

Joined-Up Early Modern Diplomacy: Linked Data from the Correspondence of Thomas Bodley  
Centre for Editing Lives and Letters  
Dr Robyn Adams, and Jaap Geraerts  
Project Report



## Project Aim

This project proposed to analyze the dataset generated by *The Diplomatic Correspondence of Thomas Bodley, 1585-97* by producing visualizations of people and geographical locations mentioned in the letters for a new project, *Joined-Up Early Modern Diplomacy: Linked Data from the Correspondence of Thomas Bodley*.

## Summary

*The Diplomatic Correspondence of Thomas Bodley 1585-97* is an online edition of the letters in English written between Bodley and his diplomatic network, completed at CELL in 2011. The bulk of these [930] letters were written during his long embassy to the Low Countries, where for nearly nine years he represented Elizabeth I in the role of English agent on the Dutch Council of State during the conflict between the United Provinces and Spain (1588 – 1597). Bodley was positioned at the centre of a correspondence network which included his political masters back home in England, the men responsible for the military activities of Elizabeth's troops, and other English personnel affected by the conflict, such as the Merchant Adventurers based in Middelburg.

The contents of the letters feature a wide range of information types, spanning military movements, political events, dynastic marriage negotiations, individuals' petitions, secret intercepted intelligence and ongoing patronage strategies between Bodley and his superiors. The letters reveal the multiple roles Bodley was required to perform, from standing firm in difficult negotiations on behalf of the queen with the Council of State, to forwarding petitions from supplicants based on the continent to prominent figures in England. As such, his correspondents represent a wide range of social hierarchy, from European royalty (Elizabeth I), through the English nobles heading up different aspects of the Low Countries campaign (Lord Treasurer Sir William Cecil, Lord Admiral Charles Howard), to individuals seeking information, restitution, repatriation or assistance (Captain Oliver Lambert) or making petitory requests (Richard Saltenstall).

## New Users, Mining our own Data

During the transcription and encoding period of the *Diplomatic Correspondence*, the research team made the decision to enrich the metadata of the project by encoding each mention of the people and geographical locations featured in the letters.<sup>1</sup> While the network of correspondents is relatively small, the nature of the correspondence – describing events unfolding in Western Europe over nearly a decade - means that the tagged references to people and geographical locations are numerous and form a generous dataset. We have behaved as new users, and have mined this dataset to produce the visualizations for *Joined-Up Early Modern Diplomacy*.

Our dataset derived from the Bodley correspondence is ideal for combining historical research and data visualization. As Ruth and Sebastian Ahnert note, 'Letters offer themselves to network visualization and analysis in a much more straightforward way than other forms of literature'.<sup>2</sup> An epistolary network rendered in visual terms depicts both the relationship between correspondents and the physical journey of the letters themselves, creating a 'material link' or trace between the nodes. A central feature of the project was to

<sup>1</sup> As Lorna Hughes states, 'the key tool in the resource discovery of a digital dataset, or indeed any dataset, is the concept of metadata, or data describing data. Metadata is used to describe a record or an archived resource in such a way, and using such descriptors, that a researcher can easily discover that it contains information relevant to their researches.' 'Resource Discovery and Curation of Complex and Interactive Digital Datasets' in Chris Bailey and Hazel Gardiner eds. *Revisualizing Visual Culture* (Farnham: Ashgate, 2010), p.48.

<sup>2</sup> Ruth and Sebastian Ahnert, 'Protestant Letter Networks in the Reign of Mary I: a Quantitative Approach', *ELH* (forthcoming).

interrogate the existing data in such a way that they might produce glimpses of patterns and behaviours by Bodley and his correspondents which were previously difficult to detect when tabulated textually.<sup>3</sup>

## Report

This project falls within an interesting moment, digitally speaking. In the months leading up to our proposal to the Open Humanities Awards (March 2013), scholarly networks were buzzing with ideas and new examples of network visualization, infographics and data visualization. The CELL project team had been interested in novel methods of digital representation which stretched traditional scholarly boundaries, and we were keen to try something new with the data we already possessed (which was in the public domain available on Github). The metadata embedded in the *Diplomatic Correspondence* appeared to be ideal material with which to test how data visualization fits into the academic skillset alongside other high-end technical skills such as palaeographical expertise. Data visualization has proved a popular tool for scholars looking for an alternative method of assessing and analyzing data, networks and groups, and network theory has gained a firm foothold in digital humanities projects. We were keen to test various features of the technique, to explore a) what new information it could tell us about a dataset we were already familiar with, and b) the positive benefits or caveats of such an approach. To emphasize, we weren't interested only in generating visualizations from the data: we wanted to assess how useful were both the process and the results of those visualizations. As the *Tooling Up* team at Stanford have noted, 'We may tend to think of visualization as a finished product, as part of presentation, but it may be more useful to think of visualization as part of the research process'.<sup>4</sup>

