

Metadata, Surveillance and the Tudor State

by *Ruth Ahnert and Sebastian E. Ahnert*

‘We kill people based on metadata.’
General Michael Hayden, former director of the NSA, 2014.¹

On 5 June 2013 the *Guardian* newspaper published an exclusive article, based on information leaked to them by Edward Snowden, that revealed a large-scale effort by the United States National Security Agency to collect domestic email and telephone metadata from the US telecommunications company Verizon. In the following days it would emerge that this was part of a more widespread and systematic programme, named Prism, which allowed the Agency to gain access from nine Internet companies to a wide range of digital information on foreign targets operating outside the United States. The NSA defended itself against allegations that its activities contravened the Fourth Amendment by arguing the difference between data and metadata: that the contents of the phone calls, emails and other communications remained private. Citing a Supreme Court ruling from 1979, however, they maintained that Americans had no reasonable expectation that the metadata produced by their telephone and Internet communications – the names of those making and receiving such communications, the times and dates when they were made, and the geo-location of each party – should remain private.² Now, in the era of the Trump administration, there is renewed unease. Trump’s appointment to head the CIA, Mike Pompeo, has previously stated that ‘Congress should pass a law re-establishing collection of all metadata, and combining it with publicly available financial and lifestyle information into a comprehensive, searchable database’.³

To a layperson it is perhaps unclear why this should be of concern. How much can metadata really reveal? This essay seeks to show that surprisingly deep insights can be gleaned from metadata by applying a range of easily available network analysis algorithms to a body of metadata collected by another government. Our source is the State Papers held at the British National Archives, specifically the correspondence held in this archive dating from the accession of Henry VIII in 1509 until the death of

Queen Mary, University of London
University of Cambridge

r.r.ahnert@qmul.ac.uk
sea31@cam.ac.uk

Elizabeth I in 1603, which comprises 132,747 unique letters (duplicates have been excluded). Analysis of this archive shows that we can observe not only broad patterns of communication but also anomalous behaviour, and can make predictions about people likely to be trading in conspiracies or illicit intelligence. These discoveries demonstrate the power of such methods for the study of history. This power, however, is merely a shadow of that wielded by government bodies and private companies. The following pages also act as a warning, therefore, about the potential uses and abuses of the metadata we generate with each of our digital communications.

NOT-READING

The Tudor State Papers constitute the working papers of the monarch's principal secretary (equivalent to the modern Secretary of State), comprising heterogeneous materials which include letters, papers, reports, memoranda, treatises, grants, commissions, state trials, treaties and ambassadors' reports. Although letters are just one among these categories, for the Tudor period alone we have 132,747 unique letters, and many more if we count duplicates in draft or copies. Upon entering the hands of the monarch's principal secretary, these state papers were regarded as his personal papers, meaning that they were often subsequently incorporated into that individual's own private archives. For this reason the state papers were scattered across numerous locations and collections, and others were lost. Whilst much of the surviving material was subsequently requisitioned by the State Papers Office, and later incorporated into the Public Record Office (now part of the National Archives), many of the state papers still reside elsewhere, most notably within the Lansdowne, Harleian and Cottonian collections of the British Library and those at Hatfield House. State Papers Online (SPO), an online repository launched by the company Gale in 2009, reunites many of these documents in a virtual environment, linking high-quality reproductions of the manuscripts to their fully text-searchable Calendars (a series of chronologically arranged catalogues, begun in 1825), which provide detailed summaries of the documents' contents.⁴ Most important, SPO provides metadata for each document, which in the case of the letters contains valuable relational information akin to that collected by the NSA, including: name of sender, name of recipient, date of composition (where given), place of writing (where given), unique document identifiers, and a content description.

The potential value of this resource is immense, but when Gale launched SPO it seems to have assumed scholars would use this data in much the same way that people use analogue archives: by surfing through the documents and Calendars (the 'browse' function), or by calling up specific letters or texts already identified (the 'search' function). However, scholars in the humanities increasingly realize that computational approaches allow the data that underlies sites like this to be analysed as a whole. Such approaches

have been described variously as ‘distant reading’, ‘macroanalysis’, and ‘cultural analytics’.⁵ Regardless of the name, the concept is the same: rather than trying to read all the material, we can use available computational models to aggregate and analyse it. The approach of ‘not-reading’ is an attractive one when you are working with large quantities of data.⁶ The NSA’s assurances that it was not reading the contents of emails were not merely a means to sidestep complaints of breaking the Fourth Amendment they also spoke to the genuine methodological and statistical necessity of working on the scale of billions of communications. Although the correspondence within SPO is nowhere near this extensive, a single person attempting to read all 132,747 letters at the rate of four letters per hour for eight hours a day during the working week would take almost sixteen years (without holidays) to make it all the way through, not allowing for very long missives, difficulties of reading illegible hands or letters in cipher, and slower reading speeds in foreign languages. And of course close reading would just generate more data, which would have to be stored in some way, because it could not effectively be kept in memory. But even if all those barriers were surmounted, it would be difficult for an individual to gain a true sense of the overarching shape of that archive and the relative significance of its constituent parts, or to describe it to a reader. In order to ‘not-read’ this entire archive, however, we needed to bypass the interface designed by SPO and get at the underlying XML data. This access was granted in 2013.⁷ Similar access to pay-walled repositories like SPO has been facilitated in the United Kingdom by the ‘text and data mining copyright exception’, which enables researchers to make copies of copyright material for computational analysis as long as they have a personal or institutional subscription.⁸

Using the XML files, we separated the letters from other kinds of document by extracting all records that had an entry in both the ‘author’ and ‘recipient’ fields. While these metadata fields made the correspondence relationships contained in SPO readily available, their contents only became useful for computational analysis after an extensive data-cleaning process. In particular, the author and recipient fields needed thorough disambiguation and de-duplication, for a number of reasons: variant spellings of early modern names; letters addressed to a titular office rather than a named individual (e.g. the Archbishop of Canterbury); changing office-holders; changing titles over a person’s lifetime as they accrued honours and offices; and women’s names changing due to marriage.⁹ Furthermore Tudor families were often unimaginative in their naming practices. Just a handful of men’s and women’s names dominated, with some recurring over several generations of one family. Simply put: single people could be referred to by multiple different names; and single names or titles could refer to multiple people. The complexity of the sender and recipient metadata fields meant that although we initially extracted 37,101 unique name entities, there were in fact only 20,656 unique correspondents. This cleaning process took eighteen months to complete (nine months of them full-time). However, once it

was finished we were able to reconstruct the entire Tudor letter network archived in SPO, and subject it to the analytical tools developed in the field of network science.

Within this field a number of publications anticipate the NSA's methods by using mobile-phone data (albeit privacy-safe and anonymized) to uncover the structure and dynamics of social networks at different levels, 'from the small-scale individual's perspective, to the large-scale collective behaviour of the masses'.¹⁰ But such methods are not restricted to electronic and cellular communication; wherever interactions produce or are reducible to metadata, network analysis is possible. In a series of key publications in the 1990s and early 2000s, scholars such as Albert-László Barabási, Réka Albert, Duncan J. Watts, and Steven Strogatz showed that a huge variety of real-world networks – for example, neural networks, transport networks, biological regulatory networks, and social networks – share an underlying order and follow simple laws, and therefore can be analysed using the same mathematical tools and models.¹¹ These methods are equally useful to our understanding of Tudor communication. They let us find out not only about patterns of communications and social structure, but also about the history of the archive, its making and its survival. Historic letter collections are subject to the vicissitudes of time, but also to the bias of collectors. The correspondence network found in SPO represents a view of the world from the perspective of the successive principal secretaries, compiled from the reports that they received. This view is therefore subject not only to the secretaries' management of their contacts, but also to their personal filing habits, and to the subsequent care of their clerks, their families, and other keepers of their papers.

GLOSSARY

Nodes: The entities in a network, in this case these are the correspondents.

Edges: The connections between those entities, which in this case are connections marked by letters.

Degree: The total number of a given node's edges. Here the degree is calculated as the total of the in-degree and out-degree.

In-degree: The total number of people who wrote to a given person.

Out-degree: The total number of people to whom a given person wrote.

Betweenness: A measure of a node's infrastructural importance. For any two nodes in a network, there is a shortest path between them, and betweenness tells us how many of these shortest paths go through a given node. Both communication hubs and 'bridges' can have high betweenness.

