (DeNardis 2009; DeNardis and Musiani 2016; Latour 1999; Musiani 2015; Star 1999). In times of ubiquitous technologies, code is becoming increasingly entangled: “Programming is no longer relegated to static coding of fixed algorithms… Programming is about dynamically enabling the code to use statistical inference and large quantities of data to itself learn, adapt, and predict” (DeNardis 2020, 50). The need to apply a critical eye to the materiality of code becomes all the greater in this context, given the social relations manifested therein, and the societal effects that these dynamics can have. As Manuel Castells put it, “The specific means of switching and programming largely determine the forms of power and counterpower in the network society” (2011, 52, own translation).

Internet governance is an interdisciplinary field dedicated to the comprehension of the power dynamics shaping how the internet works in its multiple layers, from infrastructure to applications, and ethnographic methods have been used to study it less frequently than other approaches (for exceptions see Flyverbom 2011; Omari 2020; Vicentin 2016). Nonetheless, the increasing dissemination of ethnography to disciplines beyond anthropology, including sociology (Burrell 2012; Jaton 2021), the popularization of science and technology studies within internet governance (Musiani 2015), and the way that technical objects of study familiar to internet governance scholars have generated interest in multiple disciplines (Ramos and Freitas 2017) may render this approach more present. Actually, there is a thriving discussion on the possibilities of ethnography within the digital environment (Abidin and Gabrielle 2020; Hallett and Barber 2014), especially in studies on platforms and algorithms (Christin 2020; Seaver 2017). Nick Seaver, for instance, suggests ducking the barriers behind which algorithms are kept as secret, private property, and offers research tactics that help the researcher glean a great deal from how algorithms manifest in a range of spaces, from mailing lists to patent applications, in a sort of “pragmatic bricolage” (2017, 4). There is a synergy between this approach, which focuses on the discourse and practices that surround algorithms, and rhetorical code studies (Brock 2019), opening countless research possibilities based on discourse, without access to the code itself. We can understand this type of approach as focused on the “form of expression,” which is different from the “form of content.” In the language of Deleuze and Guattari, “the form of expression is reducible not to words but to a set of statements arising in the social field considered as a stratum… The form of content is reducible not to a thing but to a complex state of things as a formation of power.
(architecture, regimentation, etc.)” (1987, 67–68). Code ethnography illuminates such a difference. Examining code in its many facets, engenders a more comprehensive understanding of the underlying layers of the architecture of power of digital infrastructures as already shown, for instance, by the study of opacity in machine learning algorithms (Burrell 2016) and of norms encoded in internet infrastructure (ten Oever 2021).

To connect expression and content, Deleuze and Guattari use the concept of “assemblage,” which suggests connections, modularity, and contingency between humans and non-humans, acts, and emotions. As Jennifer Daryl Slack and J. Macgregor Wise wrote, “an assemblage involves a mixture of bodies, actions, and passions. In this sense, an assemblage is a particular constellation of articulations that selects, draws together, stakes out and envelops a territory that exhibits some tenacity and effectiveness” (2014, 157, italics original). The concept of assemblage is used here to expand the ethnographic attention to the code, without disregarding the agency that people and institutions possess and the influences that characterize them.

Considering these dialogues, code ethnography will be explained as a tool for researchers who, like me, are outsiders in the environment being studied. Though nothing prevents inside actors—those directly connected with the code—from wanting to study it, in a sort of “native ethnography” (Kraidy 1999). Code ethnography is conceived as a bridge in the sense Gloria Anzaldúa (2012) described, between technical fields such as computer sciences and engineering and the social sciences and humanities. It is a doorway into the code for non-specialists. Moreover, it becomes an instrument that adds to the investigation of the biases and values imbricated into digital technologies (Broussard 2019; Lima et al. 2019; Noble 2018; Silva 2022; Souza et al. 2018). The sections below explain BGP ethnography in terms of code assemblage, code literacy, and code materiality. The concept of assemblage contextualizes BGP as one more actor in the political economy of the internet, shaping the internet since its beginning; the concept of literacy discusses the researcher’s interactions with the code; finally, the concept of materiality highlights what BGP communicates in its own language. The results of this investigation point to material inequities of power embedded in internet infrastructure that shape the dynamics of internet interconnection between the global North and South—dynamics that I argue the code co-produces.

