

RESEARCH ARTICLE

Latent and explicit mnemonic communities on social media: studying digital memory formation through hashtag co-occurrence analysis

Robbert-Jan Adriaansen 

Center for Historical Culture, Erasmus University Rotterdam, Rotterdam, the Netherlands
Email: adriaansen@eshcc.eur.nl

(Received 30 May 2023; revised 29 March 2024; accepted 4 April 2024)

Abstract

This article explores the nature and dynamics of mnemonic communities within the context of social media platforms and proposes to identify mnemonic communities using hashtag co-occurrence analysis. The article distinguishes between ‘explicit’ and ‘latent’ mnemonic communities, arguing that while some digital mnemonic communities may exhibit characteristics of offline communities, others exist latently as discursive spaces or semiospheres without direct awareness. On platforms like Instagram, hashtags function as semiotic markers, but also as user-chosen indexes to the content. As hashtags link the social and semantic aspects of community formation, hashtag co-occurrence analysis offers a robust framework for understanding and mapping these communities. This method allows to detect and analyse patterns of hashtag use that suggest the presence of networked community structures that may not be apparent or conscious to the social media users themselves. Additionally, a metric is introduced for determining the degree of ‘latentness’ of communities that quantifies the cohesion within communities compared to their external connections. The article demonstrates this approach by applying hashtag co-occurrence analysis to a dataset of Instagram posts tagged with #Juneteenth, a popular hashtag used to commemorate the ending of slavery in the United States. It identifies 87 mnemonic communities that reflect the diversity and complexity of how platforms facilitate memory-sharing practices and the role of semiotic markers in forming (latent) mnemonic networks.

Keywords: digital memory; hashtags; Instagram; Juneteenth; mnemonic community

Introduction

One of the difficulties scholars in digital memory studies face is the apparent fluidity of social formations and social media networks. A core axiom in memory studies is that the groups that uphold and maintain memory and derive from it a sense of identity and shared belonging are in fact communicative communities (Assmann 2008; Pentzold *et al.* 2022). Memory is understood as a social practice, occurring communicatively through collective exchange about what and how to remember. It is a contemporary, reconstructive (and recursive) process through which selective, value-laden, temporally and spatially limited, collectively and individually identity-forming memories are expressed and actualised (Erl 2017, p. 6).

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

However, the digitisation of communication processes and the rise of social media complicate the identification and understanding of the ‘communities’ within which communicative remembering takes place online, or vice versa, how communities are constituted through shared communication about the past. Digital memory formation should not be viewed in the context of traditional, geographically, and culturally bounded communities, but rather in relation to a more dynamic, often virtual, network of individuals engaged in similar semiotic practices or sharing similar experiential or mnemonic patterns. In this context, Hoskins proposed the concept of ‘networked memory’ (Hoskins 2009) to capture the idea that memory in the digital age is distributed across networks, rather than being anchored in specific location-based communities. It emphasises that digital memory is embedded in and responsive to the sociotechnical practices of digital media users, and therefore relies on the affordances of the platforms, which can dynamically add to, alter, and erase memory.

Such networked approaches to memory pose both theoretical and methodological challenges. Theoretically, we have to consider how to conceptualise the communicative ‘communities’ in digital contexts where memory can be transient, fragmented, and often co-constructed by individuals who may never interact in physical spaces. We may even wonder whether it is still meaningful to speak of communities, or whether concepts like ‘networked publics’ (boyd 2010), ‘imagined audiences’ (Marwick and boyd 2011), or ‘refracted publics’ (Abidin 2021) may be more relevant to describe the social dynamics of digital communication in pseudonymous networked environments. Methodologically, we must consider how to identify communicative mnemonic ‘communities’ in networked contexts where we may not assume that they are simply digital variants of existing ‘offline’ communities, and where an abundance of data and the fluidity of associations make it difficult to discern clear boundaries and memberships.

The main aim of this article is to propose a *digital methodology* for identifying online communities, through network analysis of hashtag co-occurrence. To achieve this, the article will also provide theoretical clarification on the nature of mnemonic communities on social media, identifying the relations and discrepancies between social or cultural concepts of online communities with social network analysis to form a conceptual framework for the methodology. Furthermore, the article will distinguish between ‘explicit’ and ‘latent’ mnemonic communities, emphasising that on social media platforms, users can be part of a communicative mnemonic community influenced by a platform’s discursive and semiotic structures ‘latently’, without the users’ awareness or conceptualisation of the community. This highlights the relative ‘agency’ of hashtags in community formation.

To demonstrate the practical application of the proposed methodology, the article provides an implementation in Python code, that uses a dataset of Instagram posts tagged with #Juneteenth; a vastly popular Instagram hashtag that continues to gain significance in the digital space for commemorating the end of slavery in the United States. This hashtag represents Juneteenth National Independence Day, a US federal holiday since 2021, which marks 19 June 1865, when the Emancipation Proclamation was enforced in Texas. The use of an actual dataset, as opposed to synthetic data, is crucial for showcasing how to navigate methodological challenges, such as handling an overly large dataset or one biased towards a specific hashtag due to data collection methods. #Juneteenth is a relevant hashtag to use to exemplify the method, as it can be assumed to be used in various contexts by many different groups that range from traditional mnemonic communities with a specific geographical base or formal organisational form to virtual communities that lack these markers. Choosing a very popular hashtag also allows to discuss methodological strategies for working with large and noisy datasets without prior assumptions about specific existing communities. Accordingly, this article does not aim to conduct a comprehensive empirical analysis of #Juneteenth, maintaining its focus on

the use of digital methods for community detection. This article is accompanied by a notebook that provides an annotated code implementation in Python with additional and more in-depth technical explanation of hashtag co-occurrence analysis. The notebook can be accessed on CoCalc <https://www.cambridge.org/S2635023824000079/MEM-Notebooks>.

Memory in new media ecologies

In memory studies, there is widespread agreement on the need to move beyond seeing mnemonic communities – such as families, nations (Nora 1996), or cities (Halbwachs 1992) – as pre-existing social frames in which memory-making takes place. Influenced by Maurice Halbwachs's notion of 'collective memory' (1980), early discussions highlighted that individual memories are shaped within social contexts. These discussions foregrounded the notion of 'community', arguing that while communities carry and shape memory collectively, they are also constituted by shared memories in turn. Jan Assmann (2007, p. 30), for example, argued in this context that communities are always mnemonic communities, *Gedächtnisgemeinschaften*, since they derive their identity from the memories that their members share.

The Halbwachsian framework, which assumes a 'framed-ness' (Erl 2011, p. 10) of collective memory in social formations has been challenged increasingly over the last decade, and new concepts have been proposed that consider the complexity and fluidity of memory across diverse communities. Concepts like 'traveling memory' denote the wandering of carriers, media, contents, forms, and practices of memory across political, cultural, and linguistic boundaries (Erl 2011); 'prosthetic memory' emphasises how forms of mass culture facilitated the assimilation of personal experience with shared collective memories beyond traditional communal, gender and ethnic distinctions (Landsberg 2004); 'multidirectional memory' refers to the idea that memory is not limited to a single, fixed direction or meaning, but is instead dynamic, interconnected, and multi-layered (Rothberg 2009); and 'palimpsestic memory' stresses the processes in which memory is constantly being reinterpreted and rewritten, both at the individual and cultural levels (Silverman 2013). All these approaches aim to 'transcend the 'container'' (Olick 2014, p. 23) – the traditional boundaries and limitations of collective memory within existing social formations.

Transcending the paradigm of community-encapsulated memory has been particularly urgent for digital memory studies, as new media ecologies have introduced novel ways of remembering, sharing, and experiencing the past that challenge the traditional entanglement of memory and community. Hoskins (2017) argues that the need for new approaches in memory studies is driven not only by theoretical shortcomings in the Halbwachsian structuralist framework of collective memory but also by the transformative nature of new media ecologies. These ecologies differ significantly from the broadcast era, which relied strongly on gatekeeper–audience relationships and the idea of memory as something produced for receptive consumers (Kansteiner 2002). In contrast, memory within new media ecologies is characterised by a dialogical character of communication, where publics now have more means to reply, comment, edit, share, and recontextualise memory. This makes memory less coherent and shattered, but also simultaneous and searchable. It also repositions the archive as something no longer kept by institutional gatekeepers but merged into platforms. Platforms keep track of all users' activities in the form of shadow archives that constitute a documented record of online representations and engagements (Hoskins 2017).