Strength: the total number of letters sent and received by a given person.

In-strength: the total number of letters received by a given person.

Out-strength: the total number of letters sent by a given person.

Power-law distribution: This is a highly skewed distribution of a quantity, such that we are n times less likely to see a value m times as large, where m is n raised to some power. As a simple example, if that power is 2, then $n = m^2$. And if the number of connections (degree) of an individual in a network has this kind of power-law distribution, then for any given value of the number of connections we will see four times fewer individuals with twice as many connections, nine times fewer individuals with three times as many connections, a hundred times fewer individuals with ten times as many connections, etc.

Eigenvector centrality: A measure of a node's proximity to power. As Stephen Borgatti puts it: 'even if a node influences just one other node, who subsequently influences many other nodes ... then the first node in that chain is highly influential'.¹²

THE TUDOR LETTER NETWORK

Networks are composed of entities, known as 'nodes', and the connections between those entities, known as 'edges'. In this letter network we have a system of 20,656 nodes (the correspondents) and 37,087 edges (communications marked by one or more letters). Once the network is reconstructed, a few lines of computer code can inform us about its overarching structure.¹³ One of the simplest things we can measure is a node's 'degree'. In this directed network the degree is calculated as the total of the number of people who write to a given person (called the in-degree), plus the total number of people to whom they wrote (their out-degree). Even this simple measure can provide us with a surprisingly helpful overview of the shape of the archive.

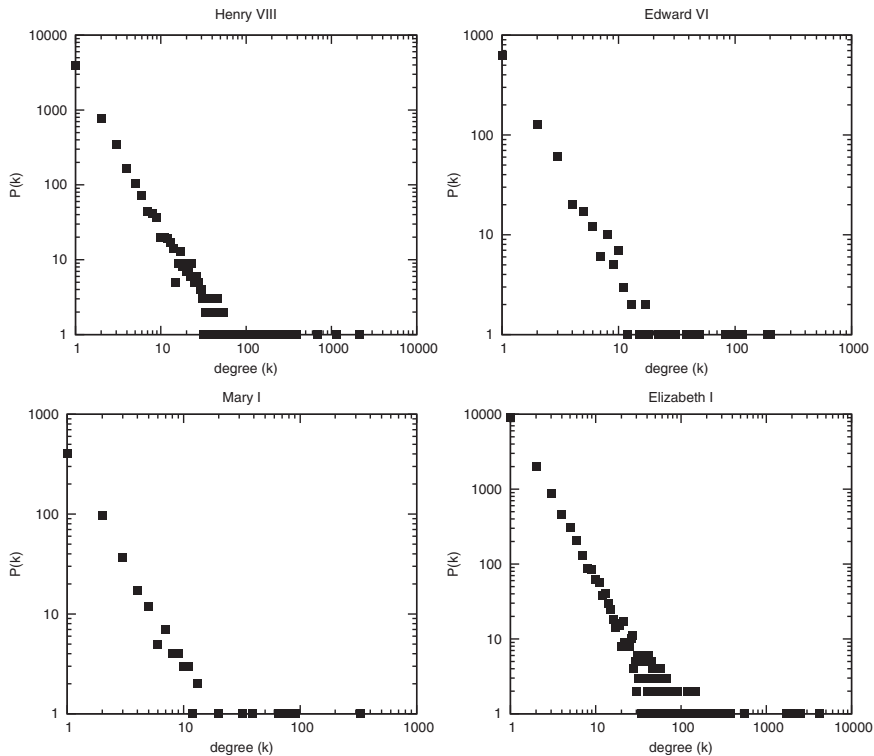
If we rank people by their degree from the highest to lowest we see an interesting distribution. For Henry VIII's reign the person with the highest degree is Thomas Cromwell (principal secretary to Henry VIII, Lord Privy Seal and Chancellor of the Exchequer) with a degree of 2,149, followed by the king (1,134), and Cardinal Thomas Wolsey his Lord Chancellor (682). Since the archive derives from the papers of the principal secretary, it is not surprising that he should have the highest number of correspondents. What is surprising, however, is that of 5,785 people in the Henrician letter-network only seventeen people (0.3%) have 100 or more correspondence connections to other individuals, and only 127 (2.2%) more than twenty. By comparison, 5,225 people (90.3%) have five or fewer connections, and 3,937 (68.1%) have only one. This is a very long tail, and it comprises many different kinds of people, from minor administrators to petitioners. The latter are

represented often by a single letter, usually directed to the Principal Secretary, seeking favour or aid of some kind.

If we plot the degrees of everyone writing or receiving letters in the reign of Henry VIII on a graph (Fig. 1a), we can see that they follow what is known as a power-law distribution. The distribution of data points within a power-law distribution is so broad across several orders of magnitude that it is normally plotted on logarithmic axes (in this case increasing in multiples of ten). On these axes a power-law distribution appears as a straight diagonal line, meaning that whether we look at the network as a whole or just within a specific region, we will always find a few very well-connected 'hubs' and a very large number of nodes with a relatively small number of connections. This power-law distribution is a classic feature of real-world networks (the same distribution is found in power grids, social networks, and the world-wide web among many other examples).¹⁴ But what is striking is that this pattern is almost identical for all Tudor reigns: Figures 1b-1d chart the distribution of the number of correspondence connections (known formally as 'degree distribution') for the reigns of Edward VI, Mary I and Elizabeth I. Even though there are a very different number of correspondents in the various reigns because of their diverse durations and variants in documentary survival rate, the proportion of people in the network with a degree of one (that is, who corresponded with only one other individual) remains remarkably constant: in Edward VI's reign that proportion is 67.8%, in Mary's it is 67.5%, and in Elizabeth's it is 66%.

By direct contrast we can see how single people can dominate the network, occupying a large majority of all edges (connections). In the reign of Henry VIII, of the 10,666 edges 2,149 link with Cromwell (20.1%); in Edward's reign the Privy Council has 201 of the 1,433 edges (14%); in Mary's reign the queen has 329 of the 906 edges (36.3%); and in Elizabeth's, William Cecil, Lord Burghley (Elizabeth I's principal secretary, her Lord Privy Seal, and Lord High Treasurer) has 4,168 of 26,813 edges (15.5%). These people are the hubs of our network, both the gravitational centre of this epistolary network and the key members of Tudor government – the monarchs, principal secretaries, the Privy Council, Treasurers, Lord Chancellors, and in Edward's reign, the Lord Protector – pulling in information and intelligence on the one hand, and fielding petitions on the other.

This overview is a powerful starting point for analysis. One might be tempted to argue these findings are obvious or just confirm what we already knew: that government figures dominate the archive. But that perception is actually misleading. As Matthew Lincoln has persuasively argued, scholars 'are particularly susceptible to confabulating these post-facto rationalizations with the idea that we somehow *knew* the results of this quantitative work already (and, implicitly, that we didn't need to waste our time doing it)'. Rather he is keen to stress the difference between what 'in retrospect sounds reasonable' from 'what we actually knew before'.¹⁵ Before the monumental data-cleaning task behind this project, it was not possible to compute



Figs 1a-1d. These four graphs chart the distribution of the number of correspondence connections (known formally as ‘degree distribution’) for the reigns of Henry VIII, Edward VI, Mary I and Elizabeth I.

the total number of correspondents (at the most basic level), let alone the distribution of communications between those correspondents. It is highly likely, for example, that estimates of the number of people in the archive corresponding with only one person would have been wrong by a factor of ten, or more. Grasping the true distribution of correspondence is vital if we are to recognize anomalous patterns of correspondence within this vast archive, or to detect people who were significant within the infrastructure of Tudor communication networks. While we need to grasp quantity to get at those significant correspondents to determine who are the statistical outliers, thereafter we need to develop bespoke computational measures that can tell us about significance beyond the obvious measure of ‘many’.