After hiring a research assistant, Jaap Geraerts, who had both technical experience and knowledge of the historical context, we set about preparing the data for visualization. The first part of the process consisted of rationalizing the data between the two databases (a Microsoft Access and a MySQL database) in which the project data is stored. The MySQL database contains the data which can be viewed on the project website (the 'online edition' as it is defined), whereas the Access database includes a larger body of material, namely deriving from a preliminary census of all letters in which Bodley was mentioned, and letters written in foreign languages dating from the period in which Bodley served as an ambassador in The Hague. Because both databases have been used for the visualizations, it was important to make sure that the data stored in both databases was identical. The differences between the data in the databases were mostly detected via queries, after which the databases were manually updated. These modifications were also applied to the relevant XML-files, thus ensuring that the MySQL-database and the website were completely in sync with one another.

In order to optimize our dataset for the creation of visualizations, where there was an unknown author or recipient, we inferred these values, thus augmenting the existing data and expanding the possibilities for analysis. At a later stage in the project specific information was added to the existing data as well, (such as the country in which the places mentioned were located), as a result of which the correspondence between specific people could be analysed in terms of their national and geographical location. In a similar fashion the whole dataset could be examined (e.g. which countries figured most prominently in the correspondence).

## Historical Data and Contemporary Software

When the process of rationalizing the data was nearing its end, we started to familiarize ourselves with the software we had selected to create the visualizations: GEPHI.<sup>5</sup> GEPHI is open source software which is mainly used for network analysis and the creation of network visualizations. The main advantages of using GEPHI, besides its being open source, is that the software is regularly updated - often based on the demands of its users - and that it is easy to use. Before we could start working with the Bodley dataset, however, we had

3 See Dan Dixon, who comments on methods of seeking patterns which 'would not be readily apparent to a human reader and require the brute force, or transformation, that computational methods bring which are usually difficult, boring, or physically improbable for human researchers to carry out', in 'Analysis Tool or Research Methodology: Is there an Epistemology for Patterns?' in David M Berry, ed. *Understanding Digital Humanities* (Basingstoke: Palgrave Macmillan, 2012), p.191.

4 'Tooling Up for Digital Humanities', Stanford University, <[http://toolingup.stanford.edu/?page\\_id=1255](http://toolingup.stanford.edu/?page_id=1255)>, accessed 21/03/14.

5 We worked with version 0.8.2.

to consider and assess our data - derived from historical sources - in relation to the capacity of the software platform to deal with it. This was probably one of the most interesting and informative parts of the whole project. For Gephi is a standardized piece of software (as are most of the tools normally used in Digital Humanities projects), which on the one hand has the advantage that it can be used to work with various sorts of data fairly easily - people have set out to import various sorts of data ranging from social connections generated by Facebook to the epistolary network centered around the Roman lawyer Pliny the Younger.<sup>6</sup> However, the fact that Gephi is generic software contains inherent drawbacks as well; one being that the historical data is often 'messier' than a neatly programmed piece of twenty-first-century software, as a result of which the complexity of the historical sources cannot always be fully captured by the software. Because of this, questions often need to be posed to the software in different ways than the traditional methods in which scholars have hitherto been trained.

The following examples will illustrate this point. Some of the letters which comprise part of the dataset were authored by more than one person or were sent to more than one recipient (there were also letters with several authors and recipients). This is a trait common to early modern epistolary networks. This was difficult to capture in Gephi, for each node represents one author or recipient, while every edge, the line connecting the nodes, represents one letter. Because one edge can only connect two nodes, (at least in the current version of Gephi), the fact that some letters had more than one author or recipient could not be visualized at all, and we had to proceed as if every letter had one author and one recipient (i.e. a letter with one author and two recipients in our dataset is seen as two letters in the data imported into Gephi - one letter from the author to each recipient). Working with this off-the-shelf software thus required an additional amount of editorial intervention as decisions about the way we treated and proceeded with our data had to be made before we could proceed with creating the visualizations.