This does not mean that humans are out of scope in digital memory – far from it – but as digital memory has become 'datafied' (Smit 2022), and turned into objects of algorithmic processing, memory formation increasingly becomes a matter of interaction with and

within these processes. A growing body of scholarship on ‘filter bubbles’ and ‘echo chambers’ highlights how social media platforms amplify specific messages and viewpoints (Gunn 2021; Jacobsen and Beer 2021a) as an effect of platform-based algorithmic curation (Terren and Borge-Bravo 2021) and distribution of content, combined with user behaviour and preferences.

The technological affordances of social media platforms also affect the relationship between memory and forgetting. In the context of new media ecologies, the difference between communicative and cultural memory collapses (Pentzold *et al.* 2022), as the technologies that shape the new media ecologies enable the preservation and dissemination of vast amounts of information and experiences beyond what was previously possible. With seemingly unlimited (cloud) storage capacities the question of memory is less about what can be remembered and more about what should be remembered (Menke and Birkner 2022), and the traditional gatekeepers of memory are increasingly less able to govern that.

Unlike the collective memory paradigm, where forgetting is often seen as memory loss, digital memory studies aims to reconceptualise forgetting as an active mechanism that allows memory to focus on specific contents to incorporate them into a meaningful ‘internal horizon of references and recursions to face the present.’ (Esposito 2017) The idea of forgetting as a function of mnemonic attention mechanisms rather than as the disappearance of memories fits the digital context well – not only because of the seemingly unlimited possibilities of data storage but also because in the digital context algorithms are largely responsible for the selection and prioritisation of recollections and information – ranging from answering search queries submitted to google to generating automated, unsolicited trips down memory lane by social media apps that synthesise ‘memories’ from past social media posts or stored photos (Jacobsen and Beer 2021a). But at the same time algorithms structure forgetting by deprioritising and obfuscating memories – either as an effect of technological decisions or on purpose as part of a specific agenda or content moderation policy.

These attention mechanisms put social media platforms in powerful and responsible positions when it comes to content distribution and moderation. Smit (2022) argues that the ‘platformisation of memory’ denotes how memory becomes contingent on platforms. This is not only a process of society becoming increasingly dependent on social media platforms for its communication and its mnemonics, but also because platforms as commercial and ideologically non-neutral meta-media transform memories and experiences into commodified quantifiable data, which they are then able to process algorithmically. This quantification and algorithmic processing affect both the shaping and construction of memory and the visibility and circulation of memory. As Jacobsen and Beer (2021b) demonstrate, social media users indicate that the quantitative properties of ‘datafied’ memory incentivise them to attach a specific value to the personal memories they share online. The number of likes and comments affects how they value their own memories and affects what and how they will post future recollections. However, such metrics also contribute to the accumulation of biographical content on platforms and an idea of competition for attention with others. In that sense, social media users also consciously act upon the either real or perceived affordances of the platforms, including their perceptions of how to create or label the content to reach larger ‘imagined audiences’ (Litt and Hargittai 2016; Marwick and boyd 2011) beyond their followers. After all, social media users are no passive recipients of algorithmically curated content. They can be understood as active curators of memories and historical representations on their own profiles, pages and regarding their own content (Adriaansen 2022a), and as such they actively engage with the platform affordances to enlarge chances for attracting views and generating virality for their content.

Conceptualising mnemonic community in new media ecologies

This implies that digital technologies do not somehow overcome issues of power and hegemony through some form of networked democratisation, but transform and reshape discursive and social structures. This brings into question whether the concept of ‘community’ can still be applied to digital contexts in the same way as traditional offline communities. The answer is not unequivocal, partly because ‘community’ is an ‘essentially contested concept’ (Gallie 1968), and its application in digital contexts adds layers of complexity. In general, we can distinguish between two approaches to understanding mnemonic community in the context of new media ecologies. The first highlights the social and cultural dimension of community formation and highlights that communities need not be bound geographically but do engage in similar mnemonic patterns or practices. The second understands community as a network as discussed in social network analysis. We will see that while these appear to be incommensurable, a network approach to digital mnemonic community formation does not exclude constructivist perspectives, but should be extended to include the agency of platform affordances, such as hashtags.

First, the social scientific or humanities approach. As mentioned in the introduction, using ‘community’ to denote social structures in new media ecologies is not always obvious. Qualitative studies of new media ecologies are often grappling with the terminology to use for online social structures, leaving Nathan Rambukkana (2015, p. 2) to wonder whether they are ‘communities, publics, discourses, discursive formations, *dispositifs*, something else’. The general answer to this question is that it depends on the context which of these concepts is most applicable. ‘Community’ is often deemed relevant when groups of people can be identified who interact online through computer-mediated communication and have a shared sense of belonging, identity, and common interests. Parks (2010, p. 107) argues that this approach works since the community has lost its connotation of being geographically limited and is now generally understood as a quality of sociality, which allows to think of ‘virtual communities’ as ‘social groups that display the psychological and cultural qualities of strong community without physical proximity’. On social network sites, platform affordances like setting profiles public or private, creating group pages, following, messaging, and liking facilitate interactions for community formation (boyd 2010). However, Kozinets (2015, p. 12) reminds us that for ‘online’ or ‘virtual’ communities we should not conceive of ‘membership’ in ‘any way similar to that of communities such as those based upon race, religion, ethnicity or gender’. Communities are not fixed or stable entities, but rather a dynamic and fluid form of association that is shaped by the interests, values, practices, and – indeed – memories of its members. In this paradigm, virtual communities exist only if its members have a shared sense of belonging or identity, or in Benedict Anderson’s words, if they are able to ‘imagine’ the community as such (Anderson 2006).

As Menke has demonstrated, nostalgia can be an important driver for community formation, defining digital mnemonic communities as groups of people who share nostalgic memories of the past through online platforms, such as social network sites. These mnemonic communities can be understood as a form of post-traditional community that can foster collective identity, emotion, and solidarity among their members, while having shared interests as the common denominator (Menke 2019, p. 161). However, interests, but also shared identities, need not precede community formation, but can be shaped in the very process of community formation through a common mode of communication or a common object of nostalgia.

Such an understanding of community would be flexible, but also somewhat unstable as the flexibility raises questions about communities’ boundaries and criteria of membership, and about the dynamics of inclusion and exclusion that shape them. After all,

membership requires communities' self-identification, as communities rely on some form of awareness of belonging, or the capacity to imagine itself as a community. Answers, like Kuzinets', that argue for a scale of rigid to fluid associations, may solve these questions conceptually, but provide little guiding for empirical research, which results in scholars resorting to assumptions about which groups or environments constitute communities, and having little to no means of identifying them inductively from large amounts of social media data.

The second conceptualisation of community has less trouble defining the boundaries of online communities by primarily focusing on patterns of relations and interactions among network actors. Instead of relying on shared interests or self-awareness it defines community in terms of performative networks, where 'nodes' (entities within a network, such as individuals) have more connections with each other within the same community than with nodes external to it. This is the principle of homophily, which suggests that nodes in closer proximity within the network are more likely to share certain characteristics or engage in similar interactions (Himmelboim 2017). The structural definition of a community is based on the density of the connections between nodes. Communities are identified as sets of nodes that are strongly interconnected among themselves and less so with nodes outside the group (Chouchani and Abed 2020). For example, if we take follower networks with social media users as nodes and their links or edges defined by who follows who, communities can be identified by analysing the patterns of these connections. Nodes within a community tend to have a higher number of connections to each other compared to nodes outside the community. Given the principle of homophily these communities are likely to have similar interests as well, but this is not a criterion for the identification of a community. Hence, from the perspective of social network analysis, communities can be both 'explicit' or 'latent'. Explicit communities can be assumed in advance as they have members who are consciously aware of their participation in a specific virtual community. In contrast, latent communities can only be identified through patterns of interactions and connections that indicate a shared interest or characteristic, but without the members necessarily being aware of their collective grouping (Fani and Bagheri 2017).