**On Politics: The Code as part of an Assemblage**

“Routers speak.” That’s what I was told at the head office of Brazil’s IX.br in São Paulo, that oversees the internet traffic exchange points in the country. Routers are fundamental devices used to establish network interconnections by reading data packets headers and forwarding data packets to the next router towards their destination (Clark 2018). Each router is associated with a different network, identified with an ASN, which make each network on the internet’s routing system unique. “Rout-

---

1 As an example, we might stress that, given the fact that algorithms are collectively produced, the feeling of being an outsider is common, even among those responsible for the programming (Seaver 2017).
ers learn” information in dialogue with other routers, identifying with each passing microsecond the best paths to specific addresses requested by the users.

BGP is the language used in communication between routers. It operates on top of the Internet Protocol (IP) and is programmed by professionals who continually configure routers so that they can ferry their data packets to their destinations as quickly and cheaply as possible. Internet providers look for ways to send their data packets at the lowest cost and greatest speed so as to minimize latency. Thus, as one might expect, underlying internet interconnection is not just a language as BGP, but money. More broadly, “the connection of different networks requires the capacity to build a cultural and organizational interface, a common language, a common medium, and the support of a universally accepted value: exchange value” (Castells 2011, 84, own translation). The more interconnected an ISP is, the more access it will have to addresses, enabling it to find the best paths, and making it more attractive to its customers, who pay to have their packets delivered efficiently.

An example can clarify this process. When an internet user decides to update an app A on their cellphone, their internet provider needs to connect to a network or an autonomous system A in order to exchange traffic: send the request and obtain the update. This connection between the ISP and content provider can be direct if both network operators agree and are physically interconnected. This type of direct interconnection is known as “peering,” and it is generally the most economical way to exchange traffic. However, if the ISP does not have any such agreement with the content provider behind the app A, it will have to send those data packets containing the update request to a transit provider to which it is a client and has a direct connection to the content provider A, or at least knows the route through which to connect to A. These relations are based on paid commercial agreements known as “transit.”

ISPs of similar sizes may also have peering relations similarly to how ISPs and content providers have. In such case, aiming to increasing their connectivity and competitiveness, ISPs share their addresses and those of their customers, allowing each peer to send data packets back and forth (usually, not always) cost free. ISPs prefer peering, and tend to reserve paid transit only for situations in which they have no direct interconnection option with a given autonomous system (for more see Faratin et al. 2008; Metz 2001). Put simply, internet providers prefer peering to transit when delivering data packets because, in the former, the relationship is between peers and may not involve any direct payment, while the latter is a customer/provider relationship based on payment. Internet users’ data packets circulate on the web effectively on the wings of these agreements and decisions based on the level of connectivity among the innumerable internet providers the packets move through in route to their destinations. This is, in a simplified way, the basics of internet interconnection.

BGP is a language that operationalizes the decisions concerning the paths to be taken. It is the logical dimension of internet interconnection. BGP not only allows the creation of databases of dynamic, constantly updated routes, but also defines how the routers will learn those routes and specifies the mechanisms by which those routes are constructed (Mathew 2016). In BGP, one of the most important properties for decision-making on pathways is AS_PATH, which defines the shortest route to a given destination. The shorter the route, the more likely it will be the path chosen. If two networks peer, and data is sent along their pathways, the number of hops (routers)
the data packets will have to take between network segments is one (1), the lowest possible and so the route most taken. The same logic applies to transit: the router at the origin verifies which of the ISPs of which it is a customer offers the shortest and/or most cost-effective route under existing commercial agreements. This is why it is so important for internet providers to increase their connectivity: the more addresses they can access via peering, the shorter their AS_PATH on different routes will be, and the greater the chance they will be paid to provide transit for other providers’ data packets. In a nutshell, this is the economic logic behind internet interconnection.

Normally the decision-making algorithm includes other properties, and some specialists describe BGP as a protocol based on “politicking,” the term one network engineer used to describe how route-choice goes beyond technical decision-making to reflect the economic interests of the companies. Here, within the context of an ethnography of code, we understand the BGP as a technopolitical property, in which the technical and political aspects are indissociable (Bruno et al. 2018; DeNardis 2014). In other words, in the context of code ethnography, understanding how a code functions technically also means grasping the politics that plays such a pivotal role in the assemblage of actors and interests that emerges from the study of it. The clarity of a code assemblage stems from a more thorough knowledge of the code in itself and of the actors surrounding it. ²

The most common way internet providers do peering, especially multilaterally, when they connect to various networks at the same time, is via the IXPs, which we can also call internet nodes, as they attract an untold number of networks, all interconnecting at the same location in search of peering opportunities. IXPs materialize one of the most important physical dimensions of internet interconnection infrastructure. Research suggests that one in five routes data packets pass through an IXP (Nomikos and Dimitropoulos 2016), showing the huge magnitude of an internet node as a passage point to data. ³ This movement should be understood in tandem with the grammar of BGP, the AS_PATH, which synthesizes the pursuit of more connectivity and shorter routes. As the next sections suggest, this grammar leads to the formation of IXPs as architectures of power, with especially high volumes of data circulation.

