**Artificial Intelligence, digital capital, and epistemic domination on Twitter:**

**A study of families affected by imprisonment**

**ABSTRACT**

Online Social Networking Sites (SNSs) and other Artificial Intelligence (AI) systems are transforming the epistemological foundations of justice systems and influencing knowledge production concerning criminal justice and its impact. This article focuses on a dimension of criminal justice which is the impact of imprisonment on families and seeks to unravel how knowledge about this problem is produced on SNSs. To this end, it draws on a study that explored conversational networks of key stakeholders on the SNS, Twitter. Building on insights from the study, the paper unravels interdependent sociotechnical dynamics that reproduce the offline marginality of affected families and operate as barriers to equitable knowledge production. Through its analysis of the dynamics, the paper provides new insights and advances the sparse criminological scholarship on the intersections of AI systems and the delivery of justice. It specifically highlights exclusionary epistemic processes that are fomented by the infrastructure of AI systems and the social contexts in which they are deployed.

**INTRODUCTION**

Social Networking Sites (SNSs) such as Twitter and Facebook are Artificial Intelligence (AI) systems that use Machine Learning algorithms to curate and structure online discourse. Studies demonstrate that the sites are important spaces of networked discursive interaction amidst a sustained rise in global usage (London School of Economics and Political Science (LSE) 2018. For example, more than 330 million people currently use Twitter globally for online communication and discursive networking (Ofcom 2020). SNSs therefore form part of the data-driven AI technologies increasingly capable of shaping discourse and knowledge production concerning social issues including criminal justice. This paper focuses on a dimension of criminal justice which is the impact of imprisonment on families. The paper’s primary aim is to unravel how knowledge about this problem is produced on SNSs. To this end, the paper draws on a study of the online conversational networks of key stakeholders either affected by, or involved in addressing, the impact of imprisonment on families: the families themselves; Third Sector Organisations (TSOs) providing family support services; and policy makers.

TSOs are, ‘registered charities and other organisations such as associations, self-help groups, community groups, social enterprises, mutuals and co-operatives’ (National Audit Office - NAO 2020). Regarding the policy makers, we identified the following in the dataset: Senior Ministers responsible for prison policy; relevant state departments; Members of Parliament; and prison services. The services can be categorised as policy makers operating at mesostructural levels of policy making (primarily at the institutional or organisational level).

The study examined these stakeholders’ conversational networks on the SNS, Twitter, and was motivated by several factors. One is the importance of investigating the mode of knowledge production to explore whether it is disempowering and exclusionary in the sense that it reproduces conditions of marginality and invisibility already affecting the families in the offline social world. These offline conditions of disadvantage are well-documented in the criminological literature (see, Condry and Smith 2018 for a comprehensive analysis). Theproliferation of SNSs and their increasing role in knowledge production also renders the study essential. Indeed, other studies show that powerful actors such policy makers and others seeking to influence criminal justice policy, increasingly rely on the knowledge produced and disseminated across data-driven SNSs, about crime and criminal justice (e.g., Walsh and O'Connor 2019). Therefore, dynamics of knowledge production on such sites should be explored and barriers to equitable knowledge production identified alongside remedial strategies.

The dearth of research in this area provides further impetus for the study. Toour knowledge, this study is the first to explore these issues. Indeed, the paucity of research is surprising considering that many families are affected, and it is a problem that is of great relevance to criminology and criminal justice research, particularly penological studies that explore the broader social consequences of state inflicted punishment.

To generate relevant insights, the study adopted an innovative triangulation of computational and social science methods in its analysis of the stakeholders’ conversational networks on Twitter.It found that offline social conditions interact with technical dimensions of SNSs to formpower-ladensociotechnical dynamics of knowledge production. The offline social conditions are characterised by TSOs’ and policy makers’ access to digital capital, positionality (higher social status), and ascribed credibility. These interact with amplificatory platform algorithms, which constitute the technical aspect of the sociotechnical dynamics. Together, the dynamics elevate the discourses of TSOs and policy makers above those produced by affected families, replicating their offline marginality. This can potentially provoke ‘digitised epistemic domination’ (see, Ugwudike 2020:483) which refers to the power and ability of influential actors to dominate discourses produced via data driven digital technologies about key social issues, thereby marginalising other discourses. It is a form of domination that occurs at the intersection of positionality and knowledge production, and it manifests as dominance over information construction and diffusion across a tableau of data-driven digital artefacts including SNSs.

**Offline marginality and inequitable knowledge production on SNSs: The case of families affected by imprisonment**

A fundamental theme arising from studies of imprisonment in Western jurisdictions is that the imprisonment of a family member can create socioeconomic problems that foment or exacerbate the marginality of many affected families (See, Condry and Smith 2018). For example, access to vital resources such as suitable education and employment can be disrupted, in part due to stigmatisation. Indeed, studies generally show that families experience marginality via a wide range of socioeconomic, practical, and psychological problems (see generally, Hutton and Moran 2019). The studies point to socioeconomic problems, for example, poverty and unemployment, practical problems such as childcare costs and the travel costs of prison visits, as well as psychological trauma and distress.

In the offline social world, the relative marginality of affected families places them in a less powerful position than other stakeholders such as policy makers. In their analysis of affected families’ experiences, Lanskey and colleagues (2018) acknowledge their relative powerlessness and demonstrate how penal power, i.e., the power of the state to punish through its laws and policies, infiltrates the lives of affected families due to the imprisonment of a family member, exacerbating their marginality. Alongside policy makers, other powerful actors capable of exerting penal power include TSOs, particularly the large national organisations, some of whom operate in collaboration with the state. Indeed, within the justice system, TSOs are primary providers of support to affected families and are as such able to institute policies that have ramifications for people in prison and their families. The role of TSOs in the justice system and across the public sector, has been likened to the emergence of a powerful ‘shadow state’ comprising non-state actors involved in delivering criminal justice and other public services traditionally reserved for the state (Maguire et al. 2019; Wolch 1990:22). The ways in which offline power differences between the three stakeholders can re-emerge online and interact with technical dimensions of SNSs to influence knowledge production has been ignored and is explored in this paper.