In digital memory studies Andrew Hoskins (2011, p. 272) has argued for a networked approach in the context of 'connective memory' – an approach to memory in new media ecologies, in which individuals and groups 'oscillate between forming more dense and more diffused nodes in a multitude of mediatized networks'. Since Hoskins' argument is consistently built upon the desire to discern digital memory from the Halbwachsian container-approach to memory, he aims to replace the very idea of 'collective memory' with that of 'memory of the multitude.' (Hoskins 2017) Hoskins challenges traditional notions of community by emphasising the role of hyperconnectivity in shaping memory in the digital age. The multitude recentres the individual as the locus of memory while acknowledging the influence of digital connectivities and traces on the constitution and communication of memory. This fluidity and interconnectedness of individual memories suggests that our understanding of memory and communities must adapt to the realities of the digital age, where hyperconnectivity itself becomes a prime model for social configuration. In this context, hashtags exemplify how hyperconnectivity facilitates the formation of fluid, networked communities around shared interests, memories, or topics. Hashtags enable individuals to connect and engage with each other based on common themes or experiences, creating ephemeral communities that transcend traditional community markers such as geography or stable group membership. As such, hashtags play a crucial role in the construction of memory and community in the digital age, by allowing individuals to coalesce around shared mnemonic content and practices, or at least by engaging with similar contexts in a shared semiotic environment.

However, while the traditional container-approach to mnemonic communities becomes irrelevant in this context, the concept of community from social network analysis can still be applied. In fact, when Hoskins – while arguing to shift focus to understanding the complexities of mediatised connectivities – speaks of people oscillating between forming ‘more dense and more diffused nodes in a multitude of mediatised networks’, this aligns with the social network analysis perspective, where the density and centrality of nodes depend on their connections and interactions within a larger network. Dense *clusters of nodes* in social media networks are what social network analysis calls ‘communities.’ These exhibit a shared engagement with other nodes, or at least a more intense engagement than with other communities. These clusters or communities can be understood as ephemeral mnemonic subnetworks that are constantly shifting and reforming, based on the interactions and exchanges of individuals who share and engage with memories and historical representations on social media platforms. These mnemonic networks are fluid and dynamic, transcending traditional boundaries of ethnicity, nationality, and geography, and are shaped by both the intentions and actions of users as well as the underlying algorithms and affordances of the platforms on which they operate.

An important advantage of this approach is that it does not assume communities to be collective agents and offers a more fluid approach. However, it runs the danger of seeing digital memory as ‘determined by technological potential’ rather than rooted in concrete social and communicative processes (Schwarzenegger and Lohmeier 2020, p. 143). While notions of the multitude as a network of interconnected individuals who in ever-changing compositions partake in fluid and ephemeral collectives resonates with our experiences of rapid digital change, we must be wary of the risk of overemphasising the distinctiveness of digital memory formations, since in reality dichotomies between broadcast era media and new media ecologies are difficult to discern, and offline and online communicative practices and identities are often intertwined (Barassi 2016). Hence, understanding mnemonic communities solely in terms of the density of nodes in a network can make them empty, positivistically established entities when not paired with any notions of intentionality on the users’ behalf.

At the same time, a recentring on individuals as the main agents of memory, even though they partake in a variety of shifting contexts and communities, fails to capture the agency of non-human elements in memory-making and mnemonic community formation, which – in the context of new media ecologies – assumes that communities are no static or fixed entities, but rather flexible and constantly evolving webs of relationships and interactions that are mediated and facilitated by digital platforms. In such a context agency is distributed among humans and non-human entities, such as platform affordances like hashtags that can carry and convey meanings beyond user agency, and can be amplified and affected by platform algorithms (Smit 2020). Laužikas *et al.* (2018) and Laužikas and Dallas (2023) have argued that communities on social media platforms could best be understood in terms of Yuri Lotman’s (2005) notion of the semiosphere, which assumes that communication is not a matter of understanding individual signs, but of the entire semiotic space – comprised of signs, texts, symbols, knowledge, representations and including their interaction with non-semiotic elements – in which communication takes place. They argue that in the context of social media platforms, a plurality of intersecting semiospheres exist in the form of ‘multiple communities sharing a common semiotic code and way of communication, as well as common beliefs, affects and attitudes’ (Laužikas and Dallas 2023). This approach offers a fruitful context for hashtag co-occurrence analysis as it takes a broader focus on agency, which includes not only individuals, but also allows to understand semiotic elements to exert or mediate agency in relationship with authors. Hashtags are a prime example since they function as semiotic codes and indexes that enable communication and aid in establishing and signifying

connections between memory objects and users, thus creating intricate networks of meaning. Functioning within a specific semiosphere, hashtags help foster dialogues, foster inter-semiospheric relationships, and aid in creating new semiotic structures. As such, they play a pivotal role in what Lotman referred to as the creolisation process, where the intersection between two semiospheres results in the generation of a shared semiotic space. These emergent spaces represent mnemonic communities that exist at least latently, but often explicitly, and are forged, and constantly reformulated through the process of semiotic translation between distinct semiospheres.

Hashtags can play a pivotal role in rendering latent communities explicit. As demonstrated by the use of Twitter hashtags during the 2011 resignation of Hosni Mubarak in Egypt, they can foster an ‘ambient’ and ‘affective’ digital environment. Here, news, opinions, and emotions blend and circulate in real time, anchored to specific locations via hashtags. This creates a sense of solidarity and shapes collective perceptions of events and their outcomes. As Papacharissi and de Fatima Oliveira (2012) note, ‘these affective attachments create feelings of community that may either reflexively drive a movement, and/or capture users in a state of engaged passivity’. Through network analysis of hashtags, we can uncover these communities, offering insight into the dynamic, interconnected, and fluid nature of digital memory formation within and between the multitude of digital semiospheres, while maintaining that memory is rooted in concrete social interactions, as hashtags are intentionally chosen by social media users to label their posts with the intent to reach target audiences.

Hashtags and mnemonic communities

Hashtags are key communicative tools for community formation and branding, functioning as dynamic elements in social media communication. They have evolved from simple content labels on Twitter into significant social forms that have been shaped by the way people use and interact with them over time. Hashtags encapsulate shared meanings, emotions, and cultural dynamics, and play a crucial role in constructing digital publics and communities (La Rocca and Boccia Artieri 2023). Importantly, hashtags mediate memory by allowing users to index and retrieve content related to specific events, themes, or sentiments. They enable the creation of networked archives of remembrance, where individual expressions of memory are embedded in larger semiotic structures. As such, hashtags are more than mere organisational tools; they represent collective ideas, trends, and movements, reflecting and shaping public opinion and social movements.

In establishing social formations, hashtags function as semiotic markers within these participatory spaces, transcending a pre-established sense of identity and fostering engagement through shared semiotic practices. Zappavigna (2015) calls the social connections that hashtags establish ‘ambient,’ in the sense that ‘other users are potentially present within the social network, but not necessarily linked together through connections between user accounts.’ This implies that the networked communities are *latently* present in social media metadata such as hashtags, and peers in the same networks need not be visible to each other but do engage in the same semiotic contexts or semiospheres. Hashtags are key metadata to uncover latent mnemonic communities online since the practice of tagging social media posts is an intrinsically social affair, which expresses itself in a twofold way.

On the one hand, hashtags are user-generated or -selected metadata that function as indexes to social media content. Contrary to taxonomies created by experts, hashtags are chosen and created by users, and have thus been regarded ‘folksonomies,’ a term coined in 2004 by Thomas Vander Wal (2007), which refers to the user-generated organisation of content through the use of tags – or in the words of Vander Wal a ‘user-created

bottom-up categorical structure development with an emergent thesaurus.’ This decentralised mode of categorisation makes that hashtags reflect users’ personal understanding and perspective of the content they are tagging. Folksonomy differs from traditional forms of indexing and categorisation, as it has no central authority but is based on the collective knowledge and interpretations of the users. Folksonomies have been called a form of ‘emergent semantics’ in the sense that over time users agree on a common vocabulary and create decentralised semantic structures in the process of tagging content (Singer *et al.* 2014). Moreover, since the tagging happens on social media platforms for the purpose of the retrieval of content by others, it has been dubbed ‘social tagging’ (Trant 2009). The practice of social tagging is, however, intrinsically wound up with platform affordances as users may choose tags they believe will work favourably for recommendation engine algorithms (Adriaansen 2022b). Nonetheless, the use of co-occurring hashtags can be seen as an ‘organizational strategy, allowing users in a hashtag community to manage information while broadcasting it.’ (Rocheleau and Milette 2015, p. 244)

On the other hand, hashtags have linguistic meaning, as they are concepts that can contain references to historical events, memories, or interpretations thereof. For example, the hashtag #RhodesMustFall was used on Twitter as an index to tweets related to the 2015 Rhodes Must Fall movement, which aimed to remove the statue of British colonialist Cecil Rhodes from the University of Cape Town. However, this hashtag was also used as a representation of the movement and of a specific understanding of history, symbolising the demand for change and the rejection of colonialist symbols. The injunction ‘must fall’ mobilised the conceptual metaphor of falling to express both the political desire to bring the statue down, and the broader struggle for the decolonisation of memory (Frassinelli 2018).