Another characteristic of the historical sources which has proved complicated to capture is the fact that letters in reality often were 'packets'; other documents were often enclosed with letters, e.g. copies of letters or more exotic objects such as maps. Sometimes recipients were asked to distribute portions of the material that was sent along with the letters they received to other people, and thus new links were created that existed external to the epistolary network (in the sense that these relationships were not directly forged by one person sending a letter to another person, but were transmitting information derived from within the letters included in the epistolary network). We have called these people **transmission agents**. It would have been extremely interesting to include these multifaceted and nuanced relationships in the network. However, the different features of the various links and activities of letter-writers and transmission agents are difficult to visualize using Gephi, for although attributes can be added to the edges connecting the nodes, only one edge can connect two nodes; hence only one sort of relationship can link two nodes. Because of these limited options to treat nodes and edges which have different attributes, we found it difficult to incorporate a range of relationships into one network and one visualization.

We found a cognate problem in trying to analyze a 'multimodal' network; the in-house Gephi algorithms do not distinguish between various types of edges - although one can work around this by filtering a network (reducing the network to one type of node/edge, i.e. one type of relationship), then running the statistical analysis over this filtered network, and comparing the results. One could circumvent these limitations by constructing separate networks, each of them based on a specific sort of relationship, and to compare the analysis of these separate networks. The limitations of the available software made us realize that the complexity of a dataset often cannot be captured in one visualization, for the specific characteristics of the visualization (or of the software used to create it) might not allow for the flexibility necessary to incorporate all aspects of the (historical) sources. In addition, there is often simply too much information, and including this would create an incredibly dense image, at once reducing the added value of visualizations (i.e. quickly discerning significant patterns). We therefore set out to think about which visualizations are best suited for depicting certain aspects of the dataset and the way in which these visualizations complement each other.

6 E.g., Sarah Joy, 'Using Netvizz & Gephi to Analyze a Facebook Network', <<https://persuasionradio.wordpress.com/2010/05/06/using-netvizz-gephi-to-analyze-a-facebook-network/>> and 'Visualizing Historical Networks: Pliny Letters', Harvard University, <<http://www.fas.harvard.edu/~histecon/visualizing/pliny/index.html>>, both accessed 21-3-2014.

This 'tension' between our historical data and our selected software made it necessary to modify the CSV-files that are used to import data into GEPHI. The files are basically lists of nodes and edges, the latter list consisting of a 'source' node and a 'target' node, and because of the discrepancies between our data and the data-format used by GEPHI, we needed to manually update these lists (in the case of multiple authorship or recipients) to ascertain that GEPHI processed the data in the correct way. So, besides the methodological and editorial decisions which had to be made, importing the data into this software required additional manual work. We stress this point because it is vital to account for the labour required prior to the actual creation of the visualizations. For although the IT-tools commonly used in Digital Humanities research, (i.e. data-mining, (network) analysis and the creation and output of visualizations), can add considerable value and extend scholarly research into other domains, to achieve this scholars are required not only to gather the corpus of material and data (activity which demands a host of expertise in itself) but often need to manipulate, disambiguate or modify the data before it can be processed by computer software.<sup>7</sup> It is essential that all these processes be executed in a methodologically robust way. Thus, research undertaken in Digital Humanities relies on the successful marriage of traditional research methods with a sound understanding and application of IT-technologies.

### Visualizations

All visualizations created in the project will be incorporated into the *Diplomatic Correspondence* website.

The aim of this project was speculative in the first instance, i.e. the visualizations were not intended to support existing research questions but rather to detect new patterns and frameworks for analysis. Our interest lay in the whole process of creating visualizations: of seeking the connective tissue which comprised Bodley's correspondence and to enhance our understanding of the advantages and pitfalls of working with IT-tools and visualizations. We wanted to analyze the dataset as a whole as well as the relationships between correspondents. As such, the primary impetus behind this project was not the visualizations themselves, but rather the lessons learned from thinking about visualizations and the process of relating our data to the software used to produce them. As well as the realization that additional editorial intervention was necessary at different times in the process, we gained new insights into the possibilities and limitations of data visualization by testing the boundaries of the software currently available.