The aim to keep data traffic local, ensuring data packets that could circulate locally do not detour through other states or nations for lack of local interconnection points, has been at the root of IXP formation worldwide since the late 1980s. IXPs help reduce the long or international routes that would substantially increase the cost of internet providers and the data latency associated with internet quality. Hence, in the sphere of public policy, IXPs are tied in with cost-reductions and heightened internet competition (ISOC 2015; Kende and Hurpy 2012).

IXPs began to form in Brazil, one of my research sites, in the 1990s, and in the early 2000s the project was imbued with the need to maintain national sovereignty,

² This is beyond the scope of this article, not only the autonomous systems and companies that operationalize the BGP affect it; but states and their laws by permitting (or forbidding) data transit across national borders and by helping local companies become global players, for example, affect BGP as well. These and other social actors comprise the complex ecosystem of this language by having their (monetized) data circulate on the network.

³ Peering is also possible outside the IXPs, an option that is quite common among large companies that decide to make private bilateral agreements.
so that the exchange of traffic between internet providers, which used to take place in the United States or be mediated by a North-American company, could play out on national territory, under local governance (Rosa 2020). Brazil has the largest public IXP ecosystem in the world (Brito et al. 2016), with over 30 units, with significant impact (Woodcock 2013), but many factors influence internet data flows and off-IXP routes. For example, if the data to be accessed is located abroad, international transit will be inevitable. Among the domain names (websites) most frequently accessed in Brazil, 77% of the requests are resolved in the United States (i.e., the traffic ends there) Furthermore, the same study shows that 84% of access requests necessarily pass through the US, regardless of where the routes end (Edmundson et al. 2016). It is important to remember that data pathways are defined by autonomous systems in a distributed manner. Each network, upon receiving a data packet, determines only the next hop, that is, the next network in-line, which, in turn, chooses the hop after that. In an interview with a BGP specialist and lecturer in Brazil, I asked how state sovereignty, which had prompted the formation of the IXP ecosystem in Brazil, influences the routing decisions taken by the country’s internet providers. He was emphatic:

The issue of sovereignty, on a more specific level than the sovereignty of states, is the sovereignty of the companies themselves. The concept of an autonomous system means precisely that, you are able to send data from one company to another, or not. Each company, each autonomous system, is either sovereign in this regard or it is not. From the very moment a company sends data, the receiver also becomes sovereign and can handle that data however it likes. Who is going to control the autonomous systems? It’s a philosophical question, because they would cease to be autonomous. (personal communication, own translation)

The reasoning presented in this interview is important if we are to understand how the BGP can help us better grasp the political economy of the internet, and how data transits globally, as the autonomous systems are the agents behind those decisions. Examining the data from the world’s largest IXPs, I found one Brazilian autonomous system interconnected with the DE-CIX in Frankfurt in 2017, my second research site. This finding went against the predominant ideas about the IXPs, which see their capacity to keep traffic local and avoid international transit as their key characteristic. Seeing Brazilian ISPs interconnected with Europe, despite Brazil having over thirty IXPs raises the question: is the South dependent on the infrastructure of the North in decolonial terms (Mignolo 2002)?

In line with studies on code and algorithms (Ananny and Crawford 2018; Kitchin and Dodge 2011), the study of internet network interconnections shows that the code needs to be understood as part of an assemblage, which, in the case studied here, includes network engineers, internet community specialists, states, traffic exchange points, internet providers, transit providers, routers, switches, and internet users. Each

---

4 In this context, being public means that governance is not private. In Brazil, it is multistakeholder, conducted by entities from academia, business, government, and civil society through the CGI.br. The DE-CIX, on the other hand, is a private, for-profit IXP governed by the Association of the Internet Industry (also known as eco).
actor in that assemblage opens up to other actors, as an actor-network, and along with that sprawl come fiber optic cables, undersea cables, and data centers, among other artifacts. This “constellation” of bodies (Deleuze and Guattari 1987) is the result of the ethnographer’s situated focus and research questions, and will always be contingent. In other words, “…[T]hese bodies only appear to be in close proximity due to a particular act of imaginative gathering and the angle of our vision through space.” And, through the “constant movement” of researchers and their subjects, “the relationship and angle keep shifting” (Slack and Wise 2014, 156).