A good network measurement for examining the organization of communication networks is ‘betweenness’. For any two randomly selected nodes in a connected network we can find a shortest path between them; betweenness is an algorithm that measures how many of these shortest paths pass through a given node or edge. It therefore tells us something about the

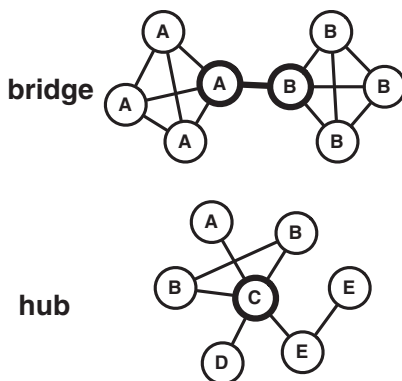


Fig. 2. Two 'toy' networks illustrating why both hubs and bridges benefit from the measure of betweenness. In both cases, the shortest path between any nodes bearing different letters (e.g. A and B, or B and D) necessarily passes through the bold outlined nodes, because there are no other possible paths.

infrastructural significance of a given node within the network as a whole, and tends to highlight two kinds of nodes: 'hubs' and 'bridges'. Figure 2 offers two 'toy' networks (not based on real data) to illustrate why both hubs and bridges benefit from this measure. In both cases, the shortest path between any nodes bearing different letters (e.g. A and B, or B and D) necessarily passes through the bold-outlined nodes, because there are no other possible paths. From these diagrams we can also extrapolate the reasons why hubs and bridges benefit from the measure of betweenness. Hubs are likely to lie on short paths because their large number of connections increases the likelihood that a short path will pass through them. Bridges, by comparison, are nodes that straddle otherwise disjointed communities, or what Ronald Burt has described as 'structural holes' in the network.¹⁶ Both the hub and the bridge are vital in communications networks for the effective movement of information.

The people with the highest betweenness (or infrastructural significance) in the reign of Henry VIII are:

- 1) Henry VIII, 2) Cardinal Thomas Wolsey, 3) Desiderius Erasmus, 4) Pope Leo X, 5) James IV of Scotland, 6) Ferdinand II of Aragon, 7) Richard Pace, 8) Francis I of France, 9) Leonardo Loredan, Doge of Venice, 10) Margaret of Savoy, 11) William Warham, 12) Privy Council, 13) James V of Scotland, 14) Louis XII of France, 15) Margaret Tudor, 16) Cuthbert Tunstall, 17) Silvestro de Gigli, 18) Thomas Dacre, Baron Dacre of Gilsland, 19) Thomas Darcy, Baron Darcy of Darcy, 20) Katherine of Aragon, 21) John Stewart, Duke of Albany, 22) Robert Wingfield, 23) Emperor Maximilian I, 24) Patrick Paniter, 25) Pope Julius II, 26) Sir Edward Poynings, 27) Tommaso

Spinelli, 28) Sir Richard Wingfield, 29) Thomas Ruthall, 30) Emperor Charles V.

These people fall into four main groups (with some appearing in two): secular rulers and popes (numbers 1, 4, 5, 6, 8, 9, 10, 13, 14, 15, 20, 21, 23, 25, 30); leading statesmen, secretaries of state, members of the royal household or privy council (2, 12, 24); religious leaders (2, 11, 16, 17, 29); and diplomats and correspondents at foreign outposts (7, 18, 19, 22, 26, 27, 28). The one person who does not fit under these headings is Erasmus, who forms his own largely self-contained intellectual community within the archive. Focusing on the other twenty-nine people, however, we can see that this list is dominated by central governing figures – hubs – and that the exceptions are the diplomats. While special ambassadors were often sent in for specific negotiations, the role of the resident ambassador was to keep open channels of two-way communication between foreign states and sovereign entities. In both cases, however, the aim of the role was to overcome ‘structural holes’: not only geographic distance, but also lacunae in information. Diplomats, then, are bridges.

The question is: how can we discover more bridges when the hubs dominate the top-ranked positions for betweenness? We already know that hubs have a high degree, and that, as a result of that high degree, hubs are likely to score highly for betweenness. We want to find the people with relatively few correspondents (that is, with low degree), who nevertheless score highly for betweenness, because such people are likely to be bridges. One way to quickly detect such people is by comparing these two measures in a graphical representation, as in Figures 3 and 4, and looking for outliers. In these two graphs the horizontal axis represents the degree, meaning that the further the point is from the intersection of the axes, the higher the degree. The vertical axis, in contrast, represents their betweenness *ranking* – the ranked place given to a node if you order them all from highest betweenness score to lowest. We use the ranking because the betweenness score is expressed as a fractional value, which is more difficult to interpret than a rank. In this case a value of ‘1’ on the vertical axis corresponds to the individual with the highest betweenness. The data points on the graph are varied in their opacity to signal the number of nodes with that score: the darker, the higher the concentration of people. The diagonal arrangement of points in the two graphs shows a direct correlation between degree and betweenness. In other words, the more correspondents a given person has, the more likely they are to have shortest paths going through them in this correspondence network. The point is made by looking at the two extremes of the graph in Figure 3, which depicts the 5,040 correspondents writing in the 1580s. The left-most and top-most point on the graph represents 3,378 people (67%) who corresponded with only one person, and have zero shortest paths going through them (so a betweenness score of zero). Sir Francis Walsingham can be found at the opposite extreme on the graph, the bottom-most and right-

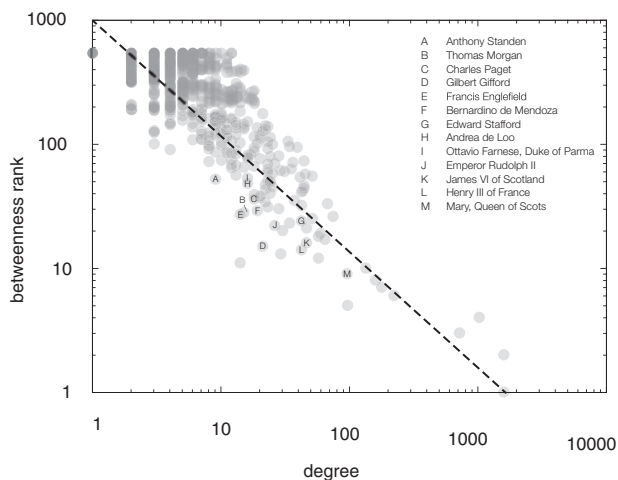


Fig. 3. This graph charts degree against the betweenness ranking of each person who corresponded in the 1580s. The data points on the graph are varied in their opacity so as to signal the number of nodes with that score: the darker, the higher the concentration of people.

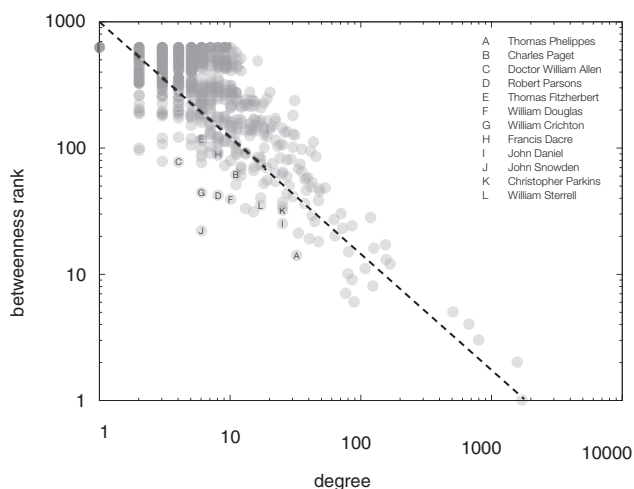


Fig. 4. This graph charts degree against the betweenness ranking of each person who corresponded in the 1590s.

most point with a total degree of 1,579, and with the highest betweenness of any correspondent in this decade. The reason for Walsingham's high number of correspondents is easily understood: he was one of Elizabeth I's principal secretaries (1573–90), well known for his use of spy networks, making him a key information hub during that decade.

Between the extremes of Walsingham and the people with no shortest paths going through them we have drawn a trend line that shows the general correlation between degree and betweenness ranking. The trend line is useful because it enables us to see deviations. Those significantly above it are less infrastructurally significant, while those below it have a higher bridging function than we might expect from the number of people with whom they corresponded. The further away we move from the origin (the intersection of the axes), the more the deviation matters because of the logarithmic scale. Focusing on outliers below the trend line, we begin to see an interesting pattern. Sir Anthony Standen (A in Fig. 3), Thomas Morgan (B), Charles Paget (C), and Gilbert Gifford (D) were Catholic exiles and agents working for Mary, Queen of Scots, who was famously implicated in a plot against Elizabeth I, which led to her execution. Gifford and Standen have also been identified as double agents in the employ of Walsingham. Sir Francis Englefield (E) was another Catholic exile who became a close confidant of Cardinal William Allen, leader of the English Catholic exiles, and the Jesuit Robert Persons, who advocated forcible intervention by Spain in English affairs, and the succession of the Infanta. Bernardino de Mendoza (F) was Phillip II of Spain's ambassador in London, but was expelled from England in 1584 after his involvement in Francis Throckmorton's plot against Elizabeth I was revealed. What unites these men is their Catholic faith, their location in countries outside their own, and their trade in intelligence as either conspirators or double agents.