The visualizations were created by using a combination of various programs such as GEPHI, Adobe Illustrator, Inkscape, and Microsoft Visio. The data on which the visualizations are based was generated by executing queries in the Access and the MySQL databases. After the results from the queries were put into the right format so that they could be imported into GEPHI, the work on the actual visualizations could finally commence.

### Constraints

One of the limits of the majority of recent network visualizations is that they can contain only one layer of information - i.e. only one sort of relationship is depicted. Users have tested the limits of the various available visualization software platforms in order to assign various types of nodes,<sup>8</sup> but so far the development of 'multimodal' networks as they are sometimes termed is still in its infancy. The limitations of networks which include only one type of relationship are obvious, for although we can depict a correspondence network, other relationships which linked correspondents to each other cannot be visualized, that is, it is difficult to capture the nuances of all these relationships in one visualization.<sup>9</sup> This project has exposed the boundaries of the visualizations currently enabled by GEPHI. However, by introducing another layer of information, namely the

7 This is not always the case: sometimes the data is available to be mined and analysed, but the specific research questions require the software to be updated or altered. See, e.g. Joris van Eijnatten et al., 'Big data for global history: the transformative promise of Digital Humanities', *Low Countries Historical Review* 128:4 (2013) 55-77.

8 See, e.g. "Visualizing Historical Networks: People and Institutions", Harvard University, <<http://www.fas.harvard.edu/~histecon/visualizing/graphing/people.html>>, accessed 21-3-2014.

9 E.g., Ruth and Sebastian Ahnert, 'Protestant Letter Networks in the Reign of Mary I: a Quantitative Approach', figure 1.

metadata of the places and people mentioned in the letters, we have managed to achieve positive results. (This was easier said than done, for the introduction of a new layer of information required the MySQL database to be modified, as well as the CSV-files used to import information into Gephi).