**Models of knowledge production on SNSs**

Several contrasting models of knowledge production on SNSs currently exist. One is the empowerment model endorsed by the so-called ‘cyber-optimists’ (Keller 2020: 175) who contend that web 2.0 is enabling high levels of interactivity, connectivity and content creation. These, it is argued, can empower marginalised groups such as families affected by imprisonment, to reverse their social invisibility and enhance their social participation, creating an unprecedented shift in social and political power (e.g., Bonilla and Rosa 2015). In the sparse criminal justice theorisation of this problem, it has been similarly noted that SNS affordances can enable less powerful groups to define key criminal justice issues in real time. The affordances it is argued, are redistributing knowledge production powers, challenging the long-established position of more powerful groups such as the police, as the authoritative definers of such issues (e.g., Greer and McLaughlin 2010). Policy makers and TSOs fall within this category. This discourse about the empowering potential of SNS affordances presumes that information produced by data-driven technologies such as SNSs is based on objective data curation and presentation, which guarantees fair representation of all. It is a discourse that evokes utopic visions of the social benefits of data-driven technologies.

However, the discourse coexists with a disempowerment model that highlights the exploitative and exclusionary potential of data-driven, web-based technologies such as SNSs (e.g., van Dijck, 2013). Fuchs’ (2013) Marx-inspired analysis of SNSs for instance, draws attention to powerful commercial and state activities that involve using data generated from the free labour of users, for political and financial gain. Such practices include presenting users with information tailored to influence their purchasing habits or the exercise of their democratic franchise. These examples of disempowering practices refute popular depictions of SNSs, and Web 2.0 generally, as participatory and democratising platforms that amplify the contributions of relatively marginal groups such as families affected by imprisonment.

Studies of the disempowerment model also reveal dynamics that structure knowledge production by using patterns in user data to amplify the discourses of the more prominent groups in society, imbuing them with epistemic power and hegemonic status (e.g., Berg 2014). The voices of already marginal groups can be stripped of any potency and silenced. In this scenario, the offline status of the more powerful as authoritative definers of social issues, capable of exerting epistemic domination, is reproduced online. Intersecting social and technical dimensions constitute sociotechnical dynamics that can produce this outcome and are discussed below.

**THEORETICAL FRAMEWORK: SOCIOTECHNICAL DYNAMICS OF KNOWLEDGE PRODUCTION ON SNSs**

**Technical dimensions: Distortive content manipulation algorithms**

Distortive content recommendation and distribution algorithms instituted by platform companies form the technical dimension of knowledge production on SNSs. The algorithms infer prominence from patterns in user data and amplify the visibility of certain posts over others (see, Petre et al. 2019). To understand this process, it is important to consider how SNSs structure information flows and influence the visibility and prominence of user content. Twitter, for example, uses ‘ranking algorithms’ to mine user data and determine degrees of visibility on the basis of presumed prominence. One way in which the algorithms perform this function is to score each tweet using a predictive ‘relevance model.’ This automated pattern recognition model considers *inter alia*, the author and the nature of the Tweet: how recent it is, whether it contains media, and the number of interactions it generates such as ‘retweets’, ‘mentions’ (names in messages) and ‘likes’ (Koumchatzky and Andryeyev 2017). Relevancy scores are allocated based on these algorithmic assumptions, and the posts that attract the most reactions are amplified, sometimes becoming trending topics regardless of authenticity. Therefore, these technical dimensions of knowledge production amplify certain discourses on the basis of presumed prominence. Social dimensions that can migrate to online spaces to reproduce the prominence of relatively powerful actors in a conversational network and spawn such algorithmic amplification are discussed later. Meanwhile, data-driven, algorithmic curation of prominent posts or trending topics to control information and knowledge production, renders social media platforms ‘algorithmic gatekeepers’ and producers of information that can evolve into accepted knowledge (Graves 2020: 342). In a Foucauldian analysis of how social media algorithms structure invisibility, Bucher (2012) describes the practice as a manifestation of ‘algorithmic power.’ This type of power positions platform companies as ‘curators of public discourse’ (Gillespie 2010: 347) as well as holders of ‘algorithmic metapower’ (Berg 2014: 22) who engage in ‘platform paternalism’ by accusing others using content optimisation techniques of ‘gaming the system’ (Petre et al. 2019: 1). Thus, critics refute the platforms’ ostensible position as neutral conduits of other people’s speech and choices, and argue that they are instead, data-driven social structures that play a fundamental role in the politics of undemocratic knowledge production (see also, Berg 2014; Petre et al. 2019).

**Social dimensions: The nexus of access to digital capital, positionality and ascribed credibility**

A key social dimension that can interact with amplificatory algorithms to influence how knowledge is produced on SNSs is digital capital. Some scholars note that access to digital capital, particularly the power to use digital technologies to influence or even dominate knowledge production, can be linked to intersecting social categories of race, gender and socioeconomic disadvantage (e.g., Benjamin 2019; Ugwudike 2020; van Dijk 2005). That is, access to digital capital can be connected to positionality (relative social status). In SNSs, a high social status and access to digital capital can empower certain actors by rendering them more amenable to algorithmic amplification and epistemic domination. This is likely to be the case where the high-status actors occupy positions of power and authority in the social world (for example, TSOs and policy makers) and as a result attract substantial attention from others on SNSs. As Becker’s (1967:141) ‘hierarchy of credibility’ suggests, the amplified discourses of high-status actors are also likely to be accorded substantial credibility, placing them in a vantage position to influence the knowledge realisable from SNSs. The concept of ‘hierarchy of credibility’ emphasises that such actors are often ascribed greater credibility and prominence than socially marginal groups whom Becker describes as underdogs.