Scholars have long associated ethnography and assemblage (Marcus and Saka 2006). The importance of this association in BGP ethnography involves situating the code so as to grasp it in context, unlike in data analytics methods where quantitative techniques prevail over social meaning. Social understanding enables us to progress in our comprehension of the materiality of code not as an isolated object, but as part of a collective, embodied in institutions and communities acting with purpose. Action is always a shared endeavor, a “hybrid agency” (Abbate 2012), distributed along a chain of relations between humans and non-humans, and each actor should be understood in terms of its attributes and characteristics symmetrically and in context. In this sense, “[p]urposeful action and intentionality may not be properties of objects, but they are not properties of humans either. They are the properties of institutions, of apparatuses, of what Foucault called dispositifs” (Latour 1999, 192). Unveiling the sharing of action in the assemblage of a code means revealing its materiality and technopolitics.

On the Technique: Code Literacy

Code is language, and the grammar of code can be unveiled and understood. As the grammar of BGP allows to understand, while sets of agreed-upon conventions (Cerf and Kahn 1974, 1), internet protocols are languages underlying contemporary communication. As Carolina Israel puts it, “Protocols are a type of language that exerts defining force on the behavior of digital data (…)” (Israel 2022, 283). They are also policies in themselves, given their impact on the flow of online information, capacity for control in distributed networks, and design, which enables them to serve as an alternative form of public policy (DeNardis 2009; Galloway 2004).

Approaching code requires some level of literacy, and much like the construct of an assemblage, code literacy deeply relates to where the individual is situated. It is ultimately plural and focused on meaning, as it has been discussed within data literacies (Fotopoulou 2020). Building on our previous research on digital literacy (Rosa and Dias 2021), code literacy can be understood as the condition that allows the individual (e.g. the researcher, the activist, the user, etc.), to interact with code to respond to their social needs. It is a result of multiple skills but also lived experiences. For instance, all indirect interactions I had with internet protocols before my fieldwork on BGP, including through technical jobs, internet use, and readings contributed to my tacit knowledge about it.

Because literacy is situated, it is also functional. In the area of education, functional literacy has less to do with the mechanics of language (mastery of its formal
aspects) than with its social practices and its use in context (Rosa and Dias 2021; Soares 2006). This distinction has been historically important to detach literacy from formal education, in order to not disregard literacy of people who can functionally use language to communicate in their social environments despite unequal or no access to school. This is key as we advance in participatory design projects with code or in recognizing, for instance, the labor of Indigenous people as internet codesign (Rosa 2022). On the other hand, while coding or computer programming skills (Vee 2017) can contribute to code literacy, code ethnography does not depend on such skills. Instead, it relies on a situated understanding. In developing BGP literacy, I have focused on BGP mechanology (what it is producing) using a myriad of practices (see Table 1) as we commonly do to understand our subjects and objects within ethnography. Other possibilities, which might require different practices, target, for example, the ontology of the code (what it is) and its archaeology (where it came from) (Berry 2015).

To collect code, and elucidate the relations between the global North and South, following the principle of symmetry (Callon and Latour 1981) I conducted the BGP ethnographic fieldwork at the largest IXPs in each region, creating a “network field” (Burrell 2009). At the time the data was collected, in 2019, the greatest amount of traffic that circulated on the world’s over 700 IXPs was DE-CIX Frankfurt in Germany, with 4.28 terabits per second (Tbps) average throughput. IX.br São Paulo is the largest IXP in the global South with an average peak data rate of 2.99 Tbps in 2019.

Conversations and interviews with BGP protocol specialists afforded an understanding of the basis underlying the functions of the routers and helped train the ethnographic lens on the code. Additionally, the code was analyzed and interpreted as a proxy for internet information flow, which would have been otherwise inaccessible due to the widely privatized ownership of information circulation infrastructure of the internet. Tiago Felipe Gonçalves, a network engineer who has worked at various IXPs worldwide, performed the processing of BGP. He quantified two BGP fields, connected networks and origin networks, and associated the countries with the networks based on two well-known databases, Cymrus and WHOIS. I then crossed that data with the databases of the World Bank and United Nations in order to operationalize the concept of global North and South (for details see Rosa and Hauge 2021).