The combination of these two aspects – the use of hashtags as metadata that function as an index to the content and as a linguistic representation of the past, or of attitudes to the past, gives hashtags a relevant advantage to study mnemonic communities on social media, as common community detection analysis methods generally focus on either of these two aspects independently. Fani and Bagheri (2017) call them, respectively, link-based and content-based approaches to community detection in social networks.

Link-based methods model social networks as a graph with users as nodes and social links as the edges linking the nodes. The example in Figure 1 shows a dummy representation of such a network with two communities represented in the differently coloured nodes. The nodes are individual social media users, the edges are the links between them. In this example the graph is bidirected, meaning that the links have directionality: the edges – the links between users – need not be reciprocal. The links could, for example, be defined in terms of who follows who. Or they could be defined in terms of who likes whose posts. In that case, the edges would also have weights, namely the number of likes. Such weights are helpful when running a community detection algorithm, as weighted edges indicate a stronger relationship between nodes. Link-based methods for community detection are intuitive since they are grounded in the principle of homophily and can be helpful in studying mnemonic communities on social media as they rely on self-established social links between people. However, they are also limited, because they do not identify communities based on shared interests, only formal ties. This can be problematic since social links ‘could be owing to sociological processes such as conformity, aspiration, and sociability or other factors such as friendship and kinship that do not necessarily point to inter-user interest similarity.’ (Fani *et al.* 2020) In other terms, people may follow each other and like each other’s posts for social reasons unrelated to shared interests.

Content-based approaches on the other hand focus on the semantic similarity of social media content using methods such as topic modelling to cluster content semantically (Fani and Bagheri 2017). These approaches differ methodologically from the graph

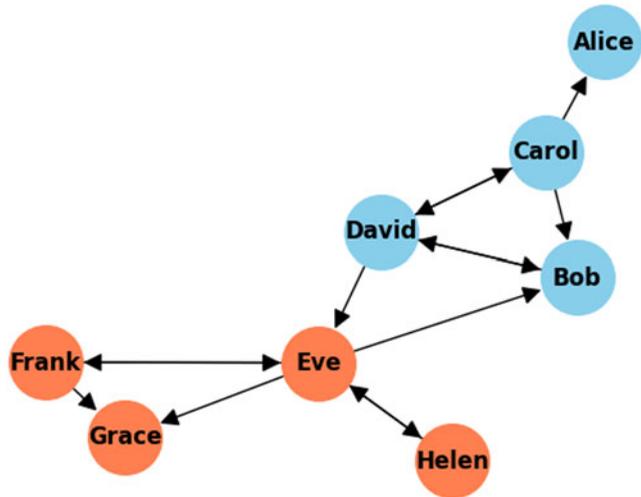


Figure 1. Illustrative link-based bidirected graph of a social media network with two communities.

approach of link-based methods, but rather use language models to identify patterns in the content itself. Although this does not take factual social links into account, it allows to identify clusters of people with similar interests. This gives good insight into the various topics that people discuss on social media, but it is a far stretch to assume communities from the topics that are being discussed.

While both approaches have valid use cases, mnemonic communities on social media networks benefit most from an approach that takes both the social and semantic dimensions into account. Hashtags therefore serve as a meaningful way to study digital mnemonic communities since they are both user-identified indexes to their content, and concepts that contain semantic information about the topics being discussed. However, studies often simply *assume* the existence of communities based around specific hashtags. Ginzarly and Srour (2022), for example, study Instagram representations of heritage in ‘hashtag communities,’ but preselect two UNESCO-launched hashtags as markers for alleged ‘virtual communities’ (Rheingold 1994). The problem with this approach is that it is entirely built upon the prior assumption of certain hashtags representing communities. This assumption is less likely to be true for hashtags launched by institutions as campaigns than hashtags that have been trending and are known to attract specific audiences. Even when studying trending hashtags that are known to facilitate specific mnemonic communities, such as #RhodesMustFall, in this way, it implies falling back to the classical ‘container’ model of collective memory in which the hashtag is seen as an instrument of an existing community, rather than the emergent network approach that integrates the study of both explicit and latent communities. It is also not useful for cases where we do not have any prior assumptions about existing communities, for example when trying to understand what mnemonic communities visitors to heritage sites engage with when representing the site on social media (Adriaansen 2021). Therefore, the question is how we can infer mnemonic communities from large quantities of social media data.

Hashtag co-occurrence analysis

This article introduces hashtag co-occurrence analysis as a simple and straightforward method to identify mnemonic communities on social media platforms. The method proposed here seeks to identify communities not based on preselected hashtags from which we assume communities a priori, but by creating an undirected graph based on the

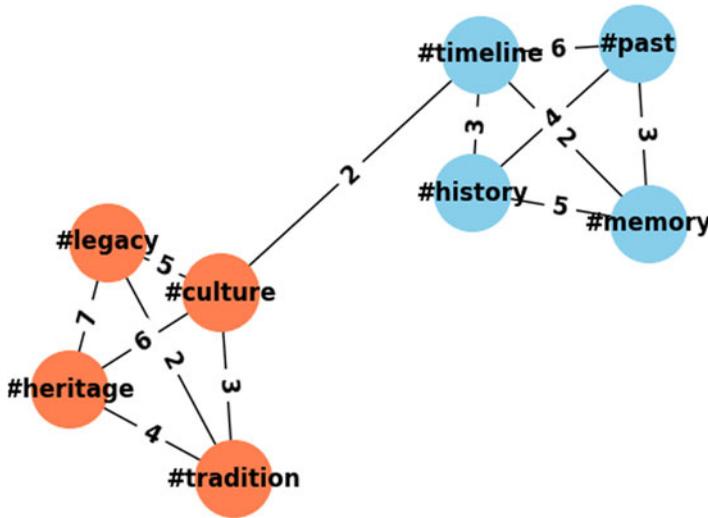


Figure 2. Illustrative hashtag co-occurrence graph.

co-occurrence of hashtags in social media posts (see Figure 2). The resulting graph network consists of hashtags as nodes, and allows to assign weight values to the edges linking the nodes. This approach makes it possible to identify the most important hashtags in the network, and to identify communities of closely linked hashtags.

Although technically hashtag co-occurrence analysis is a link-based approach, it differs from traditional link-based methods in that it takes hashtags rather than persons as nodes, and it focuses on the semantic context of the posts rather than the formal social ties between individuals. By using hashtags as the basis for the graph, the method can capture both the social dimension (since hashtags are user-generated and self-identified) and the semantic dimension (as they represent the topics being discussed). This makes hashtag co-occurrence analysis a useful tool for studying mnemonic communities on social media, as it combines the strengths of both link-based and content-based approaches. Moreover, it comes at the advantage that one does not need prior assumptions about the existence or composition of communities; instead, it allows to infer communities from the data. By linking the communities back to the individuals who engage with them through hashtag use, it is also possible to gain insight into these networks at the user level in a dynamic way, since this allows user ‘membership’ of multiple communities simultaneously.

Hashtag co-occurrence also has its limitations for the detection of (mnemonic) communities. First, since it relies on hashtags it is a method only applicable to social media platforms for which hashtags are a meaningful affordance; it is specifically relevant for social media platforms where hashtags are key to navigating content. This means that platforms like Twitter where tagging has a ‘conversational’ rather than an ‘organizational’ (Huang *et al.* 2010) purpose – where the tag is part of the message and the tag’s purpose is not always enhancing discoverability but conveying or emphasising information or connecting and participate in specific conversations – may be less meaningful to study through hashtag co-occurrence analysis. Or in any case, such platforms would require larger datasets as a smaller percentage of posts contains multiple hashtags. Second, even on platforms that use hashtags as a key affordance for retrieval not all posts will use (multiple) hashtags, which means that a subset of retrieved posts will not be considered in the analysis. Finally, hashtag co-occurrence analysis does not capture the degree to which identified communities are consciously experienced by social media users. In addition

to these limitations, it is important to note that hashtag co-occurrence analysis only captures the surface-level co-occurrence of hashtags and does not consider the actual content of the posts or the context in which they are used. This can lead to false positives, where hashtags that are used in different contexts but happen to co-occur are incorrectly grouped together. While acknowledging these limitations, hashtag co-occurrence analysis allows to gain valuable insights into the structure and dynamics of mnemonic communities inductively without having to fall back on a ‘container’ approach to collective memory, and while capturing the semantics of digital memory while ‘grounding them in users’ behaviors’ (Radicioni *et al.* 2021).