The discussion so far suggests that certain sociotechnical dynamics of knowledge production on SNSs can amplify some voices within a conversational network, undermining the visibility and contributions of others in ways that reproduce historical inequalities. In the next section, we present the aforementioned study which explored how knowledge about the impact of imprisonment on families is produced on the SNS, Twitter.

**THE STUDY**

**Methodology**

To retrieve publicly available content from Twitter, we used Web Data Research Assistant (Version 3.5.11) (Carr 2020) (henceforth WebDataRA) which is an innovative social media data mining software. WebDataRA uses Twitter’s application programming interface (API), to extract historical and real time data directly from Twitter. Initially, the study adopted a purposive sample technique. The inclusion criterion was as follows:

* Publicly available tweets posted during a data capture period of 01/01/2020-01/07/2020.

The search terms applied were ‘prisoners’ and each of the following words: families, wife, partner, children, brothers, sisters, parents, mum, dad. As well as each of those words followed by the phrase ‘of prisoner/s’. These search terms had to be deployed because the area of study has never been a ‘trending’ topic with a prominent ‘hashtag’ and extensive searches had to be conducted to locate relevant content.

**The sample**

Our initial assessment of the retrieved data revealed a key pattern: the families affected by imprisonment were the most marginal of the three stakeholders. They were the least visible in the results produced by Twitter search algorithms. We surmised that this could be an early indication of the algorithmic deamplification problem, whereby certain content that does not attract the attention of other users and as such does not appear influential, attain less visibility than others. We analysed the problem further through social network analysis and the results are presented later in this paper.

*Remedial Strategy: What can be done to identify the contribution of low capital regions of a network?*

WebDataRA provides digital research tools for identifying users who may be located at the margins of a SNS. In particular, it generates the following metadata of Twitter metrics: number of replies, mentions, retweets, likes, and followers. We located family members by scrutinising these metrics, focusing on accounts with key emblems of influence (e.g., a large number of followers) and locating families ‘following’ the biggest accounts. The accounts were held by all stakeholders: policy makers, TSOs and families.

The searches yielded 27,938 tweets and the dataset was converted from HTML format to an aggregated and extended Excel spreadsheet for cleaning. It became clear that the main groups discussing the topic of families affected by imprisonment were three key stakeholders: the families themselves; TSOs providing family support services; and policy makers. Having identified the stakeholders, we generated more focused data by retrieving their ‘tweets and replies’ (posted during the data capture period) from their profile page. This yielded a profusion of tweets and we limited the dataset to the first 10,000 tweets retrieved per stakeholder group, giving an overall total of 30,000 retrieved tweets.

**Using social network analysis to Identify influential actors in the conversational network**

To identify connections between the stakeholders and determine the accounts influencing the conversational network, we uploaded the 30,000 tweets to Netlytic2 (Gruzd 2016) for a Name Network Analysis. This type of Social Network Analysis (SNA) mines all the names (mentions) in a corpus (collection) of Tweets to generate conversational network graphs that depict connections between the authors/users, based on who mentioned whom. It is therefore useful for studies such as this since it reveals the users who are influential in the network because they are attracting the most attention (by receiving the most ‘mentions’ in tweets, retweets and replies). A name network analysis also reveals the users whose efforts to interact with the influential accounts are being unreciprocated. This can disconnect them from core conversational interactions and render their views less amenable to amplification by platform algorithms. Furthermore, a name network analysis is useful for identifying the most well-connected users who attract substantial attention *and* form reciprocal connections. Because of the substantial attention they elicit, such users have a high probability of benefiting from amplificatory platform algorithms. Using a name network analysis, researchers can investigate active and multi-directional forms of discursive participation as well as degrees of ‘information sharing, community building and collaboration’ between the members of a conversational network (Gruzd et al. 2016).

Descriptive statistics of the dataset that was used for the name network analysis are provided in Table 1. The Table shows that there were 59 duplicates in the dataset of 30,000 tweets. Therefore, the name network analysis is based on 29,941 tweets.

**Table 1: The dataset**

|  |  |
| --- | --- |
| **Statistics**  | **Value**  |
| Retrieved tweets | 30,000 |
| Duplicates | 59 |
| Corpus (retrieved tweets minus duplicates) | 29,941 |
| Nodes (users)/names found | 7,509 |
| Nodes (users)/names with ties (communication links or connections with others in the network) \* | 4,152 |
| Edges (lines indicating links or connections between nodes/users e.g., mentions) | 53,751 |

\*Unlike nodes/users without ties in the network, those with ties are connected to others in the network, for example, through mentions.

***Assessing domination through an analysis of centrality dynamics***

To conduct the name network analysis, the corpus of tweets was uploaded to Netlytic and visualisations of the network were generated. Centrality measures were used to interpret the visualisations. In directed SNSs such as Twitter where users direct ties/connections towards each other, for example through mentions in tweets, centrality analysis is useful for identifying the influential or central actor/s in a network of users tweeting about a topic (Hanneman & Riddle, 2005). Our findings regarding centrality are discussed below.