This pattern below the trend line is not just a peculiarity of the 1580s: we see it again in the graph for the 1590s (Fig. 4). In a similar region of the graph we find a cluster of people known for conspiracy and espionage. Thomas Phelippes (A in Figure 4) was a famous cryptographer and intelligence-gatherer who worked closely with Walsingham. Other figures include Paget (B), Allen (C) and Persons (D), already mentioned above. Thomas Fitzherbert (E) was a member of the Allen party within the Catholic exile community and in the service of the Duke of Feria; he later became Philip II's English secretary. William Douglas (F), the tenth earl of Angus, and William Crichton (G) were both Scottish Catholics involved in the 'Spanish blanks' plot of 1592, which sought to further the Counter-reformation in the British Isles. Francis Dacre (H) associated with various enemies of the Tudor state: he entered the service of James VI of Scotland, and was a pensioner of Spain, receiving a monthly allowance of 80 crowns from Philip II, who made him a captain of musketeers. John Daniel (I) might be described as an aspirant-intelligencer, who after a period in exile approached Justice Young with reports of a plot to blow up the Tower.¹⁷ Other people who have high betweenness despite relatively low degree include Catholic exiles who became agents for the Tudor government, like John Snowden (J), Christopher Parkins (K) and William Sterrell (L). Parkins was a Catholic exile who was subsequently recruited as an agent by Walsingham and thereafter employed as a diplomatic agent to Denmark,

Poland, Holy Roman Emperor Rudolf II, and the Hanseatic League. Sterrell is a more complex figure with numerous aliases: Stephen Alford has suggested he was a spy working under the direction of Phelippes, whereas Patrick H. Martin and John M. Finnis have suggested he was in fact a triple agent.¹⁸ Snowden is an outlier on this graph, with a strikingly high betweenness ranking of twenty-two (only twenty-one people have more shortest paths going through them during this decade), despite corresponding with only six separate people. His case helps us to see why this particular combination of network attributes is shared by conspirators and double agents.

THE OTHER SNOWDEN

Like his modern namesake Edward, John Snowden might be described as whistleblower. But whether this Catholic exile gave up information to the English crown because of a personal scruple, or simply to save his life after being captured by the Tudor authorities, has not been established. What we do know about him, however, suggests that he provides a model for understanding the status of Elizabethan double agents. Snowden was, in fact, the alias of John Cecil, an English Catholic educated at Trinity College, Oxford. Following his graduation in 1583, he attended the English Catholic seminary at Rheims, before moving to the English College at Rome, where he was ordained priest and spent much of 1587–8 as the Latin secretary to Cardinal Allen. Thereafter he travelled to Spain to assist at the seminary of Valladolid, where Father Persons was the director. Allen and Persons were two of the most prominent enemies of the Elizabethan state: they were at the centre of activities to try and restore Catholicism in England, which began with sending missionary priests to England in the 1570s and continued with a tactical alliance with Philip II King of Spain, who sought to change the faith of England by forcibly removing Elizabeth through plots such as the Throckmorton plot and the 1588 Armada.¹⁹ Under Persons's direction in 1591 Snowden and one John Fixer were sent as spies to England, but their ship, the *Adulphe*, was intercepted and they were imprisoned and interrogated by William Cecil, Lord Burghley. As Catholic priests, they stood to be tried for treason, but in order to avert this outcome Snowden provided Burghley with information on the plans of Persons and Allen, and offered himself as a secret agent for the crown.

In order to win Burghley's trust Snowden set down in a letter the principal points of the Spanish practices against England, and his own intended services, to show that 'hyt is not so impossible as it is comonly taken to be a good subiecte and a good catholique'.²⁰ In this striking statement Snowden claims that it is possible to hold two identities in tension: that he could at once be a good subject to his Protestant queen, and also practise his Catholic faith. The affirmation of his dedication to two communities that were viewed by the Tudor state as mutually exclusive, expresses a metaphorical concept of betweenness: he could be a member of both, he claimed, without

troubling his dedication to either. Such a proposal sounds like opportunism; however, in the same letter Snowden suggests that the offer arises from a genuine scruple about the plots against his monarch. He claims that he had always been ‘mynded’ to ‘haue geuen faythefull information’ to Burghley regarding the plots of the Catholics once he had safely arrived in Amsterdam, and proposed:

to persuade all men [i.e. Catholics] from favouring foraine invasions, from practices of treason against her Majestie from exasperating the superior authoritie or such like violente proceedinges, but to suffer with humility the crosse that God doth laye. To make hit evydenste to al Catholiques both at home and abroad that nether the King of Spayne meaneth them any good by his invasions pretendes[,] nor the Cardinal [Allen] or Par[sons, i.e. Persons] have eny respecte or remorse of the poor Catholique afflictions at home[.]²¹

In this passage he asserts not only his opposition to acts of treason, but also his concerns about the motivations of Philip II and the leaders, Allen and Persons. Thereby he both reasserts his status as a ‘good subjecte’ and redefines what it means to be a good Catholic: he must not raise arms, but rather suffer his ‘crosse’. Based on this reassessment, Snowden proposes that by granting religious freedom Burghley could win over current opponents to the English cause. In exchange for liberty of conscience, he suggests, exiled Catholics like him would readily take an oath neither to hear nor suffer any practice of treason, or deal in matters of State.

Snowden’s proposal was audacious. As Stephen Alford points out, such a claim ‘went against the grain of over thirty years of Elizabethan thinking. Burghley had long viewed all the English priests trained in the seminaries of France, Italy, and Spain as conspirators and traitors’.²² But Burghley clearly took it seriously. After recovery of Snowden’s papers from the Adulphe, which confirmed many of his claims, he and his companion were recruited as double agents on 3 May 1591. To be effective in this mission Snowden was required to seem like a loyal member of the Catholic exile community whilst gathering intelligence for the English crown. Communicating with such resolutely opposed communities is what gives Snowden, and others like him, such high betweenness. Although he corresponded with relatively few people, the record of his correspondence in the State Papers archive places him between key leaders in these respective communities, as can be seen in Figure 5. In fact, this diagram shows just how similar Snowden’s community looks to the toy example of the ‘bridge’ in Figure 2.

Snowden had a clear idea of how information should flow between these two communities, making him not only a bridge but also a gatekeeper of information.²³ In one direction, towards the Tudor government, he sent details of Spain’s military capabilities, names of Allen’s principal agents in Rouen, Paris, Madrid and Flanders, a long list of priests who opposed

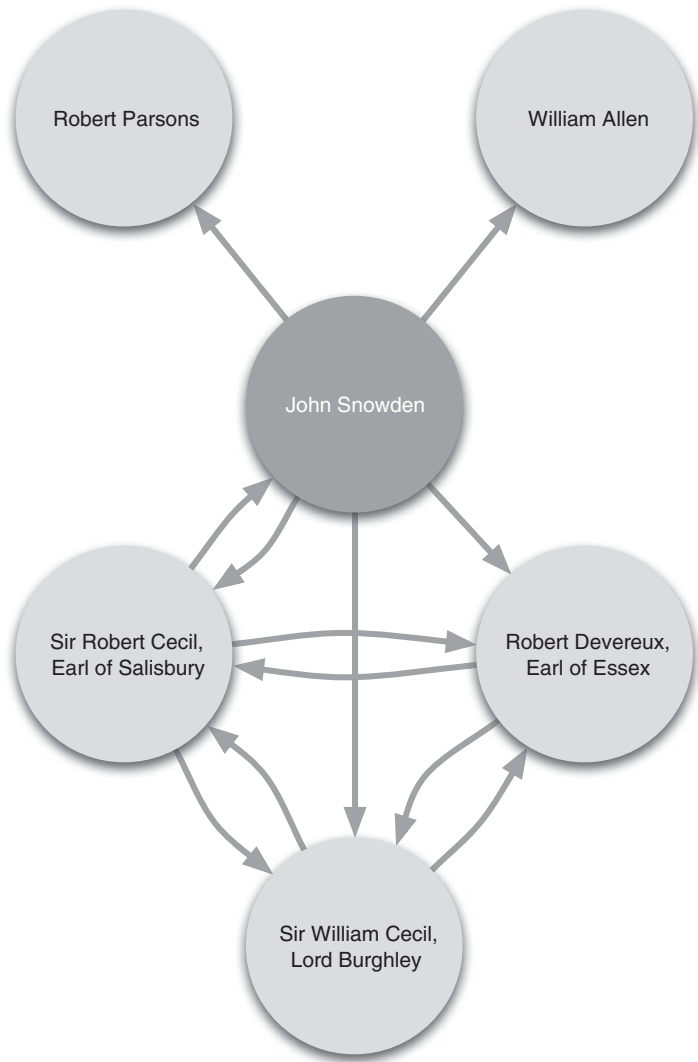


Fig. 5. John Snowden’s letter network.