Regarding the collection of code, it is important to mention the barriers that code privately administered can create. For instance, the data from IX.br and DE-CIX Frankfurt had to be collected differently. Data for the former was downloaded from its route server, which, unlike that of DE-CIX, makes that information available to download. To gather data from DE-CIX, the alternative was to use the data the Packet Clearing House, which collects BGP routes snapshots from all the IXPs it connects to and makes it public. The decision on this alternative was made after discussions with the technical community. Given the private manner in which the DE-CIX data are handled, collecting them from the PCH router rather than directly from their route server was the best way—second best, that is—to ensure the viability of a

---

5 The number of IXPs in the world is unknown. The public and online data bases most used to determine the number are EuroIX, Packet Clearing House, PeeringDB, and Telegeography.
Techniques |
--- |
Analysis of legal, historical, and technical documents |
Analysis of literature from other fields |
Analysis of public databases |
Code collection |
In-depth interviews |

| Techniques | Description |
|---|---|
| Analysis of legal, historical, and technical documents | What are the key documents to any understanding of the code under study? In the case of BGP, the Requests for Comments (RFCs), which are used by the technical community on the internet to propose and establish standards, play a fundamental role. In addition, company directives, supplier tutorials, IXP reports and regulations issued in relation to internet interconnection or the internet in general have been also fundamental (examples: Baker 1995; Cisco n.d.; Clark 1982; Rekhter, Li, and Hares 2006). |
| Analysis of literature from other fields | In which venues and communicative spaces is such code discussed? Network engineering and computer science scholars share their research mostly through conference proceedings. In the study of BGP, added to debates on BGP and internet routing (e.g., Karlin et al. 2009; Testart et al. 2019), IXPs served as the physical gateway into it, with a growing body of literature (e.g., Carisimo et al. 2015; Chatzis et al. 2013; Fanou et al. 2017; Gill et al. 2008; Labovitz et al. 2010). |
| Analysis of public databases | Are there datasets available for conducting secondary data research? In internet interconnection, which BGP structures, for data to circulate online, commercial interconnection agreements need to be made between internet providers (like the one that serves your home or cellphone network) and content providers (your email provider, for example). For this to happen, these actors need to be physically interconnected at their respective geographical locations. The online databases used to identify public interconnection sites, such as IXPs, have been important to understand the physical BGP infrastructure. PeeringDB is the best known networks database, but Euro-IX, Pack-et Clearing House (PCH) and Telegeography were also analyzed. |
| Code collection | One of the characteristics of ethnography within the digital environment is that it can often be done remotely, even if this might seem methodologically counterintuitive, given ethnography’s association with “going into the field.” For example, the BGP collection was done online, just as occurs whenever the user downloads a database. However, it is important to note that the data represents localized information about IXPs in Germany and Brazil, and it links up to various other data collected in the ethnographic field through interviews, online information gathering about these nodes, visits in loco, etc. |
| In-depth interviews | This technique allows the ethnographer to understand the assemblage being studied from the perspective of its participants. In-depth interviews are also a good indicator of field saturation (Lingel and Gray 2018), when convergences arise between the findings from different interlocutors, rendering counterpositions and arguments all the more consistent. In this study, interviews were conducted in person and remotely, by voice, email, or social media. |

Table 1 Code Literacy Techniques Applied in this Study
While “ethnography is an art of the possible, and it may be better to have some of it than none at all” (Hannerz 2003, 212), in which concerns internet interconnection, there is data that should be considered of public interest and be regulated as such, as it affects competition among telecommunication incumbents and small ISPs, especially in the global South (Rosa and Hauge 2021). In bringing code to the fore, code ethnography research may join efforts to provoke critical considerations of privatized areas of internet governance.

### On Materiality: What the Code has to Say

In this final section, the aim is to analyze the materiality of the code by expanding our understanding on the interdependency between the global South and North in terms of internet interconnection infrastructure. This is the point where code ethnography deepens digital ethnography, dialoguing with software and critical code studies (Chun 2011; Marino 2020), where, especially in the latter, the emphasis on the interpretation of, and consequent interaction with, code is fundamental.

Figure 1 is a visual representation of a “prefix” announcement on BGP as a way to both materialize and explain what kind of data is the input for the analysis that follows. Figure 1 contains information collected at an IXP route server, and is presented

---

6 The IX.br Looking Glass where its data is available is: https://lg.ix.br/. The directory with PCH data for March 2019 is: https://www.pch.net/resources/Routing_Data/IPv4_daily_snapshots/2019/03/route-collector.fra.pch.net/. Analysis is based on IPv4 (Internet Protocol version four), the most used on the internet to date.
as it is seen from the standpoint of an internet network router connected to an IXP. I describe below how to read this data.