As detailed in the accompanying notebook on CoCalc, the proposed hashtag co-occurrence analysis workflow comprises several steps:

1. Data collection: Collect a dataset of social media posts containing hashtags. In the notebook, we will not cover this step but work with a pre-established dataset of Instagram posts tagged with #Juneteenth.
2. Create an edge list: Process the dataset to create a list containing all combinations of two co-occurring hashtags and their respective weights.
3. Generate a hashtag graph: Create an undirected graph object using the edge list.
4. Community detection: Apply a community detection algorithm, in this case, the Louvain method, to identify clusters of hashtags that tend to co-occur frequently with each other.¹ These clusters represent latent mnemonic communities.
5. Data exploration and labelling: Analyse the detected communities to understand the topics and themes they represent. Provide unique labels for the different communities in a semi-automated way by passing the most central hashtags through a language model and checking for duplicates.
6. Data organisation and analysis: Organise the data by assigning the community numbers and labels back to the posts, and by creating spreadsheets indicating the user membership of communities, and the communities themselves, accompanied by some basic statistics and a community cohesion metric to interpret relative cohesion or ‘explicitness’ versus diffusion or ‘latentness’ of communities. We achieve this by comparing the average strength of internal ties (within a community) to the average strength of external ties (between the community and other nodes) for each community. This approach is detailed more in-depth below.

#Juneteenth on Instagram

In the example we will use a dataset of 40,056 Instagram captions, sourced by querying the hashtag #Juneteenth, a key hashtag that references the annual celebration of the emancipation of enslaved African Americans in the United States, covering June 2021, when Juneteenth was first celebrated as a national holiday. The steps taken for the analysis are explained in more detail in the Jupyter Notebook hosted on CoCalc. The directory including the data and the Jupyter Notebook for the Python implementation can be accessed via CoCalc <https://www.cambridge.org/S2635023824000079/MEM-Notebooks>. Only a high-level description of the procedure will be discussed here.

Given the availability of a dataset containing the captions and some metadata such as user id, the first step would be to extract hashtags from the Instagram captions. Once we have extracted the hashtags, we must consider the network structure of our graph. Data

¹ The Louvain method is chosen for its efficiency in detecting community structures in networks by maximizing the modularity score. This means that it excels at identifying groups with many internal connections and fewer external connections, which makes it suitable for analysing the interrelation of hashtags.

collection methods affect graph structures and, in this case, because the data is sourced from one hashtag only, we know that all nodes have at least an edge connected to the main node: #Juneteenth. This will affect community detection in different ways, depending on the community detection algorithm. In most cases, it will construct a very large community around this main hub node, since it is the most central node in the community, which obfuscates potential communities in the process.

However, since we are interested in understanding what communities remember *within the context of #Juneteenth*, we can simply remove this node and from the network. Its presence in the network is not meaningful for our research purpose as it is not supposed to be part of a community; this also solves the centrality issue for community detection. In the example, we remove Juneteenth from the extracted hashtags for this purpose. In cases where data has been collected in different ways, for example, through place-based approaches that collect posts from a location tag, for example of a contested heritage site (Adriaansen 2021), this issue will be less urgent and this step can be skipped. The resulting network is now disconnected: since #Juneteenth connected to all other hashtags, removing it will result in a network of clusters of hashtags that are not all connected to each other anymore. This is not problematic if we choose a community detection algorithm that does not rely on the entire network being connected.

Once we have the lists of hashtags per post we create the edge list, which contains all unique combinations of hashtags in the dataset, including their weights. The main question here is how to define the weights, which should be determined with the research purpose in mind. We could simply take the frequency of hashtag pairs as the weights, which would give good insight into the popularity of certain hashtag combinations – something we could want the community detection algorithm to consider. However, such weights can also be misleading. Consider, for example, an individual, organisation or company using the popularity of #Juneteenth to promote some kind of service in many different posts. While doing so they combine #Juneteenth with several idiosyncratic hashtags representing the service. These hashtags would appear as a single, well-connected community in the network when we use frequency of occurrence as the edge weights. However, this misses our purpose as we depart from the vantage point of communities being established through user tagging practices to engage with imagined audiences, so a more meaningful weight would be the number of unique users that use a specific hashtag pair. This captures how many users use the hashtag combinations.

The resulting network is still rather ‘noisy’ in the sense that it contains many nodes that may be irrelevant for community detection because they are used infrequently in combination with other hashtags or only by a few users. Therefore, we will prune the network to filter out edges (and nodes that are left isolated) that are of less importance to the overall network structure. This process is also called network backboning (Coscia 2021; Coscia and Neffke 2017), and a variety of backboning approaches exist. A straightforward one would be to introduce a threshold and remove edges below a certain weight value. But this disregards the fact that for each hashtag the significance of its edges is distributed differently, so we need to consider a more sophisticated approach that considers the individual characteristics of each node. An established method that does this is the disparity filter (Serrano *et al.* 2009), which retains edges based on their weight relative to the total weight of the node’s edges. This method considers the fact that different nodes may have different degrees of connectivity and different overall weights, so it adjusts the threshold for edge removal dynamically based on the particular distribution of edge weights around each node. The resulting pruned backbone network still contains the core structure of the network but reduces the number of nodes and edges so the computational cost will be reduced as well, and the network will be easier to analyse. For smaller networks this step can be skipped.

Now we can run the community detection algorithm on the pruned network to identify latent mnemonic communities. We opt for the Louvain community detection method for its efficiency with large-scale networks (Blondel *et al.* 2008), but also because it considers the edge weights and can work with disconnected graphs. The resulting 87 communities provide insight into the different ways that users engage with the #Juneteenth hashtag and how they express different memories and commemorative practices around it by engaging with specific audiences. Each community is not determined by a shared memory narrative, but by a shared will to engage and target each other while relying on a shared vocabulary and a shared semiotic space. In our case this shared vocabulary is represented by the common use of hashtags, which function both as topic indicators and as linguistic markers of community engagement and identity. To proceed with the communities, we will pass the top 30 nodes (hashtags) with the highest degree (with most connections to other nodes) to a language model to generate labels and do a check for potential duplicates. Table 1 shows the 25 largest communities measured by the number of hashtags they contain. Community size (total number of hashtags in the community) is not the best indicator for community importance, so a column with the average edge weight is added, which indicates the average number of users that use hashtag pairs in the community.

Moreover, the average edge weight also provides an entry into establishing a measure for the ‘latentness’ of the mnemonic communities in the network. In user-based networks, homophily could be an indicator, as it refers to the tendency of individuals to associate and bond with similar others, which could be indicated by edge weights if they represent follower networks. However, in the context of this network defined by hashtag co-occurrence, and with the number of unique users as weights, homophily should be understood as the affinity between hashtags based on the similarity of user engagement patterns, reflecting their cohesion in terms of the semiosphere – rather than personal similarity among users. This form of homophily is crucial for identifying not just the explicit communities formed around specific, highly interacted hashtags but also the more diffuse, latent communities that represent broader, thematic semiospheres that users engage with, possibly without a conscious sense of belonging or explicit community identification.