Allen’s aggressive policy and Elizabethan subjects who favoured the Spanish. In the other direction Snowden promised to send information to exiled Catholics about the situation of English Catholics, challenging the perception amongst that community that in England there is ‘practysed vppon Catholics a most rigorous persecution’ such that ‘the oulde prisons wyll not houlde them but new ar buylded[; and] that the tormentes they suffer ar infinite and the maner of theyre deathes intollerable’. He proposed to inform them that he found there, instead, ‘greate peace, tollerance,

tranquillitye, and moderation with wonderfull clemency and mercy, and almost a general libertye and dimission of all'.²⁴ What is interesting about this passage is that Snowden does not suggest circulating misinformation within the exile community, as we might expect in a programme of espionage; rather, he states the power of truth. This may simply be to appease his conscience. But this approach forcefully communicates the betweenness both of Snowden's identity and his strategy: he can at once correct the misapprehensions of his fellow believers, and also arouse doubt and dissent amongst the treasonous opponents of the Tudor state.

By what means he spread this information (if indeed he did) remains unclear as none of the letters sent to Catholics that survive in the State Papers archive contain such sentiments. It is more likely, however, that he transmitted it through other means. Snowden had a clear understanding of the best way to move information between countries and discrete communities: by circuitous, and thus less traceable, routes. In his early letters to Cecil he proposes the safest paths of information across the network. In a letter dated 3 July 1591 he suggests the 'most compendious and comodious' way of informing the King of Spain of 'the false foundations and wayne hopes' he receives of the state of things in England is 'by the Duke of Savoye, and he have hyt from Morgan and those he favoreth in Flaunders, and they wyll easely be brought to performe any suche office'.²⁵ In another letter on 7 July he proposes similar routes to get information to the Pope's Nuncio in Madrid (via the Bishop of Cassiano and 'D. Luys'), to the Duchess of Feria (via Lord Hungerford), and, by her son, to Philip II.²⁶ This strategy shows how Snowden's value to Burghley and Cecil derived not only from his access to intelligence within the exile community, but also from his masterful understanding of how intelligence should be moved. In these examples we can see that although Snowden provides the most direct path across the network (the source of his high betweenness), he understood that a longer, meandering path provided him with a valuable form of protection.

Snowden's full story is a fascinating one, which, unfortunately, there is not space here to tell. Nevertheless, this brief overview of his activities, and his description of how he proposed to manage his contacts, is enough to show the kind of network significance that Snowden wielded: he was able to make short paths across the network, to channel news between communities, but to keep those communities successfully separate. These are the attributes that made Snowden such an effective double agent, and they are ones that can be measured: they result in low degree and a very high betweenness. In other words, this is the network profile of someone trading in intelligence, and it is why in Figure 4 we find him in the company of so many others involved in conspiracy, espionage, and counter-espionage.

THE POWER OF PREDICTION

If there is a network profile for those who traded in secret information, then a predictive model can be developed. We constructed such a predictive approach by creating a network ‘fingerprint’ for each correspondent in the dataset using eight network measures, to see if those with the most similar network attributes amongst the 20,656 people in the archive might also have been involved in conspiracy or espionage. This approach has clear parallels with the NSA’s use of metadata analysis to identify citizens or foreign targets that might merit closer scrutiny in the form of wire-taps and other systematic surveillance.

We chose to use eight network measures because a larger range of measurements gives us a higher-dimensional space, which in turn allows us to distinguish a greater variety of potential network roles. For each correspondent we measured their in-degree (total number of people writing to a given individual), out-degree (total number of people to whom a given individual wrote), degree (the total of the previous two measures), in-strength (an individual’s total incoming letters), out-strength (an individual’s total outgoing letters), strength (the total of the previous two measures), betweenness and ‘eigenvector centrality’. Eigenvector centrality in this context may be described crudely as a measure of a given node’s (person’s) proximity to power. As Stephen Borgatti puts it: ‘even if a node influences just one other node, who subsequently influences many other nodes ... then the first node in that chain is highly influential’.²⁷ Hubs score highly for eigenvector centrality, but so do nodes with few connections that are ‘well placed’. Because of the nature of this archive, which is constructed from the perspective of the principal secretary, an edge shared with a monarch or principal secretary will ensure a high eigenvector centrality score. We can therefore see eigenvector centrality as a rough indicator of distance from central government. Taken together, these eight measurements provide a unique fingerprint for every person in the network. It is then possible, beginning with one person of interest, to find individuals with similar values for those measures by calculating a mathematically-derived similarity score, and ranking individuals according to it (from most similar, to least).²⁸

This predictive approach was tested with the spies and conspirators who appear in Figures 3 and 4 and was found to be remarkably successful in finding other people with a similar role in the network within the top fifteen to twenty search results. For example, the fifteen people most similar to Cardinal Allen writing in the Elizabethan reign are (in rank order):

- 1) William Douglas, Earl of Angus, 2) Robert Persons, 3) Anthonio de Guarás, 4) Sir Gelli Meyrick, 5) Francis Dacre, 6) Anthony Babington, 7) Juan de Idiáquez, 8) Mathew Carnsew, 9) Pedro de Zubiaur, 10) Hugh Owen, 11) Thomas Paget, Lord Paget, 12) Sir William Catesby, 13) Gilbert Curll, 14) Charles of Lorraine, Duke of Mayenne, 15) Francesco de Mendoza.

We have already encountered Douglas, Parsons and Dacre. Guarás was a Spanish agent accused by the English spy William Herle of exchanging intelligence with rebels, Catholics and other discontented parties.²⁹ Meyrick was a Welsh conspirator in Essex's rebellion, and executed for his part in it. An English Catholic, Babington gave his name to a plot against Elizabeth for which he was executed. Idiáquez was one of Philip II's chief advisors, and oversaw English and Irish affairs. Zubiaur was a soldier who won several victories over the English for Philip II. Owen and Paget (Charles's brother) were Catholic exiles involved in various conspiracies against the throne. Catesby was a Jesuit who was arrested for having harboured the Catholic martyr Edmund Campion. Curll was Mary Queen of Scots's secretary and undoubtedly complicit in her secret correspondence with the key conspirators who plotted to place her on the throne. Mayenne was Leader of the Catholic Alliance in Europe. And Mendoza, after serving Phillip II, entered the diplomatic and military service of Archduke Albert of Austria, governor of the Spanish Netherlands. These men, then, fall into two main categories: eight of them are Catholic conspirators from the British Isles, and five are continental Catholics, of whom four are Spanish and figures of diplomatic and military leadership. What unites them is that all thirteen were perceived to present foreign threats to England's security, and most of their correspondence entered the archive through interception. Meyrick might also be loosely grouped with them as a conspirator. Only Carnsew fails to conform to this pattern. Out of the highest fifteen matches, then, the similarity score succeeded in fourteen cases (93%).