In the first column, the prefix “1.0.6.0/24” has been announced. A prefix is a block of addresses that routers announce in a BGP speaking session when communicating with other routers to share which addresses they can reach. “1.0.6.0” identifies the network, “/24” indicates the block has 256 IP (Internet Protocol) addresses, each of which may be a specific computer host to be reached, or a destination of data packets sent in that direction. In the second column, the network routers that know how to reach such prefix and are announcing them are identified. In total two different routers are announcing it, the first “80.81.192.30” has announced one route to reach the prefix, while the network router “80.81.192.172” knows three options (paths) to reach it. The routes are given on the right, in four columns containing autonomous system numbers indicate four paths available to reach that prefix—they are the AS_PATH attribute of BGP. The autonomous system numbers in red rectangle are the networks connected to the IXP. The autonomous system numbers in yellow rectangle are the origin networks who own the prefix announced. Importantly, they have customer-provider relations with the autonomous systems on their left, which also have customer-provider relations with their left autonomous systems and so on. Through this, one can understand that the autonomous systems connected to the IXP bring with them these commercially established relations. Once an autonomous system connects to a certain IXP, and peers with autonomous systems also connected there, they are also connected to the chain of customer-provider relations that their peers bring. This increases their connectivity and optimizes their paths to more addresses, or prefixes, which ultimately brings economic and competitive advantages. The best AS_PATH selected in this example is the last one, which contains a “>” symbol. Although three out of the four paths seem to be identical, indicating to traverse the exact same autonomous systems in order to reach that prefix, they are not the same if they have different attributes, as for example, different routers under them, generating differences in the paths, not visible in the illustrative BGP script. Considering its affordances and

Fig. 1 A Border Gateway Protocol Script, Source: Extracted from PCH Routing Data
design, the question I put to the BGP, the lingua franca in which internet networks communicate, is: do network routers connected to an IXP in the North and connected to an IXP in the South say different things when speaking through one node or the other? The answer is yes. First, comparing the number of autonomous systems peering on both of the studied IXPs, IX.br São Paulo has almost twice as many connected networks as the IXP in Frankfurt (DE-CIX)—1,720 versus 879. In fact, IX.br is the world’s largest IXP in terms of quantity of participant networks at the time of writing. This is due in part to the large number of autonomous networks registered in Brazil but also to the IXP’s not-for-profit business model; until recently, its services to autonomous systems were rendered free of charge, and its present rates are much lower than those practiced by commercial IXPs worldwide (IX.br São Paulo charges USD 130 per 10 Gigabits per second port, roughly 690 reais, while the average for the US and Mexico is USD 2,400 per 10 Gbps port). As shown in Fig. 2, the São Paulo node is extremely attractive in terms of BGP routes in Brazil, most of which come from other states across the country (64.6%, or 1,095). This is because the main internet content providers, such as Amazon, Facebook, and Google are at IX.br São Paulo. As explained by the founder of the social organization Associação Nacional para Inclusão Digital (National Association for Digital Inclusion), in the Brazilian Northeast region, they have worked:

…to bring down the cost of internet infrastructure and make the small internet providers more independent of the telecom companies [that provide transit]. After the IXPs opened [in the Northeast], they spent years exchanging little traffic amongst themselves… So, an alternative way to bring down the cost of broadband would be to exchange traffic in São Paulo… We added 80 new participants in São Paulo because there was already content there, Google, Globo [national open and cable company], while there was none elsewhere. (personal communication, own translation)

For small providers, having direct access to these behemoths on an IXP means being able to peer with them and so avoid having to pay other providers for that access via transit. Thus the physical location of the major content providers, or in which IXPs they are based, confers enormous power, allowing them to shape the flows of information online, because, wherever they are, they draw new participants to the IXPs. Julio Sirota, the infrastructure manager at IX.br told me that “the three largest content providers generally represent 60% of traffic on the IX in São Paulo” (personal communication).

If we look at the data on the nationality of the networks interconnected through the two IXPs under study here, we will see that, while IX.br is largely a national node serving Brazilian networks (96%), the DE-CIX is highly internationalized, with 70% of the connected networks located abroad (see Figs. 2 and 3).

7 Networks connected to IX.br SP and DE-CIX Frankfurt. Source: IXP sites, March 24, 2019.
8 Brazil is the second-largest country in terms of ASNs, with 8,814 versus the United States’ 29,899 (IPInfo.io 2021).
Fig. 2 Level of International Appeal of IX.br

Fig. 3 Level of International Appeal of DE-CIX
The different levels of international attractiveness between the largest IXPs in the global South and global North, and also regionally, as in the Brazilian context, raise questions about the current roles of the IXPs within the internet ecosystem. While these structures were originally built to facilitate the exchange of data locally, keeping internet traffic local and with low latency, and therefore helping to boost the quality of the internet services provided, the research results indicate a continued flow of information from the global South to the global North.

As the DE-CIX is located in the North and the IX.br in the South, the majority of the networks connected to them would likely be from the same regions, and this is in fact the case, as shown in Fig. 4. However, the volume of Southern networks drawn to DE-CIX is far higher (23.2%, or 204 networks) than the Northern networks attracted to IX.br (3.3%, or 57 networks).