By comparing the normalised internal average weight versus the normalised external average weight for each community, we can establish an idea of the ‘latentness’ or ‘explicitness’ of communities in the hashtag co-occurrence network. This comparison provides insight into the strength and focus of ties within communities (internal) versus those that extend beyond the community boundaries (external), while considering the network’s overall engagement levels. When the normalised internal average weight is higher than the external average, this suggests that a community is a more explicit, unified semiosphere, with strong internal cohesion around specific hashtags. This reflects a higher degree of homophily within the community, where the strength of connections among hashtags is higher relative to interactions that extend beyond the community. This indicates a more focussed user engagement with a bounded semiotic space. On the other hand, when the normalised external weight is higher than the internal one, the community can be seen as more latent, with its hashtags engaging more broadly across the network. This indicates that the community is more semiotically diffuse and connected to a wider range of other communities, which suggests that user engagement is less concentrated. This does not mean that these latent mnemonic communities are less important; they could function as ‘border’ zones in Lotman’s terms (Monticelli 2019). These are sites of dynamic semiotic processes where meaning is negotiated, transformed, or ‘translated’ through interactions with other semiospheres. Rather than having a clear, unified identity, these communities can serve a crucial role in facilitating semiotic exchange and

Table 1. Top 25 communities engaging with #Juneteenth on Instagram, sorted by size

	Community	Size	Average edge weight	Normalised internal average edge weight	Normalised external average edge weight	Difference
1	Unity	638	12	0.433	0.575	-0.142
2	Fashion	380	7	0.257	0.304	-0.046
3	Celebratejuneteenth	351	9	0.349	0.562	-0.212
4	Blackhistory	297	27	1	0.597	0.403
5	Hiphop	282	7	0.274	0.37	-0.096
6	Naturalhair	227	7	0.265	0.288	-0.024
7	Fathersday	136	7	0.242	0.278	-0.036
8	Bookstagrammer	99	7	0.246	0.228	0.018
9	Pridemonth2021	69	10	0.366	0.345	0.021
10	Charlotte	14	5	0.189	0.185	0.004
11	Nailart	10	5	0.19	0.148	0.043
12	Megantheestallion	8	5	0.172	0.148	0.025
13	Vaccine	7	5	0.189	0.148	0.041
14	Chadwick boseman	6	5	0.201	0.172	0.028
15	Charleston	6	4	0.16	0.185	-0.025
16	Oklahomacity	5	4	0.148	0.148	0
17	Lastcouplestanding	5	5	0.189	0.148	0.042
18	Meghan markle	4	4	0.157	0	0.157
19	Coffeelover	4	5	0.194	0.185	0.009
20	Zphib	4	6	0.213	0.148	0.065
21	Union	4	6	0.222	0	0.222
22	Findyourpark	4	8	0.305	0.166	0.139
23	Cannabis Community	4	6	0.209	0	0.209
24	Sicklecell	4	6	0.24	0	0.24
25	Erykahbadu	4	6	0.203	0	0.203

the generation of new meanings through the creative recombination and reinterpretation of elements from various semiospheres.

The value of our approach over other metrics that allow to quantify the cohesion or dispersion of communities within a network lies in its ability to account for weighted edges and provide an interpretable measure of the 'latentness' or 'explicitness' of communities. Unlike the proximity matrix discussed by Freelon (2020), which measures the proportion or number of ties spanning between pairs of communities, or the E-I index (Krackhardt and Stern 1988), which quantifies the ratio of external to internal ties for

a single community, our approach compares the average strength of internal ties to the average strength of external ties for each community. Normalising these average weights makes communities of different sizes and overall engagement levels comparable. Moreover, the difference between the normalised internal and external averages provides a clear interpretation: positive values indicate more ‘explicit’ or focussed communities with strong internal cohesion, while negative values suggest more ‘latent’ or diffuse communities with broader external connections across the network.

Additional columns in [Table 1](#) show the normalised internal and external average weights, and their difference, with positive values representing more explicit, and negative values more latent communities.

The detected communities reflect the diversity of memory practices and discourses that coalesce around the commemoration of Juneteenth on Instagram. Several of the larger communities appear to function as distinct semiospheres, each with its own set of central hashtags that frame the meaning and memory of Juneteenth within that community. For example, the prominent ‘blackhistory’ community centres around hashtags like #blackowned, #supportblackbusiness, and #blackentrepreneur and situates the memory of Juneteenth within a discourse of black economic empowerment and entrepreneurship. The frequent co-occurrence of these hashtags suggests a shared semiotic space where the historical legacy of slavery and emancipation is linked to contemporary issues of economic inequality and efforts to support black-owned businesses. The strongly positive difference between internal and external average edge weights suggests that this is an explicit mnemonic community, a focussed semiotic space and audience that is deliberately targeted by social media users through tagging practices. This community likely also shares a historical metanarrative of black social and economic empowerment, although this cannot be inferred from the hashtags alone.

Other pronouncedly explicit communities, particularly those with fewer nodes, exhibit characteristics reminiscent of mnemonic communities in the classical definition. They rely on specific networks of organisations or businesses that function more as ‘containers’ of memory than the ‘blackhistory’ community does. The ‘union’ community for example revolves around labour union accounts and related posts that celebrate Juneteenth as an act of working-class solidarity with the holiday’s historical significance for Black Americans. These accounts represent a formally organised network of social movements that demonstrate a shared sense of identity and belonging centred around advocacy for worker’ rights and racial justice. Other explicit communities operate in an equally focussed, but different, semiotic space as they are evidently leveraging potential audiences of #Juneteenth for marketing purposes. This is the case for the ‘cannabis’ community, which, upon manual inspection, comprises mainly of posts targeting cannabis products to a black audience. Similarly, the ‘vaccine’ community consists of hashtags used by a variety of organisations using the Juneteenth celebrations to advertise and offer free COVID-19 testing and vaccinations, as an orchestrated campaign to tackle vaccination hesitancy among Black Americans.

On the other hand, communities that are evidently latent have negative values in the ‘difference’ column of [Table 1](#), which indicates that the connections within the community are weaker on average than the connections bridging out to other parts of the network. The largest cluster, the ‘unity’ cluster, has the most the largest negative difference between its internal and external average edge weights in the top 25. This suggests that the hashtags central to this community, such as #blacklivesmatter, #blm, #freedom, #socialjustice, and #equality, are used widely across the dataset and not exclusively within a tight-knit community. The broad appeal and usage of these hashtags reflects the mainstreaming of the Black Lives Matter (BLM) movement and its associated language in the wake of the 2020 racial justice protests. Many users include these hashtags to express

solidarity and align themselves with the ideas of antiracism and social justice, without necessarily engaging with BLM as a movement. This underlines one of the reasons why these latent communities are so interesting: the aforementioned potential to function as semiotic ‘border’ zones, where new meanings can be negotiated and constructed in contact with other semiospheres. While meaning and narratives have a more fixed character in explicit communities, for example in the labour unions, a latent semiotic space like the ‘unity’ cluster allows to explore modes of intersectional thought and the negotiation or translation of meaning.

Other large latent mnemonic communities like ‘fashion’, ‘celebratejuneteenth’, and ‘hiphop’ show the incorporation of Juneteenth in different semiospheres. The ‘fashion’ community interestingly combines Juneteenth-related content with more general lifestyle and influencer hashtags like #ootd, #instadaily, #summervibes, which points to the commodification and aestheticisation of Juneteenth memory, as the holiday becomes incorporated into broader patterns of influencer marketing and personal branding practices on the platform. Users posting about Juneteenth in this community may not be consciously engaging in a shared attempt at remembrance but are nonetheless connected by the platform’s hashtag infrastructure.

There are three important things to consider when interpreting communities. First, they are dynamic and ever-changing. The identified communities are the result of a snapshot of #Juneteenth engagement spanning June 2021. This means that the temporal aspect should be considered when interpreting the data: the fluid nature of hashtags and their usage allows for continuous evolution and transformation of these communities. Factors such as the changing economy of hashtag virality, changes in platform algorithms that affect tagging behaviour, the aging of a platform’s user base, and socio-political developments that affect general memory discourses will affect community dynamics and persistence over time. Second, we should avoid overinterpreting the results of this algorithmic community detection approach. The communities detected by the Louvain method are largely dependent on the specifics of the dataset and the way the edge weights are defined. Hence, while this method can provide valuable insights into the mnemonic communities within the #Juneteenth network, it is just one way of understanding the complex and multifaceted nature of online mnemonic practices and would ideally be applied in a mixed-methods approach combined with qualitative analysis, in a form of ‘digital hermeneutics’ as suggested by Tuters and Willaert (2022). Third, given the fact that in the context of the semiosphere hashtags have a degree of semiotic agency – hence we took them as nodes – the latent communities they establish should not be confused with some form of digital social structure, rather as semiospheres. The communities do, for example, not fully overlap with community membership, as users can actively engage in different communities. When we link the communities back to the users, we see that the average user is involved in approximately 2 communities, with a standard deviation of 1.21. The most active user engages with 13 communities in total.

Conclusion

Identifying and understanding mnemonic communities on social media platforms is a challenge to scholars working with large datasets. Moreover, the fluid and often elusive nature of social interactions online make that one cannot assume online mnemonic communities to be a mere digital equivalent of offline communities. This poses both the need to reconceptualise the notion of community in the context of new media ecologies and to rethink methodological approaches to identifying mnemonic communities using social media data.