With other conspirators the similarity score does almost as well. For Persons, eleven out of the top fifteen are fellow conspirators if we examine the whole Elizabethan period (73%). The profile of Sterrell returns results for ten conspirators and foreign leaders who plotted at different times with the Spanish among the top fifteen (67%). For others this method becomes more accurate if we limit the time window to a decade: Englefield and Morgan results include six conspirators of the top ten in the 1580s (60%); and Crichton's include seven spies or conspirators out of ten in the 1590s (70%). The low score for eigenvector centrality – a measure that, as discussed above, we can use to think about proximity to, or distance from, central government – is a distinctive aspect of this network fingerprint. Government interception essentially brings into the network communication lines that might otherwise be separate from official government intelligence networks. Inclusion of these letters might boost the degree of individuals by adding data, and it may uncover connections between sub-communities in the network that might otherwise have been hidden. However, the centres of those communities are still, necessarily, separated from the crown, thus keeping the eigenvector centrality score idiosyncratically low.

The similarity score does not only find patterns of conspiracy and interception, however. Its use is more general, helping us to understand the commonalities in network properties within and between particular

groups of people. In this way we can, for example, find clusters of people sharing particular correspondence behaviours that correlate with certain activities. For example, if we look at Tommaso Spinelli – one of England's earliest resident ambassadors, serving at the court of Margaret of Austria – the twenty most similar individuals seem at first to be heterogeneous in their official roles: statesmen, diplomats, churchmen. On closer scrutiny, however, we discover that all but one of them served on diplomatic missions during the reign of Henry VIII (95%).³⁰ Similarly, we can use the method to highlight new categories of people that lack a clear label. One example of this can be seen with the people most similar to Pietro Bizzarri, a fascinating figure who came to our attention in the first instance because of his unusual network profile: despite a very high number of outgoing letters (in the top 0.58%) he has zero eigenvector centrality, because he only receives letters from one individual who does not write to anyone else. The significance of the similarity list he produces is best understood, however, with a brief overview of his biography.

Bizzarri was an Italian convert to Protestantism who first came to England as a fellow of St John's College at Cambridge, and subsequently sought a career as courtier-poet in the courts of Edward VI, Mary Stuart and Elizabeth I, before requesting permission from Burghley in 1564 to travel to Venice with the continuation of his pension.³¹ Bizzarri made a compelling case for his payment to continue: 'Et auidem existimo nullum esse in europa locum qui sit magis aptus et opportunus rebus cognoscendis, quam uero istud omnibus Principibus ac Regibus perutile sit, quotidiana experientia testatur' (because I think there is no place better or more apt in Europe to gain knowledge of affairs that are beneficial to kings and princes, as everyday experience shows us).³² In return for permission to travel, then, Bizzarri offered himself as an 'intelligencer' (an early modern term used to describe people who furnished information on an *ad hoc* basis, or people we might now categorize as spies). It was an offer Burghley readily accepted, having no diplomatic presence in Venice at that time, and so began Bizzarri's lifelong career as an intelligencer, passing political and diplomatic information to the Tudor government.

We find the fifteen people most similar to Bizzarri in the Elizabethan period are:

- 1) Edmund Palmer, 2) Thomas Stokes, 3) Roger Aston, 4) William Stallenge, 5) Sir Ferdinando Gorges, 6) Sir William Browne, 7) Georges Fremyn, 8) Thomas Copley, 9) William Lyly, 10) Captain Cockburn, 11) Sir Dennis O'Roughan, 12) John Lee, 13) Sir Edmund Uvedale, 14) Edward Burnham, 15) Edward Grimeston.

From this list emerge two small groups: military leaders (6, 7, 10, 13), and men in the service of diplomats (3, 9, 14, 15); beyond that, the list appears heterogeneous, including agents, priests, and merchants. As we look

carefully at their letters, however, it emerges that all but one (Copley) of these fifteen men are involved in reporting to the government. Palmer, described in the *Calendar of State Papers* as a priest, was a double agent located mainly in Saint-Jean-de-Luz, a port on the French-Spanish border, and sent intelligence reports on the Spanish to Walsingham (nine letters), Burghley (eighteen) and his son Robert Cecil (twenty-four). Stokes was an English merchant at Bruges who sent 111 letters to Walsingham on military developments in the Low Countries. Stallenge was a merchant involved with the new Plymouth haven, where he collected intelligence from incoming ships from Brittany, Spain and Portugal, and sent it, in 188 dispatches, to Robert Cecil. O'Roughan was a double-agent Catholic priest, first used by the Lord Deputy of Ireland Sir John Perrott as a priest-catcher, and then by Perrott's successor, Sir William Fitzwilliam, to accuse Perrott of being a crypto-papist Spanish spy, on which topic he wrote thirty-six letters to Burghley and three to Elizabeth I. And Lee was an agent in Antwerp, posted there perhaps first in a military capacity, who sent reports to Burghley (thirty-eight letters) and his son Cecil (twelve) on the activities of the English exiles. Taken together we see a shared profile: like the spies and conspirators they have low degree measures and low eigenvector centrality; but unlike them they have high strength, especially out-strength (total number of letters sent) due to the high volume of intelligence they sent to the queen's principal secretary.

What is also notable about this list of fourteen intelligencers is that only five of them have any kind of biographical record, either in the *Oxford Dictionary of National Biography* or *The History of Parliament*, or even in Wikipedia: Aston, Stallenge, Gorges, Lee and Uvedale. Lacking biographical entries does not, of course, mean that figures are unknown to historians, but in addition these men are almost absent from scholarship apart from the occasional passing reference (normally following the formulation: 'X sent a letter to Walsingham/Burghley/Cecil with the information that...').³³ The focus is on the events reported by these men, rather than on the men themselves and their intelligence roles. However, while individually these men may not be deemed worthy of their own histories, we would contend that considered as a group they are. By using the similarity score we are encouraged to understand the commonalities between those men, and how they constitute a category of extra-diplomatic intelligencers.

A call to write a collective history of this extra-diplomatic category, if answered, would allow us to extend, and problematize, the existing scholarship on early modern diplomacy.³⁴ Individuals sent on official embassy could offer only so much intelligence coverage: they had a set geographic posting, and certain social networks remained closed to them as formal representatives of a foreign power. As a result scholars including William Sherman, Lisa Jardine, Stephen Alford and Elizabeth Williamson have gestured to the ways in which secretaries of state sought to diversify the sources of their information – from educational travellers (or 'scholarly readers'), to

merchants, turncoats and opportunists – and the problems this has created for labelling the men who served the intelligence-arm of the government.³⁵ Alford writes: ‘there was no kind of distinction between the gathering of intelligence at home or abroad or between intelligence, security work and counter-espionage. The secret reports on Walsingham’s desk in the 1580s could range from letters by diplomats to the reports of informants working close to English Catholic families’.³⁶ By using similarity measures, however, we can begin to quantify the differences between those diplomats and informants, and to see where our category of the extra-diplomatic intelligencer fits into that taxonomy.

SURVEILLANCE OF SURVEILLANCE

The discussion of taxonomies brings us to a more general, and perhaps more important point: that we might need to think more broadly about what it means for two people to be similar. Our tendency is to want to create bounded categories with labels, and ontologies, and to think about similarity in terms of those. This, however, can be counter-productive when thinking about people. Individuals can hold multiple official or unofficial roles in the network through time or simultaneously; the definition of roles may change through time and may not even be consistent at any one time; and roles may overlap. Instead, we propose, by thinking about similarity in terms of network properties rather than human-assigned categories we can begin to understand group identities in different ways, not least the shadings and slippage between those categories, thereby destabilizing them in productive ways.

One way of thinking about this slippage is to focus on the people who do not seem to fit the general pattern of people with whom they co-appear, either on the graphs charting betweenness against degree, or using the similarity measure. If we look back at Figures 3 and 4 we see that amongst the spies and conspirators that cluster below the trend line are scattered a number of foreign leaders. In Figure 3, which depicts the correspondents writing in the 1580s we see that appearing close to the Catholic conspirator Englefield are Ottavio Farnese, Duke of Parma (I), and Rudolph II, Holy Roman Emperor (J). We see other foreign leaders in the bottom right-hand of the graph, closer to the horizontal axis because of their high betweenness, who include James VI of Scotland (K), Henry III of France (L), and Mary Queen of Scots (M). All of these figures are represented in the State Papers archive because they wrote directly to Elizabeth and/or her principal secretaries; but they also sent a considerable body of correspondence to other contacts outside the natural ambit of the Tudor government. Rudolph II for instance wrote to a wide range of continental monarchs, members of the nobility and archbishops. The most likely reason we have these documents is because they were intercepted. For example, the 1580 letter from Rudolph to Edzard II, Count of East Friesland, concerning trade agreements with the English and the desire of the Merchant Adventurers to resettle in Emden,

appears in the archive in an English translation annotated by Christopher Hoddesdon.³⁷ Hoddesdon was master of the Merchant Adventurers at Hamburg, and sent regular political intelligence from Emden and Antwerp to Walsingham and Burghley.