While DE-CIX Frankfurt has become a major internet node on an international level, IX.br São Paulo has become major nationally. We might name them giant internet nodes to call attention to their social, economic, and political implications in data circulation infrastructure. In the Brazilian case, the IXP ecosystem that operates on a not-for-profit basis was created with an eye on the regional development of the internet in the country, in which IXPs ought to serve their respective regions equally. However, that vision collides with the political challenges of an internet that is highly concentrated on the web layer of international content providers that have no commercial interest in smaller IXPs, located in other regions. Hence internet providers from other parts of the country and continent increase their costs because they have to transport data through IX.br São Paulo, quite unlike the local internet providers. In this context, the BGP grammar, based on the shortest possible AS_PATH, sustains the

![Networks Connected to DE-CIX Fra and IX.br SP per Global North and Global South](image)

**Fig. 4** Networks Connected to DE-CIX Fra and IX.br SP per Global North and Global South, Source: IXP sites, March 24, 2019
logic of the economy of interconnection and contributes towards this centralization and the maintenance of regional imbalances.

On the international level, the concentration dynamic is more evident (Fig. 5). Here, rather than analyze the autonomous systems that peer up at the IXPs under study, we see the autonomous systems that they “announce” (origin networks) there, and the addresses, or prefixes, reachable through these networks. Once announced, these addresses become accessible to the participants connected to the IXP via peering, thus increasing their connectivity. The numbers for Frankfurt and São Paulo are very different. They show the power of DE-CIX, which has almost five times the number of addresses shared by the participants on IX.br. On one hand, this explains, materially, why Brazilian internet providers have opted to exchange traffic in Germany—as they increase their connectivity, they improve their AS_PATH options and economic returns. On the other, it shows the levels of centralization and unequal data flows on the internet, previously undocumented in the literature. This data reveals a clear infrasstructural interdependency between the global North giant internet nodes and the global South internet service providers. So many Brazilian internet providers connect to the IXP in Germany is precisely because of the disparity between Northern and Southern connectivity levels, and this lack of resources in the South in terms of connectivity and presence of content providers continues to make the giant nodes bigger. So while São Paulo is the “North” to the rest of Brazil, Germany plays the same role for the global South, illustrating forms of coloniality of power (Quijano 2007) embedded in digital infrastructures.
Conclusions

The results of the code ethnography presented herein should be understood as a snapshot of a highly dynamic setting, where digital ethnography open up space for us to unveil classical sociological questions related to power, dependencies, and inequalities now mediated by digital infrastructures. They show hidden inequalities in internet interconnection dynamics and the possible consequences of information flows gravitating towards national and international internet nodes, especially if we think in terms of the data of citizens of the global South circulating in the North under Northern jurisdiction. Even with progress in the legal protections of personal data on a national level—such as Brazil’s Personal Data Protection Law—sensitive data from citizens in the global South continue to circulate in other nations where they will be handled in accordance with local law. The physical attributes of the internet are therefore just as important as the logical, and need to be investigated more thoroughly.

The concentration of power on the internet goes beyond what we have intensely seen on platform studies. The power of content providers and the BGP grammar has generated inequities on an infrastructural level that are firmly stacked against the global South. If the data of Brazilian citizens journeys to Germany or elsewhere because of decisions taken by autonomous systems, how can we rethink our public policies so that the public interest can prevail over the private? This is the kind of question that code ethnography begins to address by placing BGP, as an agent, in context.

The study of internet interconnection reveals that to advance our understanding of the digital with ethnography, attention to its materiality and affordances is also crucial. Given that communication mediated by technologies is always a shared action, digital infrastructures are always there to be unveiled. Such a task, obviously, has limitations. Leading computer operational systems depend on over 50 million lines of code, while leading conglomerates that include major search engines are built over 2 billion lines (Metz 2015). The main challenges come, however, not only from the magnitude of these numbers, but from the privatized way that code is created and maintained, and how complexly it is transformed over the time. Characteristics such as these, described as opacity (Burrell 2016; Christin 2020), raise barriers that cannot be ignored or dismissed. While regulation is necessary to circumvent them, to better navigate these challenges, a code ethnographer should look at logical and physical dimensions in relation to each other. The physical infrastructures (e.g. IXPs) co-produces and, at the same time, help elucidate the logical components (e.g. BGP). Physical artifacts can also serve as gateways to get access to the code.