This article illustrates that the seemingly divergent conceptions of online (mnemonic) communities – one grounded in social and cultural dimensions, the other in network analysis – are not incommensurable. Instead, they can be integrated through the lens of semiospheres. The semiosphere, a concept adapted from Yuri Lotman's semiotics, provides a framework for understanding communities as dynamic, fluid entities shaped by both human interactions and the semiotic structures of digital platforms. In this context, hashtags function as semiotic markers that link the social and semantic aspects of community formation.

This integrated approach enables a more nuanced understanding of digital mnemonic communities. It acknowledges that these communities are not just clusters of individuals with shared interests but are also shaped by the semiotic affordances of social media platforms, such as hashtags. These semiotic elements carry and convey meanings, sometimes beyond the direct intention of users, and are influenced by the underlying algorithms of the platforms. This perspective allows us to capture the complex interplay between individual agency and the structured, yet fluid, nature of digital communication spaces.

The proposed methodology of hashtag co-occurrence analysis for mnemonic community detection exemplifies this integrated approach. By understanding the communicative context of social media platforms in terms of semiospheres, hashtags can be understood as semiotic markers that are human-generated and human-chosen, but also have a degree of semiotic agency themselves as they function within a platform environment governed by algorithmic processes. Hashtags also have the advantage that they function both as index to the content and are concepts that conceptually capture memory topics, contents, or attitudes to the past. By not taking social media users, but hashtags as the nodes in the network we can capture these dynamics. By creating a network graph based on the co-occurrence of hashtags in social media posts, we can identify latent mnemonic communities that are not necessarily visible or conscious to the users, but that reveal the complex semiotic engagements of users with memories and historical representations as visible in social media metadata. Hashtag co-occurrence analysis allows us to study digital memory formation in terms of networks rather than containers, and to capture the fluidity and diversity of mnemonic communities on social media platforms.

The example of #Juneteenth on Instagram demonstrated the potential of this approach to reveal the different ways that users engage with nostalgic content and how they construct shared memories from various contexts and in relation to various topics and themes within the context of #Juneteenth. The 87 identified communities showcase the diversity and complexity of memory-sharing practices on Instagram. By taking the difference between the normalised internal and external average edge weights as a measure for the 'latentness' of mnemonic communities, we were able to identify both concrete, explicit, communities engaging with #Juneteenth for specific political or commercial purposes, and more latent communities that operate as shared semiospheres and discursive spaces on Instagram. The findings from the #Juneteenth dataset underscore the importance of considering both the explicit and latent dimensions of community formation, which is crucial for understanding the dynamics of digital mnemonic practices.

The proposed approach is not without limitations and challenges. It relies on hashtags as the main source of information, which means that it is only applicable to social media platforms where hashtags are a main affordance. It also does not consider the actual content or context of the posts or the hashtags, which can lead to misinterpretations. Moreover, it does not account for the temporal dynamics of hashtag usage and community formation, which can change over time due to various factors. Furthermore, it does not capture the subjective experiences or perceptions of the users who participate in these communities, and it is therefore advised to use hashtag co-occurrence analysis in a mixed-methods approach combined with qualitative content analysis. Nonetheless,

hashtag co-occurrence analysis can be a valuable tool for identifying and interpreting digital mnemonic communities inductively and comprehensively, without having to rely on prior assumptions or predefined categories about social communities as ‘containers’ of memory. This enables us to explore the complex and diverse landscape of digital memory formation on social media platforms and to gain insights into the semiotic contexts and processes that shape it.

Supplementary material. Computational Notebook files are available as supplementary material at <https://doi.org/10.1017/mem.2024.7> and online at <https://www.cambridge.org/S2635023824000079/MEM-Notebooks>.

Data availability statement. The data that support the findings of this study are openly available in CoCalc at <https://www.cambridge.org/S2635023824000079/MEM-Notebooks>.

Funding statement. This work received no specific grant from any funding agency, commercial or not-for-profit sectors.

Competing interests. No competing interests are to be declared.

References

- Abidin C** (2021) From “networked publics” to “refracted publics”: a companion framework for researching “below the radar” studies. *Social Media + Society* 7(1), 1–13. doi: 10.1177/2056305120984458
- Adriaansen R-J** (2021) Social media and multimodal historical representation: depicting Auschwitz on Instagram. In Berger S, Brauch N and Lorenz C (eds), *Analysing Historical Narratives. On Academic, Popular and Educational Framings of the Past*. New York: Berghahn Books, 192–208.
- Adriaansen R-J** (2022a) Dark academia: curating affective history in a COVID-era internet aesthetic. *International Public History* 5(2), 105–114. doi: 10.1515/iph-2022-2047
- Adriaansen R-J** (2022b) Historical analogies and historical consciousness: user-generated history lessons on TikTok. In Carretero M, Cantabrana M and Parellada C (eds), *History Education in the Digital Age*. Cham: Springer International Publishing, 43–62.
- Anderson B** (2006) *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Revised edition. London: Verso.
- Assmann J** (2007) *Das kulturelle Gedächtnis*. München: C.H.Beck.
- Assmann J** (2008) Communicative and cultural memory. In Erll A and Nünning A (eds), *Communicative and Cultural Memory. An International and Interdisciplinary Handbook*. Berlin: De Gruyter, 109–118.
- Barassi V** (2016) Ethnography beyond and within digital structures and the study of social media activism. In Hjorth L, Horst H, Galloway A and Bell G (eds), *The Routledge Companion to Digital Ethnography*. New York: Routledge, 406–415.
- Blondel VD, Guillaume J-L, Lambiotte R and Lefebvre E** (2008) Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008(10), 10008. doi: 10.1088/1742-5468/2008/10/P10008
- boyd d** (2010) Social network sites as networked publics: affordances, dynamics, and implications. In Papacharissi Z (eds), *Networked Self: Identity, Community, and Culture on Social Network Sites*. New York: Routledge, 39–58.
- Chouchani N and Abed M** (2020) Online social network analysis: detection of communities of interest. *Journal of Intelligent Information Systems* 54(1), 5–21. doi: 10.1007/s10844-018-0522-7
- Coscia M** (2021) The Atlas for the Aspiring Network Scientist, arXiv. doi: 10.48550/arXiv.2101.00863
- Coscia M and Neffke FMH** (2017) Network backboning with noisy data. In *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, pp. 425–436.
- Erll A** (2011) Travelling memory. *Parallax* 17(4), 4–18.
- Erll A** (2017) *Kollektives Gedächtnis und Erinnerungskulturen: eine Einführung, 3. aktualisierte und erweiterte Auflage*. Stuttgart: J.B. Metzler Verlag.
- Esposito E** (2017) Algorithmic memory and the right to be forgotten on the web. *Big Data & Society* 4(1), 1–11. doi: 10.1177/2053951717703996
- Fani H and Bagheri E** (2017) Community detection in social networks. *Encyclopedia with Semantic Computing and Robotic Intelligence* 01, 01. doi: 10.1142/S2425038416300019
- Fani H, Jiang E, Bagheri E, Al-Obeidat F, Du W and Kargar M** (2020) User community detection via embedding of social network structure and temporal content. *Information Processing & Management* 57(2), 102056. doi: 10.1016/j.ipm.2019.102056