**RESULTS**

*Indegree centrality*

Netlytic’s indegree centrality measure identifies prominent users whose centrality in a network stems from the fact that they receive the most ties/connections or efforts to communicate with them, through ‘mentions’. In SNSs, ‘mentions’ in tweets, replies and retweets, indicate that one user acknowledges or wishes to communicate with another user. Therefore, users who receive the most mentions can be considered popular or prominent within a network (Hanneman & Riddle, 2005). The SNA revealed that policy makers are overrepresented in this category. Figure 1 below presents a visualisation of the network, showing indegree centrality3. Policy makers have the largest node sizes4 coded mainly in green to the left of the visualization and are as such leading indegree centrality followed by TSOs who are also represented mostly by large green nodes located towards the left of the visualisation. There are some large orange nodes to the right of the network, but these are also led by policy makers and TSOs.

[insert Figure 1.]

Policy makers leading indegree centrality as indicated by the profusion or density of large nodes in that space, consist of: four senior ministers; one shadow minister; two state departments; and a male prison service, while the TSOs comprise seven large national charities. Both policy makers and TSOs also dominate the biggest cluster5 towards the centre of the network where most inbound conversations/connections are occurring. The families of people in prison are not similarly represented, indicating a degree of marginality affecting this group.

*Outdegree centrality*

Our analysis of network behaviour using Netlytic’s outdegree value which measures the outbound connections/messages that users send, revealed further instances of marginality. In the visualisation of outdegree centrality (see Figure 2), the same policy makers with indegree centrality also have the largest node sizes, coded mainly in green to the left of the visualization. They are as such, also leading outdegree centrality, followed by TSOs who are again also represented mostly by large green and turquoise nodes located towards the left of the visualisation. Additional policy makers (six prison services) and TSOs (two TSOs). also emerge as outdegree leaders, represented by larger orange nodes and larger blue nodes to the right of the visualisation.

[insert Figure 2.]

However, an important change in pattern has occurred. In the above outdegree visualisation, we begin to see some family members emerge as large nodes (coded in turquoise) towards the lower left of the network. This means that whilst TSOs and policy makers are over-represented in both the indegree and outdegree categories, the family users emerge primarily in the outdegree category with few *inbound* connections from the other stakeholders. An analysis of their outbound connections revealed that they mostly sought to connect with powerful stakeholders (TSOs and policy makers) through direct messages (tweets and retweets) in which they mentioned the names of the powerful stakeholders. Figure 3 further illustrates the finding about the families’ outdegree status and shows that a family member is leading the top posters in the network, followed (in a clockwise direction) by TSOs and prison services.

[insert Figure 3.]

One of the functions of the outdegree measure is to identify who initiates the most connections/ties/conversations with others. It is a measure of influence in that it reveals the users who have good knowledge of other stakeholders in the network and can disseminate information across fragmented networks (Gruzd et al. 2016). That said, in a network of stakeholders where less powerful members rely on the more powerful for information and support, connections should ideally be reciprocal and conversational ties inclusive. In the absence of this, less powerful actors with a low indegree value (low inbound connections) but a high outdegree status (high outbound connections) can be considered marginal if their efforts to communicate with the powerful stakeholders or interact with them are being unreciprocated, hence their low indegree status. Affected families are the main stakeholders experiencing this form of marginality. They are active in their unreciprocated efforts to communicate with others, mainly with policy makers and TSOs, which also explains the high indegree status of these two users.

*Total degree centrality*

A user’s total degree value reflects a combination of their outdegree (outward connections) *and* indegree (inward connections) counts. Those with high total degree values are the most well-connected actors. In Figure 4 below, the same policy makers and TSOs with high indegree and out degree centrality are leading total degree centrality and are located towards the centre of network. They are represented by the largest green, and turquoise nodes to the left, as well as the orange and blue nodes to the right.

[insert Figure 4.]

However, the visualisation shows another change in pattern: the families no longer have a strong presence in the visualisation, indicating that they lack total degree centrality. This is in part a reflection of the families’ low, average indegree value of 27 which is 89% *lower* than their average outdegree count of 115. In contrast, policy makers and TSOs with high total degree centrality, had high indegree counts. For example, the most prominent policy makers in the network (the four senior ministers, one shadow minister, two state departments) had a high, average indegree count which, at 1453, was 90% *higher* than their average outdegree count of 144. The high indegree values attained by the policy makers and TSOs contributed to their high total degree centrality, which designated them as the most prominent and well-connected members of the network. Further analysis of the inbound and outbound connections of both parties revealed that whilst their inbound connections were initiated by the other stakeholder including the families, the outbound connections or messages (mentions, tweets, retweets, and replies) initiated by TSOs and policy makers were directed mainly at each other. Thus, they created an institutionalized vertical connection within the network population. Some outbound connections made by TSOs and policy makers were also directed at: the media; other organisations that provide services to people in prison and their families; and law firms. In addition, one national TSO also directed messages towards financial donors. Families were largely ignored, and although as mentioned earlier they reached out to both TSOs and policy makers (through mentions in tweets and retweets), their efforts were generally unreciprocated.

**Assessing domination and marginality through macro-level analysis of network structure**

Whilst centrality analysis focuses on micro-level measures, which analyse information flows at the level of each node/user, macro-level (network wide) measures are useful for analysing information flows and the interconnectedness of a conversational network. The measures include centralisation, modularity, density, and reciprocity (Gruzd et al. 2016) and the relevant values for our dataset are discussed below.