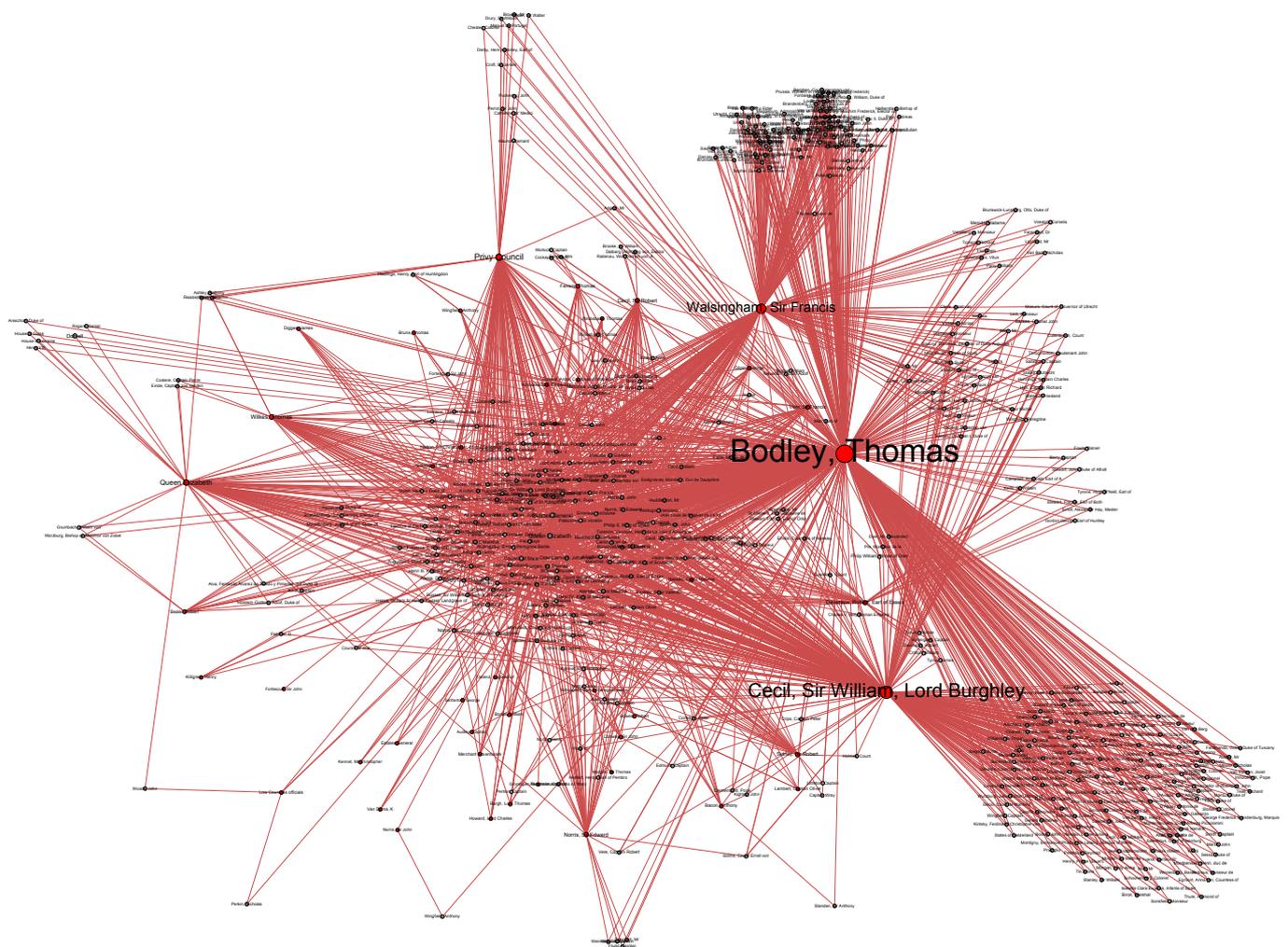


Figure 1. People: entire network 1 (hairball)

### The Visualizations in Detail

Importing the comprehensive metadata to Gephi results in an extremely dense image which has required manual modification in order to produce a visualization in which the relationships are clear. Because of the density of this image, as an overview it is not extremely useful. We could not resort to algorithms which can normally be used to improve the 'readability' of networks, (such as the Force Atlas algorithm),<sup>10</sup> because these are fine-tuned to work with straightforward one-to-one relationships. Introducing another layer of information (the people and geographical locations) creates a one-to-one-to-one relationship (between the author; the information mentioned, and the recipient, respectively). However, although at first instance the scholarly advantages of such a visualization are difficult to grasp, when zooming in and looking at the specific correspondence between two people, useful patterns start to emerge. It became immediately apparent, for instance, that there were a number of Scottish noblemen that were only mentioned in the correspondence between Bodley and Sir Robert Devereux, earl of Essex.

<sup>10</sup> 'New Tutorial: Layouts in Gephi', <<http://gephi.org/tag/force-atlas/>>, accessed 21-3-2014.