- Frassinelli PP** (2018) Hashtags: #RhodesMustFall, #FeesMustFall and the temporalities of a meme event. In Mutsvairo B and Karam B (eds), *Perspectives on Political Communication in Africa*. Cham: Springer International Publishing, 61–76.
- Freelon D** (2020) Partition-specific network analysis of digital trace data: research questions and tools. In Foucault Welles B and González-Bailón S (eds), *The Oxford Handbook of Networked Communication*. Oxford: Oxford University Press, 90–110.
- Gallie WB** (1968) Essentially contested concepts. In Gallie WB (ed.), *Philosophy and the Historical Understanding*. New York: Schocken Books, 157–191.
- Ginzarly M and Srouf FJ** (2022) Cultural heritage through the lens of COVID-19. *Poetics* **92**, 101622. doi: 10.1016/j.poetic.2021.101622
- Gunn HK** (2021) Filter bubbles, echo chambers, online communities. In Hannon M and de Ridder J (eds), *The Routledge Handbook of Political Epistemology*. Abingdon, OX: Routledge, 192–202.
- Halbwachs M** (1980) *The Collective Memory*. New York: Harper & Row.
- Halbwachs M** (1992) *On Collective Memory*. Chicago: University of Chicago Press.
- Himmelboim I** (2017) Social network analysis (social media). In Matthes J (ed.), *The International Encyclopedia of Communication Research Methods*. Hoboken, NJ: John Wiley & Sons, Ltd, 1–15.
- Hoskins A** (2009) The mediatization of memory. In Garde-Hansen J, Hoskins A and Reading A (eds), *Save As ... Digital Memories*. London: Palgrave Macmillan UK, 27–43.
- Hoskins A** (2011) 7/7 and connective memory: interactional trajectories of remembering in post-scarcity culture. *Memory Studies* **4**(3), 269–280.
- Hoskins A** (2017) Memory of the multitude : the end of collective memory. In Hoskins A (eds), *Digital Memory Studies: Media Pasts in Transition*. New York & London: Routledge, 85–109.
- Huang J, Thornton KM and Efthimiadis EN** (2010) Conversational tagging in Twitter. In *Proceedings of the 21st ACM conference on hypertext and hypermedia*, New York, NY, USA: Association for Computing Machinery, pp. 173–178.
- Jacobsen B and Beer D** (2021a) *Social Media and the Automatic Production of Memory: Classification, Ranking and the Sorting of the Past*. Bristol: Bristol University Press.
- Jacobsen BN and Beer D** (2021b) Quantified nostalgia: social media, metrics, and memory. *Social Media + Society* **7**(2), 1–9. doi: 10.1177/20563051211008822
- Kansteiner W** (2002) Finding meaning in memory: a methodological critique of collective memory studies. *History and Theory* **41**(2), 179–197. doi: 10.1111/0018-2656.00198
- Kozinets RV** (2015) *Netnography: redefined*, 2nd Edn. Los Angeles: Sage.
- Krackhardt D and Stern RN** (1988) Informal networks and organizational crises: an experimental simulation. *Social Psychology Quarterly* **51**(2), 123–140. doi: 10.2307/2786835
- Landsberg A** (2004) *Prosthetic Memory: The Transformation of American Remembrance in the Age of Mass Culture*. New York: Columbia University Press. p. 240.
- La Rocca G and Boccia Artieri G** (2023) Interpreting the changeable meaning of hashtags: toward the theorization of a model. *Frontiers in Sociology* **7**, 1–14. Retrieved from: <https://www.frontiersin.org/articles/10.3389/fsoc.2022.1104686>
- Laužikas R and Dallas C** (2023) The message is the agent: nexus and semiosphere in social media communication. *Zenodo (preprint)*, 1–29. doi: 10.5281/zenodo.8129305
- Laužikas R, Dallas C, Thomas S and Vaitkevičius V** (2018) Archaeological knowledge production and global communities: boundaries and structure of the field. *Open Archaeology* **4**(1), 350–364. doi: 10.1515/opar-2018-0022
- Litt E and Hargittai E** (2016) The imagined audience on social network sites. *Social Media + Society* **2**(1), 1–12. doi: 10.1177/2056305116633482
- Lotman Y** (2005) On the semiosphere. *Sign Systems Studies* **33**(1), 215–239. doi: 10.12697/SSS.2005.33.1.09
- Marwick AE and boyd d** (2011) I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society* **13**(1), 114–133. doi: 10.1177/1461444810365313
- Menke M** (2019) *Mediennostalgie in digitalen Öffentlichkeiten: zum kollektiven Umgang mit Medien- und Gesellschaftswandel*. Köln: Herbert von Halem Verlag.
- Menke M and Birkner T** (2022) Digitales Erinnern: Eine Synthese anhand der analytischen Kategorien Speichern, Datafizierung, Vernetzung und Affektivität. In Pentzold C and Lohmeier C (eds), *Handbuch kommunikationswissenschaftliche Erinnerungsforschung*. Berlin, Boston: De Gruyter, 495–516.
- Monticelli D** (2019) Borders and translation: revisiting Juri Lotman's semiosphere. *Semiotica* **2019**(230), 389–406. doi: 10.1515/sem-2017-0125
- Nora P** (Ed.) (1996) *Realms of Memory: Rethinking the French Past*. New York: Columbia University Press.
- Olick JK** (2014) Reflections on the underdeveloped relations between journalism and memory studies. In Zelizer B and Tenenboim-Weinblatt K (eds), *Journalism and Memory*. London: Palgrave Macmillan UK, 17–31.

- Papacharissi Z and de Fatima Oliveira M** (2012) Affective news and networked publics: the rhythms of news storytelling on #Egypt. *Journal of Communication* **62**(2), 266–282. doi: 10.1111/j.1460-2466.2012.01630.x
- Parks MR** (2010) Social network sites as virtual communities. In Papacharissi Z (eds), *A Networked Self*. New York: Routledge, 105–123.
- Pentzold C, Lohmeier C and Birkner T** (2022) Kommunikatives Erinnern. In Pentzold C and Lohmeier C (eds), *Handbuch kommunikationswissenschaftliche Erinnerungsforschung*. Berlin, Boston: De Gruyter, 47–70.
- Radicioni T, Saracco F, Pavan E and Squartini T** (2021) Analysing Twitter semantic networks: the case of 2018 Italian elections. *Scientific Reports* **11**(1), 13207. doi: 10.1038/s41598-021-92337-2
- Rambukkana N** (2015) #Introduction: hashtags as technosocial events. In Rambukkana N (eds), *Hashtag Publics: The Power and Politics of Discursive Networks*. New York: Peter Lang, 1–12.
- Rheingold H** (1994) *The Virtual Community: Homesteading on the Electronic Frontier*. New York, NY: HarperPerennial.
- Rocheleau S and Milette M** (2015) Meta-hashtag and tag co-occurrence: from organization to politics in the French-Canadian Twittersphere. In Rambukkana N (eds), *Hashtag Publics: The Power and Politics of Discursive Networks*. Toronto: Peter Lang, 243–254.
- Rothberg M** (2009) *Multidirectional Memory: Remembering the Holocaust in the Age of Decolonization*. Palo Alto: Stanford University Press.
- Schwarzenegger C and Lohmeier C** (2020) Reimagining memory : digital media and a new polyphony of memory. In Filimowicz M and Tzankova V (eds), *Reimagining Communication: Experience*. New York: Routledge, 132–146.
- Serrano MÁ, Boguñá M and Vespignani A** (2009) Extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences* **106**(16), 6483–6488. doi: 10.1073/pnas.0808904106
- Silverman M** (2013) *Palimpsestic Memory: The Holocaust and Colonialism in French and Francophone Fiction and Film*, 1st Edn. New York: Berghahn Books.
- Singer P, Niebler T, Hotho A and Strohmaier M** (2014) Folksonomies. In Alhaji R and Rokne J (eds), *Encyclopedia of Social Network Analysis and Mining*. New York, NY: Springer, 542–547.
- Smit R** (2020) Connective memory work on justice for Mike Brown. In Merrill S, Keightley E and Daphi P (eds), *Social Movements, Cultural Memory and Digital Media: Mobilising Mediated Remembrance*. Cham: Springer International Publishing, 85–108.
- Smit R** (2022) Die Plattformisierung des Erinnerns. In Pentzold C and Lohmeier C (eds), *Handbuch kommunikationswissenschaftliche Erinnerungsforschung*. Berlin, Boston: De Gruyter, 471–494.
- Terren L and Borge-Bravo R** (2021) Echo chambers on social media: a systematic review of the literature. *Review of Communication Research* **9**, 99–118.
- Trant J** (2009) Studying social tagging and folksonomy: a review and framework. *Journal of Digital Information* **10**(1), 1–44. Retrieved from: <https://jodi-ojs-tdl.tdl.org/jodi/article/view/269>
- Tuters M and Willaert T** (2022) Deep state phobia: narrative convergence in coronavirus conspiracism on Instagram. *Convergence* **28**(4), 1214–1238. doi: 10.1177/13548565221118751
- Vander Wal T** (2007) Folksonomy. Retrieved from: <https://vanderwal.net/folksonomy.html>
- Zappavigna M** (2015) Searchable talk: the linguistic functions of hashtags. *Social Semiotics* **25**(3), 274–291. doi: 10.1080/10350330.2014.996948

Robbert-Jan Adriaansen (Erasmus University Rotterdam) is a tenured assistant professor at the History Department of Erasmus University Rotterdam and is an executive director of the Center for Historical Culture. His research interests include the dynamics of historical consciousness and historical representation on social media, in social movements and in education.

Cite this article: Adriaansen R-J (2024). Latent and explicit mnemonic communities on social media: studying digital memory formation through hashtag co-occurrence analysis. *Memory, Mind & Media* **3**, e13, 1–21. <https://doi.org/10.1017/mem.2024.7>