Centralisation assesses whether one or few central nodes/users dominate the flow of information in a network, indicating a high centralisation of information. It is measured on a scale of 0-1 and if the score is closer to 1, the network has a high centralisation of information. Our network’s centralisation value of 0.048780 is lower than 1, which indicates that it is a decentralised network with no nodes/users leading the entire flow of information. Instead, information is being transmitted between many actors across the network. That said, TSOs and policy makers are leading total degree centrality in several clusters within the network and are therefore prominent actors who are also likely to be identified as such by amplificatory platform algorithms.

Another useful measure of network structure is modularity which assesses whether the clusters within a network are distinct communities. A higher modularity score (above 0.5) indicates that the clusters in a network are not close-knit communities. Instead, they are divided and less likely to overlap. In terms of our network’s modularity, as Figure 5 below shows, the network of stakeholders consists of separate colour-coded clusters. It also has a moderate modularity value of (0.629200). This indicates that the network is fragmented, although there are areas of overlap comprising strong connections between some nodes/users who are engaging in the same conversations and interacting closely with each other. However, this is happening mainly within, but not across, clusters (Gruzd et al. 2016). Again, this not surprising given the fragmentation of the network into clusters. As noted above, TSOs and policy makers formed insular connections with each other. In the network visualisation of total degree centrality (see Figure 5), TSOs and policy makers dominate the denser green, turquoise and orange clusters, whilst families lack total degree centrality, and are more likely to be found in the turquoise cluster located towards the lower left margins of the network.

[insert Figure 5.]

Density is another macro-structural measure. It compares the proportion of existing ties to the total number of ties (connections with others e.g., through mentions) users can make within a network. It is therefore useful for assessing how close-knit users are in the network. A density score closer to 1 indicates the network is a more cohesive, high-density network. Within such closely connected networks, all or most users are connected to each other. But a caveat concerning the reliability of the density measure is that it is difficult to sustain numerous connections in large social networks and this can in some cases, account for low density values (Gruzd et al. 2016).

Our network attained a low-density value of 0.001038 which indicates that overall, it is not as well-connected as it can be. This low-density value of the network of stakeholders discussing the impact of imprisonment on families points to the overall low-level connectivity and inclusivity within the network. It is in part, likely to be a reflection of the cohesion between the major actors (TSOs and policy makers) which marginalises families (most of whom strive to connect with them). If TSOs and policy makers were to reciprocate, there would be more multidimensional ties within the network. In other words, the network would be more well-connected and inclusive, and information transmission more rapid. However, because of the current fragmentation into clusters, information diffusion and flow will be more rapid at intra-cluster, rather than inter-cluster levels. Users disconnected due to limited reciprocal connections/ties with the major influencers are left behind. As noted earlier, most families in the dataset fall within the disconnected category.

Diameter, another measure of network structure, further illustrates the dynamics of information transmission. It calculates how far information travels across the network, and measures the distance (path length) between the two nodes located farthest away from each other (Gruzd et al. 2016). In the network of stakeholders discussing the impact of imprisonment on families, the average path length is 28 indicating that information may have to travel through 28 connections to get from one end of the network to the other. This represents a low rate of information transmission or spread and reinforces the earlier observed low level of network density and overall cohesion.

Another important, macro-level measure of graph network cohesiveness is ‘reciprocity’ which assesses the extent to which mutual ties/conversations exist in the network. A reciprocity value of 1 indicates that all the ties are mutual (all users reply to each other) which is unlikely in a large social network. But our network’s low reciprocity value of 0.185500 suggests that many connections are uni-directional or one-sided. This complements the findings regarding centrality, particularly the position of the families as the marginal, outdegree group who were sending many unreciprocated messages/ties/connections.

*Methodological limitations*

Social media analysis using the computational methods presented here, necessarily focuses on a proportion of available content and cannot capture the full range of large-scale data increasingly available amidst rising SNS usage. Furthermore, such analysis excludes non-users and private accounts from analysis. Therefore, whilst the methods may be useful for analysing large scale data to detect network dynamics, their main limitation is that the data may be limited in the ways explained above. Social media analysis is also affected by platform filtering algorithms that restrict the parameters of knowledge production. Although this study seeks to demonstrate how the algorithms operate and suggest some remedial strategies, it is itself not free of their limiting influence. Another limitation concerns the presence of bogus accounts, some of which are operated by software robots known colloquially as bots (Ferarra et al. 2016). However, WebDataRA’s technical enhancements provide several account authentication techniques. For example, the tool provides metadata that can be used to ascertain the veracity of an account. High frequency posts by accounts with limited connections to others (through ‘follows’ for example) can indicate unreliability (Yang et al. 2019). But in our study, we were mindful that many families could fall within this category. Our verification processes therefore, involved a painstaking process of examining the content posted by accounts we tagged as suspicious and this lasted several months.

An additional methodological issue relates to the implications of focusing on a subset of extracted tweets. However, the decision to select the same number of tweets (10,000 per stakeholder group) did not impose any artificial changes on the network structure. This was evident because, we used Netlytic’s name network analysis to test how different numbers of tweets would affect network structure, and overall, the pilots revealed the same microstructural and macrostructural network dynamics and patterns, regardless of the number of tweets selected for each group. We eventually selected a subset of 10,000 tweets per group, to provide a broad view of network dynamics and patterns whilst ensuring that the selection was not too voluminous and unmanageable. Furthermore, the decision to select the same number of tweets per stakeholder was useful for ensuring that any observed differences in levels or degrees of prominence amongst the stakeholders (e.g., via indegree and/or total degree status), could not be attributed to differences in the frequency of tweets per stakeholder group.

The geographical location of the extracted tweets is yet another potential methodological issue since the experience of imprisonment and its impact on families varies across different jurisdictions. The analysis presented here focuses on tweets from one jurisdiction: England and Wales. Future research can explore other jurisdictions.