The other foreign leaders were subject to the same scrutiny and practices of interception. It is this fact that places them alongside conspirators and spies. Foreign leaders are, of course, in no sense similar to Catholic conspirators in terms of standard labels or ontologies we might apply to them. They look similar on a graph comparing degree and betweenness centrality because of two factors. The first is their international reach: conspirators and foreign leaders both correspond with people at a geographical remove, and to whom the government may have no direct epistolary access. The second factor is not one of their own making; rather, the seizure of letters – whether by the covert and targeted interception of particular letters, or the wholesale requisitioning of a personal archive – leaves behind a particular signal in the archive. Our use of betweenness to understand the resulting infrastructure of the archive shows that the surveilled individuals provide short paths to the regions of the network extended by the interception and seizure of correspondence. It allows us to see that the structural similarity we see is only partly about the epistolary practices and social position of individuals; it is also about the making of the archive and how the process by which it comes to us in the present, with all its complex motivations and biases, necessarily shapes the evidence as we receive it.

Our predictive methods, therefore, do not only reveal something about the network fingerprint of individuals and the people most similar to them. In addition, and perhaps more importantly, they show the fingerprint left behind by government surveillance, and tell us precisely who was being targeted in any given period. If we compare betweenness and degree for each of the decades preceding 1580 (going back to Henry VIII's accession in 1509) we see a different pattern from the one observed in Figures 3 and 4. While the deviations below the trend line consistently contain foreign leaders across all decades (showing that they remain a perennial target of surveillance), we only begin seeing Catholic conspirators in this region of the graphs from the 1570s onwards. Following the Ridolfi plot and the Rising of the North, both in 1569, and the secret arrival of Catholic priests in England in the 1570s, panic about a co-ordinated Catholic conspiracy between Philip II and Mary Queen of Scots increased exponentially. Figures 3 and 4, then, act as a barometer not only of foreign threats but also of the government's fears.

By using the data-driven methods similar to those employed by modern governments and agencies, we are able to understand something about the surveillance practices of historical governments. As such it might be called surveillance of surveillance. This approach helps us to look deeper into the historical archive. The cases outlined above illustrate how quantitative methods can help scholars to navigate massive archives by suggesting

areas that may merit close reading. This is the ultimate aim of our methods. A variety of simple and more advanced network analysis measures allows us to observe broad trends, such as the power-law distribution of correspondents; to examine the way in which particular individuals bridged disparate communities; and to construct a similarity measure that helps us to detect structural similarities between people in the network. More generally the methods we describe enable researchers both to make discoveries about individual histories, and to examine their role in the broader narrative of Elizabethan foreign and domestic policy.

But aside from the application of these methods to historical scholarship, these findings also lead us to consider what we can infer about the methods of government agencies and private companies undertaking network analysis on citizens. Their methods undoubtedly go far beyond ours, resourced as they are with greater funding, labour, computational expertise and data. Nevertheless, we can easily imagine how the ‘distant reading’ methods demonstrated above provide ways of profiling citizens and foreign targets in order to predict likely security threats. The data will undoubtedly show similar power-law distributions, and members of known terrorist groups will serve as reference data to help build predictive models similar to our own. The results of such predictions, however, have different repercussions: the decision to read a historical letter has few if any ethical ramifications, whereas the act of surveillance necessarily compromises the privacy of the target. The justification for this, however, is one of national security.

The promise that such methods might help to avert future terrorist attacks means that many people have come to accept that some level of surveillance is necessary for our safety: we are comforted by the assertion that if we are good citizens we have nothing to fear. It may be for these reasons that the Investigatory Powers Bill (known as the Snoopers’ Charter Bill) introduced in the UK in 2016 was passed with little resistance.³⁸ These arguments, however, make the assumption that governments will act on such intelligence only when a threat is posed. Yet even with the best intentions political organizations and private companies may tip towards the slippery slope between a careful overseeing eye and the use of surveillance as a tool of manipulation or dictatorship. In 2018 the whistle-blower Christopher Wylie revealed that data harvested from the social-networking site Facebook was utilized by political campaigns to understand the social profiles of citizens not simply for the purposes of close-reading, but for its active counterpart, micro-targeting – the generation of individualized political messages designed to serve a particular agenda. This confirms the warning issued in 2016 by Alex Younger, the head of MI6 in the UK, that ‘the connectivity that is the heart of globalisation can be exploited by states with hostile intent to further their aims. ...The risks at stake are profound and represent a fundamental threat to our sovereignty’.³⁹ It is now abundantly clear that the gathering, analysis, and manipulation of network data are as much a challenge to national security as they are an aid.

Subsequent attempts to place checks on these activities – such as the Supreme Court case *Carpenter v. United States* in 2018, or the inquiry into what Facebook knew about Cambridge Analytica's activities – show the growing realization that we need to place the surveillance (or 'dataveillance')⁴⁰ practices of governments and companies themselves under surveillance. However, the onus should not just be left to political and judicial institutions. The scholarly community has a duty to think deeply about the ethics and practicalities of data gathering, and the question of 'where digital security ends and national security begins'.⁴¹ This is already happening in computer science and the digital humanities, but historians have an important voice to add. The history of surveillance offers vital lessons for the current moment. But it is important that we also look to the future. The more we engage with the power that big data and digital methods hold for us, the more we can help to shape the development of theoretical discourse around metadata and surveillance in ways that make sense for the humanities, and for society at large.

Ruth Ahnert is a senior lecturer in Renaissance Studies at Queen Mary University of London. Her research interests include Tudor literary history, epistolary culture, book history and digital humanities. She is the author of *The Rise of Prison Literature in the Sixteenth Century* (Cambridge University Press, 2013), and editor of *Reforming the Psalms in Tudor England* (special issue of *Renaissance Studies*, 2015). With Sebastian E. Ahnert she is writing a book called *Tudor Networks of Power*, and with Thomas S. Freeman she is editing *The Letters of the Marian Martyrs* (under contract with Oxford University Press).

Sebastian E. Ahnert is a Gatsby Career Development Fellow at the University of Cambridge. He holds a joint appointment at the Sainsbury Laboratory and Cavendish Laboratory, where he leads a research group studying complex networks and algorithmic information theory on the interface of theoretical physics and biology. He is an author of over fifty publications in a wide range of journals including *Science*, *Nature*, *Cell*, *PNAS*, *Nature Communications*, *Journal of the Royal Society Interface*, *Physical Review*, and *English Literary History*.

NOTES AND REFERENCES

The authors would like to thank the Arts and Humanities Research Council, Stanford Humanities Center, and the Royal Society for funding fellowships during which the research for this article was undertaken.

1 David Cole, 'We Kill People Based on Metadata', *New York Review of Books*, 10 May 2014: <http://www.nybooks.com/daily/2014/05/10/we-kill-people-based-metadata/>, accessed 10 Nov. 2017.

2 This contentious argument, however, was dismantled on 22 June 2018 in the case *Carpenter v. United States*, when the Supreme Court ruled that the government needs a warrant to access a person's cellphone location history. The court found in a 5 to 4 decision that

obtaining such information is a search under the Fourth Amendment and that a warrant from a judge based on probable cause is required. See https://www.supremecourt.gov/opinions/17pdf/16-402_h315.pdf, accessed 26 Oct. 2018.

3 Mike Pompeo and David B. Rivkin Jr, 'Time for a Rigorous National Debate About Surveillance', *Wall Street Journal*, 3 Jan. 2016, <https://www.wsj.com/articles/time-for-a-rigorous-national-debate-about-surveillance-1451856106>, accessed 26 Oct. 2018.

4 See <https://www.gale.com/uk/primary-sources/state-papers-online>, accessed 26 Oct. 2018.

5 For an introduction to this terminology, see the books and journals of the same names: Franco Moretti, *Distant Reading*, London, 2013; Matthew Jockers, *Macroanalysis: Digital Methods and Literary History*, Urbana, Chicago and Springfield, 2013; and *Journal of Cultural Analytics*, <http://culturalanalytics.org/>.