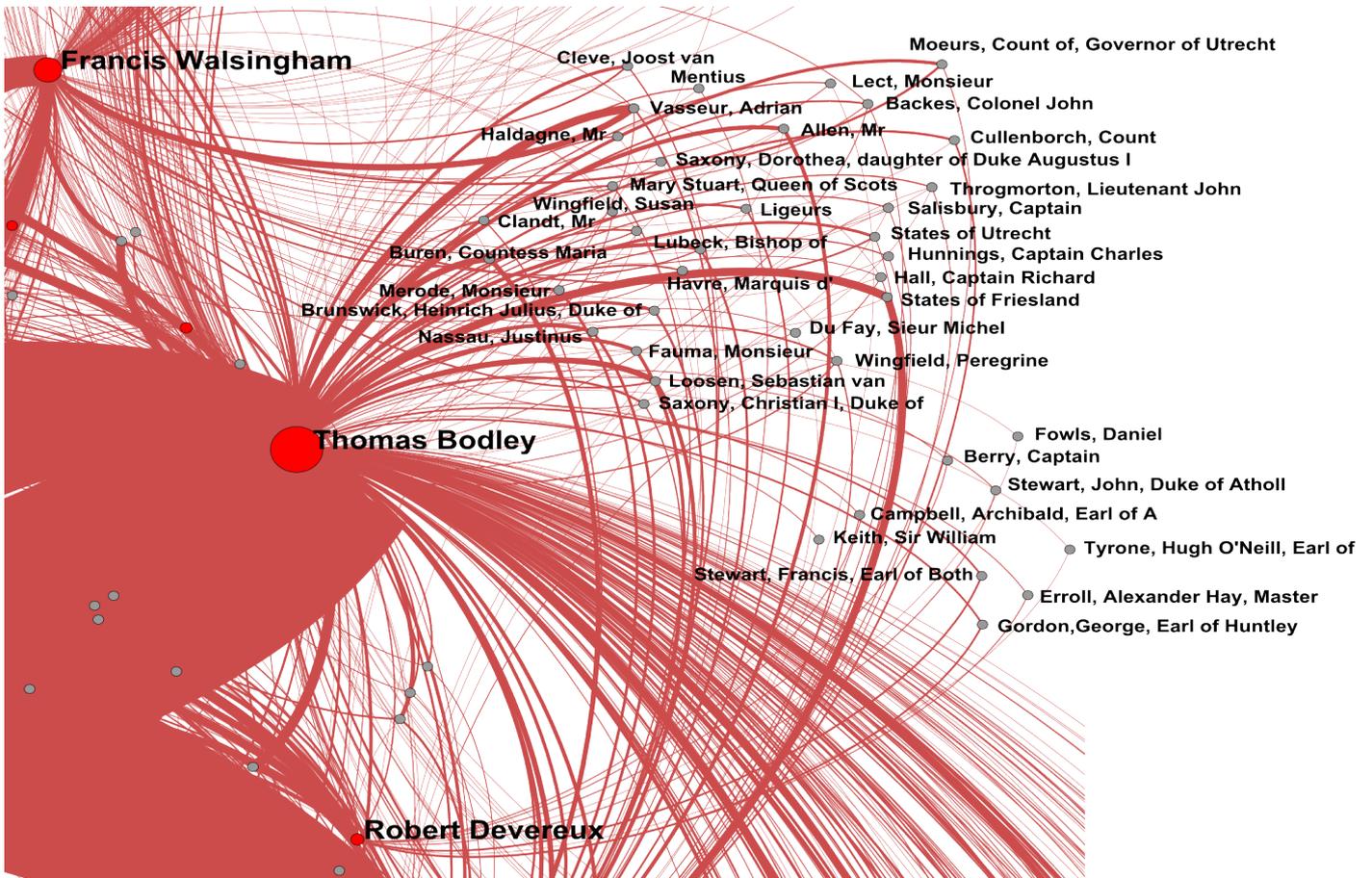


Figure 2. People: Bodley and Devereux (zoom)

To increase the readability of this visualization we applied filters to the network, as has been done in the previous image, where the in-and-out degree (the number of directed edges which go to or from a node, or the number of connections associated with a node) is set to 11. This setting of 11 filters out a number of nodes (people mentioned as well as authors/recipients) which did not meet this criterion leaving a much less dense network, and providing a concise overview of the most important correspondents and the people they mentioned in their letters.