Despite these methodological limitations and issues, data-driven computational methods of the kind applied in the study are useful for retrieving and analysing the large-scale datasets (or big data) increasingly emerging from growing human interaction with data-driven technologies such as SNSs. The methods are now applied extensively across a range of social science, computer science, and other disciplines, to study social problems. Furthermore, a key methodological strength is that such methods can be used to reach stigmatised and typically hidden populations, and can also be used to generate and analyse more substantial data than other research methods regularly applied in social science research.

**DISCUSSION AND CONCLUSION**

This paper’s aim was to unravel how knowledge about the impact of imprisonment on families is produced via AI technologies such as SNSs, and the paper drew on a study of conversational networks on the SNS, Twitter. It is important to explorewhether the knowledge production process reproduces the relative powerlessness and marginality of affected families, increasing the potential for epistemic domination.This possibility provided significant motivation for the study which employed computational Social Network Analysis (SNA) techniques.

The SNA revealed that families appeared to be more prolific at reaching out to others and posting content. A family member was identified as the top poster in the network. However, the families were the least central members of the network because they were directing most of their messages at the key actors: policy makers and TSOs, who were, as we have seen, typically not reciprocating. Therefore, the marginalisation of affected families in the form of lack of recognition, limited influence, and lack of reciprocity are social dimensions that appeared to re-emerge in the online world, disconnecting them from knowledge production and policy making spaces.

Meanwhile the SNA revealed an insular, reciprocal connection between TSOs and policy makers. This insularity is evident in the finding that their outbound messages were mainly directed at each other. Regarding the TSOs, this can in part be explained by their altruistic efforts to target their campaigns for ameliorative family support, at policy makers who are the most powerful group able to introduce such measures. An unintended consequence, however, is that such insularity can distort knowledge production processes by triggering the algorithmically-produced ‘filter bubbles’ phenomenon (DiFranzo and Gloria-Garcia, 2017) whereby oppositional discourses or information that seem irrelevant to a user are filtered out by platform algorithms. This can foment digital ‘echo chambers’ (Sunstein, 2018) which entrench insularity, erode discursive pluralism, reinforce social divisions, and trap like-minded social actors within the narrow confines of entrenched ideologies. Another unintended consequence is that the families’ voices can become disconnected and marginalised from the centre of the conversational network, thus silencing and disempowering them. Indeed, our findings point to a disempowerment model of knowledge production, which excludes affected families from the discursive space occupied by the major users (policy makers and TSOs) whose discourses and actions impact directly on the experiences and overall wellbeing of the families**.** An implication of this finding is that despite the aforementioned claims of the empowerment model, studies such as this are needed to reveal how existing social conditions and platform algorithms interactively automate historical patterns of marginalisation and disempowerment, calling into question the assumptions of the empowerment model.

The SNA in this study further shows that TSOs and policy makers are the most well connected given their reciprocal ties and their high indegree status. Studies of reciprocity on SNSs have similarly found that high-profile groups such as politicians are sometimes insular users who mostly communicate with each other or with others of a similar status (e.g., Keller 2020). They do not readily respond to efforts from ordinary citizens to engage with them, although there is evidence that where they do, it builds goodwill and public trust (Tromble 2018). Unreciprocated mentions can be an indication of lack of recognition and marginalisation regardless of size of ‘likes’, ‘followers’ and so forth. Some families in our SNA had many followers, running into thousands. But they were almost completely marginalised from spaces at the centre of the conversational network dominated by TSOs and policy makers. These findings point to the influence of sociotechnical dynamics elucidated below.

*Intersecting sociotechnical barriers to equitable knowledge production: The nexus of algorithmic amplification, digital capital, positionality, and ascribed credibility*

The study shows how algorithmic amplification, a technical aspect of AI-driven knowledge production, can intersect with social dimensions to form sociotechnical dynamics of knowledge production. First, since TSOs and policy makers are more prominent members of the conversational network, amplificatory algorithms are more likely to enhance the visibility of their discourses. By highlighting their discourses, the amplificatory algorithms can reproduce an existing social condition which is unequal access to digital capital. They can confer even greater digital capital on TSOs and policy makers, particularly the power and ability to discursively frame key issues and create social impact using digital technologies such as SNSs (Van Dijk 2005). Relatedly, amplificatory algorithms are likely to highlight their discourses regardless of authenticity or credibility. This would reinforce the higher position that TSOs and policy makers already occupy in the ‘hierarchy of credibility’ given their positionality (higher social status) compared with affected families. The study’s findings therefore expand Becker’s (1967) concept of the ‘hierarchy of credibility’ by demonstrating how it can transcend the offline world and manifest in online spaces where much social interaction occurs. The study also demonstrates that just as ‘opinion leaders’ exist in offline social networks (see, Dubois et al. 2020), in the online conversational network of stakeholders discussing the impact of imprisonment on families, TSOs and policy makers emerge as higher status ‘opinion leaders’. They possess some characteristics that studies associate with that position in both offline and online contexts. One such quality is ascribed credibility based on their social status and their ostensibly good knowledge of, or expertise in, the subject matter of interest to the network. They also have the digital capital to develop and apply strong communicative strategies. An additional characteristic is total degree centrality. Studies show that this form of centrality in a network, empowers opinion leaders to expedite information diffusion (e.g., Dubois et al. 2020). In SNSs, these characteristics enhance the potential for TSOs and policy makers to benefit from algorithmic amplification which can provoke ‘digitised epistemic domination’ (Ugwudike 2020: 483) and foment the disempowerment model of knowledge production.