The inequalities built into the internet infrastructure call attention to the layers of materiality that a political economy of the internet can plumb so as to take a critical look at technical systems that are, in principle, inaccessible. In digitally mediated communication, the inductive and qualitative methods of the social sciences applied to any code broaden the horizons of investigation into the social relations of power. Clearly, there are no formulas for an ethnography of code, other than the paths toward building new knowledge and new research possibilities.
Acknowledgements The author would like to thank all the interviewees, Tiago Gonçalves, Francisco Badaró, NIC.br and Packet Clearing House for their invaluable contributions and collaborations, as well as friends and colleagues from the Rede de Pesquisa em Governança da Internet (REDE) and the Annenberg School for Communication’s Center for Advanced Research in Global Communication (CARGC) for engaging discussions on previous versions of this paper.

Funding This research was possible thanks to Columbia University’s School of International and Public Affairs and Carnegie Corporation of New York Next Generation Cyber Fellowship (2016–2017).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Abbate, Janet. 2012. L’histoire de l’Internet au prisme des STS, Internet history through an STS lens. *Le Temps des médias* 18 (June): 170–80. https://doi.org/10.3917/tdm.018.0170.

Abidin, Crystal, and Gabriele de Seta. 2020. Doing digital ethnography: Private messages from the field. *Journal of Digital Social Research* 2 (1): 1–97.

Adachi, Tomi. 2011. Comitê Gestor da Internet no Brasil (CGI.br): Uma evolução do Sistema de Informação Nacional moldada socialmente (Brazilian Internet Steering Committee: An evolution of the National System of Information). Doctoral Dissertation, Department of Administration, São Paulo, Brasil: Universidade de São Paulo. https://www.teses.usp.br/teses/disponiveis/12/12139/tde-10102011-165732/publico/TomiAdachi.pdf. Accessed on 09 July 2022.

Ananny, Mike, and Kate Crawford. 2018. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society* 20 (3): 973–89. https://doi.org/10.1177/1461444816676645.

Anzaldúa, Gloria. 2012. *Borderlands / La frontera: The new Mestiza*. 4th Edition. San Francisco: Aunt Lute Books.

Baker, F. 1995. RFC 1812 - Requirements for IP version 4 routers. Internet Engineering Task Force (IETF). https://tools.ietf.org/html/rfc1812. Accessed on 09 July 2022.

Berry, David M. 2015. *The philosophy of software: Code and mediation in the digital age*. 1st Edition. Basingstoke; New York: Palgrave Macmillan.

Brito, Samuel Henrique, and Bucke, Mateus Augusto Silva Santos, Ramon dos Reis Fontes, Danny Alex Lachos Perez, Hirley Dayan Lourenço da Silva, and Christian Rodolfo Esteve Rothenberg. 2016. An analysis of the largest national ecosystem of public Internet EXchange Points: The case of Brazil. *Journal of Communication and Information Systems* 31 (1): 256–71. https://doi.org/10.14209/jcis.2016.23.

Brock, Kevin. 2019. *Rhetorical code studies: Discovering arguments in and around code*. Ann Arbor: University of Michigan Press.

Broussard, Meredith. 2019. *Artificial unintelligence: How computers misunderstand the world*. 1st Edition. Cambridge, MA: MIT Press.

Bruno, Fernanda, Bruno Cardoso, Marta Kanashiro, Luciana Guilhon, and Lucas Melgaço, eds. 2018. *Tecnopolíticas da vigilância: Perspectivas da margem* (Technopolitics of Surveillance: Perspectives from the margin). 1st Edition. São Paulo, SP: Boitempo.

Burrell, Jenna. 2009. The field site as a network: A strategy for locating ethnographic research. *Field Methods* 21 (2): 181–99. https://doi.org/10.1177/1525822X08329699.

Burrell, Jenna. 2012. *Invisible users: Youth in the Internet cafés of urban Ghana*. Cambridge, MA: The MIT Press.
Vicentin, Diego Jair. 2016. A reticulação da banda larga móvel: definindo padrões, informando a rede. Doctoral Dissertation, Department of Sociology, Campinas, São Paulo: Universidade Estadual de Campinas. http://repositorio.unicamp.br/jspui/handle/REPOSIP/320968. Accessed on 09 July 2022.

Woodcock, Bill. 2013. On Internet, Brazil is beating US at its own game. Al Jazeera, September 20. http://america.aljazeera.com/articles/2013/9/20/brazil-internet-dilmarousseffnsa.html. Accessed on 25 April 2021.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Fernanda R. Rosa** Fernanda R. Rosa is an Assistant Professor of Science, Technology, and Society at Virginia Tech. Her research interests include Internet governance and design, social justice, and the global South.