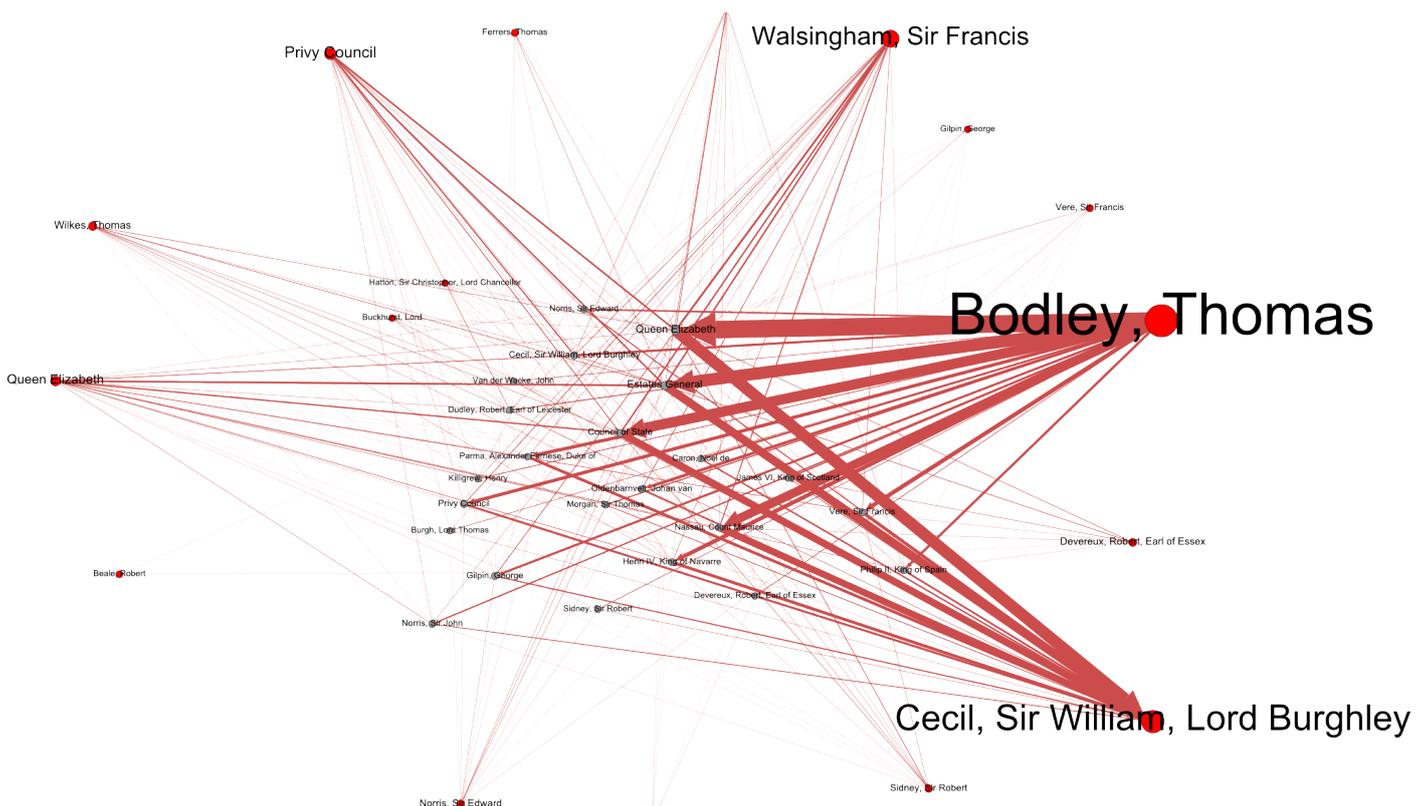


Figure 3. People: (filter with 11 degree)

As patterns within the dataset are detected by using different kinds of (statistical) analysis, we soon realized that, due to the limits of specific types of visualizations, it was better to depict the outcomes of these analyses using different *kinds* of visualizations. The dataset as a whole can be analyzed, for instance, by looking at the frequency with which places and people were mentioned, showing the degree to which the correspondence covered a wide or rather small range of people and geographical locations. This can be done quite simply by employing bar graphs.