Indeed, the influence of algorithmic amplification is evident in our earlier finding that policy makers and national TSOs were the actors most readily identifiable by platform algorithms during our search for relevant content. Families affected by imprisonment were less visible, suggesting that their offline marginality is being replicated online. In addition, although the retrieved dataset comprised the same number of tweets from each stakeholder (10,000 tweets each), the families experienced lack of reciprocity (low inbound connections, hence their low indegree status). This contributed to their lack of total degree centrality which in turn meant that their discourses were unlikely to benefit from algorithmic amplification. The impact of algorithmic amplification can therefore be such that the positionality (higher social status) of prominent actors, such as TSOs and policy makers, is reproduced and maintained online.

An implication of these power-laden dynamics is that they can exacerbate the invisibility of less powerful groups such as affected families and limit the comprehensiveness of knowledge production. Influential members of a network can, however, support the efforts of marginal contributors to attain visibility and participate in knowledge production. For example, our study suggests that in digital spaces where discourses about criminal justice and its impact are being produced, influential users such as TSOs and policy makers can replicate the efforts of marginal but key stakeholders such as affected families to engage with them. They can do so via a ‘feed forward’ process that involves mentioning the families in tweets, retweets and replies. These will amplify the families’ discourses to a wider audience given the significant attention afforded TSOs and policy makers. Such practices will help replace the current disempowerment model with an empowerment alternative that places affected families in good stead to inform dominant discourses and frame key issues. This can elevate them in Becker’s (1967) ‘hierarchy of credibility’.

The concept of ‘linking’ social capital is also relevant here. It is a form of capital that is based on ties formed within an ecosystem where ordinary citizens, for example, families affected by imprisonment, are able to interact with key institutional actors such as policy makers (Szreter and Woolcock, 2004). This form of capital can ensure that the voices of those most affected by a social problem are embedded in decisions made by powerful individuals and institutions. In proposing greater reciprocity from TSOs and policy makers, and in its description of SNSs as sociotechnical assemblages, this paper subscribes to the view that technological design and application are mediated by human rationality and choice. TSOs and policy makers can choose to support the amplification of affected families. Our study shows that many affected families do strive for online visibility given their high outdegree centrality which is driven mainly by their efforts to engage with key actors: policy makers and TSOs. This became more apparent as the COVID-19 global pandemic escalated from March 2020 onwards when most families adopted a yellow background and the hashtag ‘prisonerspeopletoo’. The families adopted these devices to increase the visibility of their tweets about the harms and distress caused by the embargo on prison visits as well as the prison lockdown implemented during the pandemic. But they experienced deamplification and subordination through sociotechnical barriers. What this indicates is that as an SNS, Twitter is a space where struggles for influence and visibility occur. In addition, it draws attention to the fact that, with the advent of the pandemic and social distancing measures, SNSs are more likely than ever to become important sites of social connection, participation, and activism. But algorithmic amplification, differential access to digital capital, positionality, and ascribed credibility are intersecting sociotechnical dynamics that can hinder efforts to garner visibility.

In sum, this paper expands the scholarship on intersections of AI and criminal justice. It also advances the current scholarship on using web science methods to study how marginality and epistemic domination are reproduced in digital society with implications for knowledge production concerning criminal justice. The paper focuses on Twitter, but future research can explore other SNSs. Future research can also use the methodological approach presented here to explore how other social categories such as race can influence online knowledge production concerning the impact of imprisonment on families. This warrants criminological attention given the longstanding overrepresentation of black people in the prison population (e.g Ministry of Justice – MOJ 2019). Furthermore, their noted vulnerability to the harms of data-driven technologies increasingly informing criminal justice policy and practice (Ugwudike 2020; Benjamin 2019) may also heighten their susceptibility to the disempowerment model of knowledge production on SNSs. As Benjamin (2019: 26) notes in her incisive analysis of biases entrenched in some data-driven technologies, ‘platforms like Twitter, Instagram, and YouTube…encode more insidious forms of inequity in the very design of their products and services’. She provides examples of how Twitter algorithms can for example, perhaps inadvertently amplify certain discourses thatreinforce racial discrimination.

Additional penological issues that can be explored using the computational SNA applied here are vast and include: the humanitarian and other impacts of the COVID-19 pandemic on people in prison and their families; the ‘joint enterprise’ doctrine; the Imprisonment for Public Protection sentence; and post-prison resettlement. These are generating much online debate, particularly among affected families, policy makers, TSOS, and other key stakeholders.

**NOTES**

1. All platform filters were removed before searches to limit algorithmic personalisation and tailoring of results to the account holder’s inferred identity and preferences. But it must be acknowledged that algorithmic automation influences the content visible on SNS.

Netlytic is an opensource web-based tool that analyses datasets retrieved from online/digital communications data.

Figures 1-5 do not show the edges/lines indicating connections between nodes/users because the edges have been removed to enhance the visibility of node sizes. Node labels (names) have also been removed for ethical reasons; to preserve the anonymity of Twitter users.

A node’s size denotes how well connected it is and its centrality value: the bigger the node size the more well connected it is and the higher the centrality value.

‘Cluster’ in this context refers to groups of nodes/users that form homologous communities and are more connected to each other than others in the network. They share similar qualities and perform a similar function or role in the network.

**REFERENCES**

Becker H (1967) Whose Side Are We On? *Social Problems* 14(3): 239-247.

Benjamin R (2019) *Race After Technology: Abolitionist Tools for the New Jim Code*.

Cambridge: Polity Press.

Berg M (2014) The production of online space: Facebook and algorithmic metapower.

Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.981.475&rep=rep1&type=pdf> (Accessed June 2020)

Bonilla Y and Rosa J (2015) #Ferguson: Digital protest, hashtag ethnography, and the racial

politics of social media in the United States. *American Ethnologist* 42(1): 4–17.