6 Franco Moretti, 'Conjectures on World Literature', *New Left Review* 1, 2000, pp. 54–68.

7 XML stands for Extensible Markup Language, a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.

8 The UK copyright exception and its ramifications for researchers, last modified 9 Feb. 2016: <https://www.jisc.ac.uk/guides/text-and-data-mining-copyright-exception>, accessed 26 Oct. 2018.

9 While women make up only four percent of the people in the network due to the political nature of the archive, this still represents a large number of people when you are working at the scale of 20,656 correspondents. Women's letters make their way into the archive through various means, which include (but are not limited to): acting as petitioners to the government; the seizure of personal and family archives; or using their position of proximity power, as queen's consort, perhaps, or as ladies-in-waiting to Mary or Elizabeth.

10 Julián Candia, Marta C. González, Pu Wang, Timothy Schoenharl, Greg Madey and Albert-László Barabási, 'Uncovering Individual and Collective Human Dynamics from Mobile Phone Records', *Journal of Physics A: Mathematical and Theoretical* 41: 22, 2008, p. 224,015.

11 See Duncan Watts and Steven Strogatz, 'Collective Dynamics of "Small-world" Networks', *Nature* 393, 1998, pp. 440–42; Albert-László Barabási and Reka Albert, 'Emergence of Scaling in Random Networks', *Science* 286, 1999, pp. 509–12; Albert and Barabási, 'Statistical Mechanics of Complex Networks', *Reviews of Modern Physics* 74, 2002, pp. 47–97; and Mark E. J. Newman, *Networks: an Introduction*, Oxford, 2010. For a general overview, see Barabási, *Linked: the New Science of Networks*, Cambridge MA, 2002.

12 Stephen P. Borgatti, 'Centrality and Network Flow', *Social Networks* 27, 2005, pp. 55–71 (p. 61).

13 We used Python, which has a 'library' of network analysis measures called NetworkX. For documentation on the NetworkX library, see this resource (last modified 31 Jan. 2016): <https://networkx.readthedocs.io/en/stable/>, accessed 26 Oct. 2018.

14 Barabási and Albert, 'Emergence of Scaling'.

15 Matthew Lincoln, 'Confabulation in the Humanities', <https://matthewlincoln.net/2015/03/21/confabulation-in-the-humanities.html>, accessed 26 Oct. 2018.

16 Ronald S. Burt, *Structural Holes: the Social Structure of Competition*, Cambridge MA, 1992.

17 Stephen Alford, *The Watchers: a Secret History of the Reign of Elizabeth I*, London, 2012, pp. 306–8.

18 See Alford, *Watchers*, pp. 285–94; Patrick Martin and John Finnis, 'The Secret Sharers: "Anthony Rivers" and the Appellant Controversy, 1601–2', *Huntington Library Quarterly* 69: 2, 2006, pp. 195–238; Patrick Martin, *Elizabethan Espionage: Plotters and Spies in the Struggle between Catholicism and the Crown*, Jefferson NC, 2016.

19 For an overview of their activities, see Thomas H. Clancy, *Papist Pamphleteers: the Allen-Persons Party and the Political Thought of the Counter-Reformation in England, 1572–1615*, Chicago, 1964; Francis Edwards, *Robert Persons: the Biography of an Elizabethan Jesuit, 1546–1610*, St Louis Missouri, 1995; Victor Houlston, *Catholic Resistance in Elizabethan England: Robert Persons's Jesuit Polemic, 1580–1610*, Aldershot, 2007; and Eamon Duffy, 'William, Cardinal Allen, 1532–1594', *Recusant History* 22, 1995, pp. 265–90.

20 Snowden to Burghley, 23 May 1591: State Papers (henceforth SP) 12/238 f.257r.

21 SP 12/238 f.257r–v.

22 Alford, *Watchers*, p. 276.

23 A formal definition of a 'gatekeeper' in social-network analysis is given in Roger V. Gould and Roberto M. Fernandez, 'Structures of Mediation: a Formal Approach to Brokerage in Transaction Networks', *Sociological Methodology* 19, 1989, pp. 89–126.

24 SP 12/239 f.59r.

25 SP 12/239 f.102v.

26 SP 12/239 f.115r.

27 Borgatti, 'Centrality and Network Flow', p. 61.

28 Such a score can be calculated in several ways. Our approach was to use the Euclidean distance between the eight-dimensional points defined by the logarithms of an individual's ranks for the different measurements.

29 Rory Rapple, *Martial Power and Elizabethan Political Culture: Military Men in England and Ireland, 1558–1594*, Cambridge, 2009, pp. 102–4.

30 On Spinelli see Betty Behrens, 'The Office of the English Resident Ambassador: Its Evolution as Illustrated by the Career of Sir Thomas Spinelly, 1509–22', *Transactions of the Royal Historical Society* 16, 1933, pp. 161–95, p. 162.

31 On Bizzarri, see Kenneth R. Bartlett, 'Bizzarri, Pietro (b. 1525, d. in or after 1586)', *Oxford Dictionary of National Biography*, Oxford, 2004; online edn, January 2008: <http://www.oxforddnb.com/view/article/2487>, accessed 17 Sept. 2016; and Massimo Firpo, *Pietro Bizzarri, esule italiano del cinquecento*, Torino, 1971.

32 SP 15/12 f.23r.

33 See for example the reference to Stokes in Mack P. Holt, *The Duke of Anjou and the Politique Struggle during the Wars of Religion*, Cambridge, 1986, p. 199; to Fremyn in Peter C. Mancell, *Hakluyt's Promise: an Elizabethan's Obsession for an English America*, New Haven, 2010, p. 62; or to Grimestone in John Nichols's *The Progresses and Public Processions of Queen Elizabeth: Volume III: 1579 to 1595*, ed. Elizabeth Goldring, Jayne Elisabeth Archer, and Elizabeth Clarke, Oxford, 2013, p. 546, n. 94.

34 See for example, *Practices of Diplomacy in the Early Modern World c.1410–1800*, ed. Tracey A. Sowerby and Jan Hennings, Routledge, 2017; Catherine Fletcher, *Diplomacy in Renaissance Rome: the Rise of the Resident Ambassador* Cambridge, 2015; *Diplomacy and Early Modern Culture*, ed. Robyn Adams and Rosanna Cox, Basingstoke, 2011; Douglas Biow, *Doctors, Ambassadors, Secretaries Humanism and Professions in Renaissance Italy*, Chicago, 2002; Daniela Frigo, *Politics and Diplomacy in Early Modern Italy: the Structure of Diplomatic Practice, 1450–1800*, Cambridge, 2002.

35 William H. Sherman, 'Research Intelligence in Early Modern England', *Studies in Intelligence* 37: 5, 1994, pp. 95–104 (p. 100); Lisa Jardine and William Sherman, 'Pragmatic Readers: Knowledge Transactions and Scholarly Services in late Elizabethan England', in *Religion, Culture and Society in Early Modern Britain: Essays in honour of Patrick Collinson*, ed. Anthony Fletcher and Peter Roberts, Cambridge, 1994, pp. 102–24; and Elizabeth Williamson, "'Fishing after News" and the *Ars Apodemica*: the Intelligencing Role of the Educational Traveller in the Late Sixteenth Century', in *News Networks in Early Modern Europe*, ed. Noah Moxham and Joad Raymond, Brill, 2016, pp. 542–62.

36 Stephen Alford, 'Some Elizabethan Spies in the Office of Sir Francis Walsingham', in *Diplomacy and Early Modern Culture*, ed. Adams and Cox, pp. 46–62, p. 48.

37 SP 82/1 f.61r-v.

38 For discussion see Tim Ring, 'Your Data in Their Hands: Big Data, Mass Surveillance and Privacy', *Computer Fraud and Security* 8, 2016, pp. 5–10.

39 Quoted in Commissioner Julian King's 'Speech to the National Cybersecurity Centre, In London', 2017. Full text available here: https://ec.europa.eu/commission/commissioners/2014-2019/king/announcements/commissioner-kings-speech-national-cybersecurity-centre-london_en, accessed 26 Oct. 2018.

40 José van Dijck, 'Datafication, Dataism and Dataveillance: Big Data between Scientific Paradigm and Ideology', *Surveillance & Society* 12, 2014, pp. 197–208.

41 See Michael Hack, 'The Implications of Apple's Battle with the FBI', *Network Security* 7, 2016, pp. 8–10.