Bucher T (2012) Want to be on the top? Algorithmic power and the threat of invisibility on

Facebook. *New Media & Society* 14(7): 1164–1180.

Carr L (2020) Web Data Research Assistant (Version 3.5.11). Available at

<https://chrome.google.com/webstore/detail/web-data-research-assista/kcdbekkmigohaijilebpaeoopcgjbbdm> (Accessed May 2020)

Condry R and Smith PS (eds) (2018) *Prisons, Punishment, and the Family: Towards a New*

*Sociology of Punishment?* Oxford: Oxford University Press.

DiFranzo D and Gloria-Garcia K (2017) Filter bubbles and fake news. Crossroads: *The*

*ACM magazine for students.* Available at: <https://xrds.acm.org/article.cfm?aid=3055153> (Accessed December 2018).

Dubois E, Minaeian S, Paquet-Labelle and A, Beaudry S (2020) Who to trust on social

media: How opinion leaders and seekers avoid disinformation and echo chambers. *Social Media and Society* 6 (2) 1-13.

Ferrara E, Varol A, Davis, C, Menczer F, and Flammini A (2016) The rise of social bots.

Available at: <https://dl.acm.org/doi/pdf/10.1145/2818717> (Accessed September 2020).

Fuchs C (2013) *Social Media: A Critical Introduction*. London: Sage

Gillespie T (2010). The politics of ‘platforms.’ *New Media & Society* 12(3): 347–364.

Graves L and Anderson C (2020) Discipline and promote: Building infrastructure and

managing algorithms in a “structured journalism” project by professional fact-checking groups. *New Media and Society* 22(2): 342–360

Greer C and McLaughlin E (2011) Trial by media: Policing the 24–7 news media sphere and

the politics of outrage. *Theoretical Criminology* 15, 23–46.

Gruzd A (2016) *Netlytic:* Software for automated text and social network analysis*.*

Available at <http://Netlytic.org> (Accessed May 2020)

Gruzd A, Paulin, D and Haythornwaite, C (2016) Analysing social media and learning

through content and social network analysis: A faceted methodological approach. *Journal of Learning Analytics* 3(3): 46–71.

Hanneman RA and Riddle M (2005) *Introduction to Social Network Methods*. Riverside, CA:

University of California.

Hutton M, Moran D (eds) *The Palgrave Handbook of Prison and the Family*.

Basingstoke: Palgrave Macmillan.

Keller (2020) To whom do politicians talk and listen? Mapping Swiss politicians’ public

sphere on Twitter. *Computational Communication Research* 2 (2): 175-202.

Koumchatzky N and Andryeyev A (2017) Using deep learning at scale in Twitter’s

timelines.Available at: <https://blog.twitter.com/engineering/en_us/topics/insights/2017/using-deep-learning-at-scale-in-twitters-timelines.html> (Accessed September 2020).

Lanskey C, Losel F, Markson L and Souza, K (2018) ‘Prisoners Families, Penal Power, and

the Referred Pains of Imprisonment’. In: Condry R and Scharff Smith P (eds) *Prisons, Punishment, and the Family: Towards a New Sociology of Punishment?* Oxford: Oxford University press.

LSE Digital Communications (2018) Social media platforms and demographics*.* Available

at:<https://info.lse.ac.uk/staff/divisions/communications-division/digital-communications-team/assets/documents/guides/A-Guide-To-Social-Media-Platforms-and-Demographics.pdf>(Accessed January 2020).

NAO (2020) What are third sector organisations and their benefits for commissioners?

Available at: <https://www.nao.org.uk/successful-commissioning/introduction/what-are-civil-society-organisations-and-their-benefits-for-commissioners/> (Accessed September 2020).

Maguire M, Williams K, and Corcoran, M (2019) ‘Penal drift’ and the voluntary sector. *The*

*Howard Journal* 58 (3): 430-449.

MOJ (2019) *Statistics on Race and the Criminal Justice System 2018: A*

*Ministry of Justice Publication under Section 95 of the Criminal Justice Act 1991*. London: MOJ.

Ofcom (2020) *Online Nation: 2020 Report*. London: Ofcom.

Petre C, Duffy BE, and Hund E (2019) “Gaming the System”: Platform paternalism and the

politics of algorithmic visibility. *Social Media and Society* 1-12.

Sunstein C R (2018) *#Republic: Divided Democracy in the Age of Social Media*. Princeton,

NJ: Princeton University Press.

Szreter S and Woolcock M (2004) Health by association? Social capital, social theory, and

the political economy of public health. *International Epidemiological Association* 33(4) 650-667.

Tromble R (2018) ‘Thanks for (actually) responding! How citizen demand shapes politicians’

interactive practices on Twitter’. *New Media & Society* 20(2): 676–697.

Ugwudike P (2020) Digital prediction technologies in the justice system: The implications

of a ‘race-neutral’ agenda. *Theoretical Criminology* 24 (3): 482–501

van Dijk J (2005) *The Deepening Divide: Inequality in the Information Society*.

Thousand Oaks, CA: Sage.

Walsh, J. and O’Connor, C. (2019) Social media and policing: A review of recent

research. *Sociology Compass* 13(1): 1-14.

Wolch J (1990) *The Shadow State: Government and Voluntary Sector in Transition*, New

York: Foundation Centre.

Yang K C, Varol O, Hui P M, et al. (2019) Scalable and Generalizable Social Bot Detection

through data selection. ArXiv:1911.09179 [Cs]. Available at: <http://arxiv.org/abs/1911.09179> (Accessed October 2020).