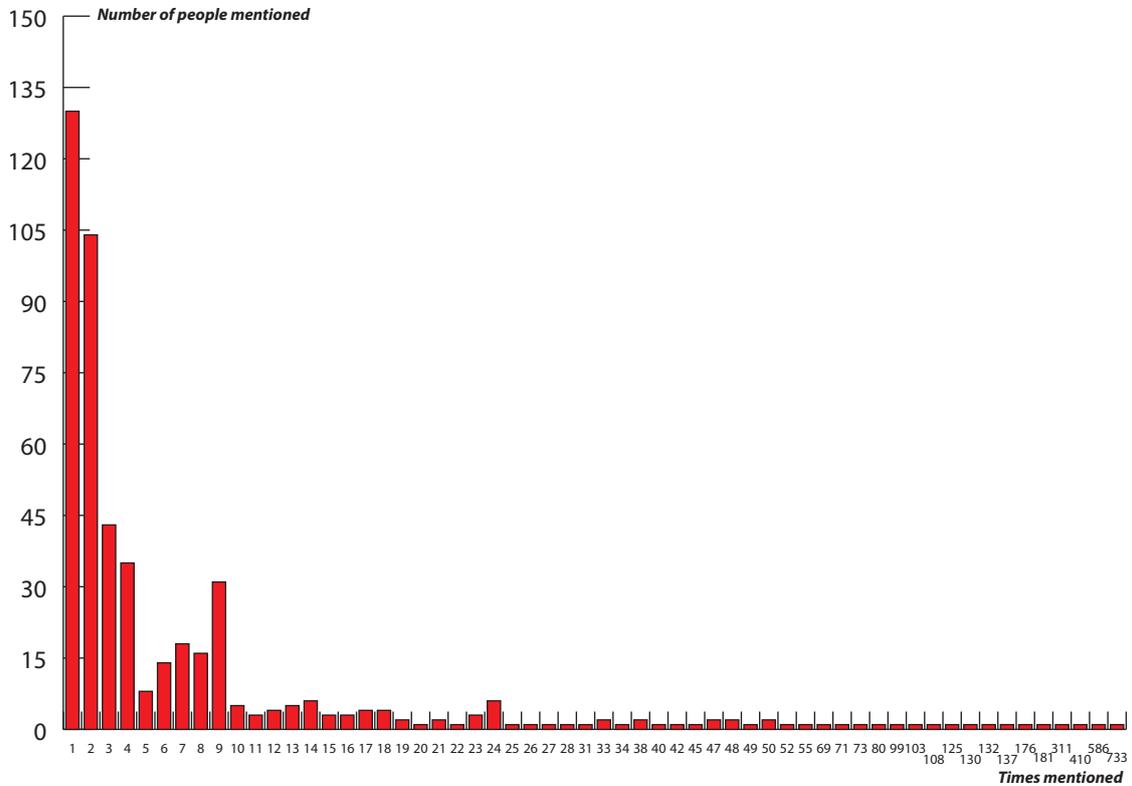


Figure 4. People: frequency of mentions (bar chart)

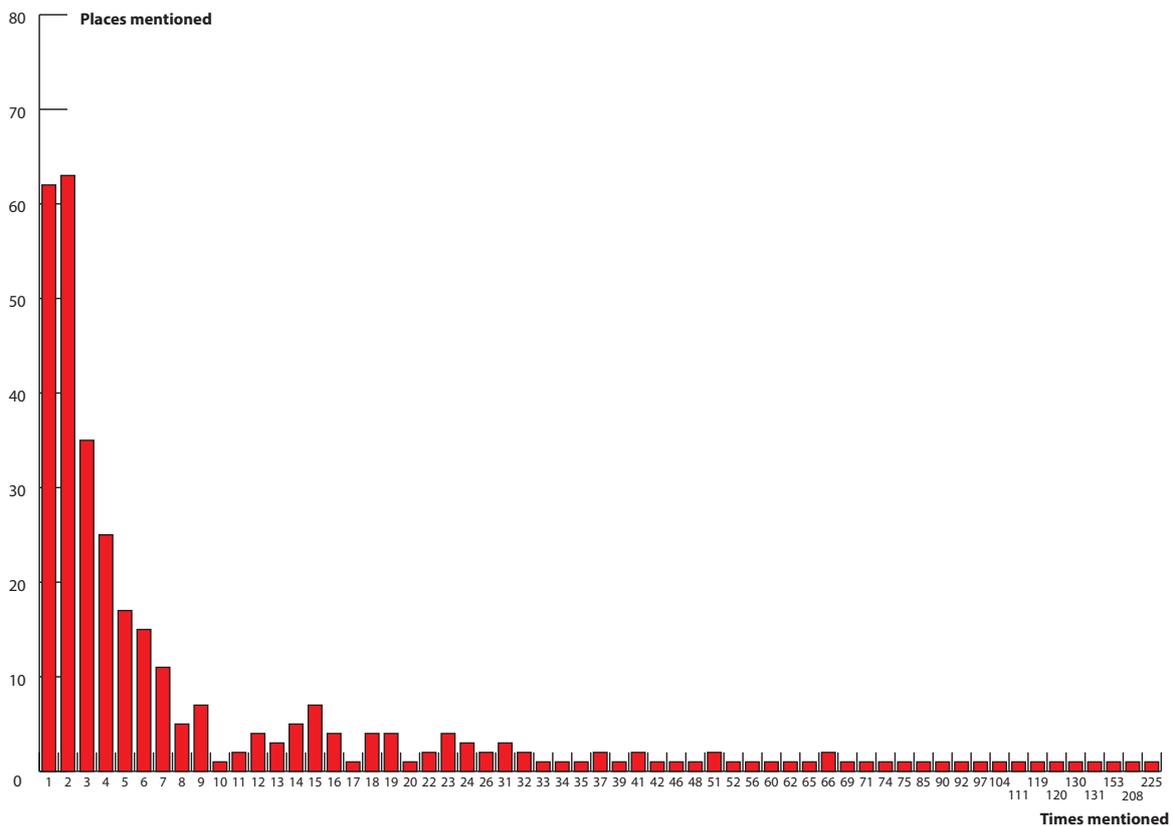


Figure 5. Places: frequency of mentions (bar chart)

Both charts show that most of the places and people were mentioned just one or two times, and that there only were a very limited number of places and people which were mentioned more than a hundred times. The correspondence, although covering a wide range of people and places, clearly centered on a select number of them (understandable, considering Bodley’s mission in the Low Countries, focused on western European affairs), and such visualizations, which give a general impression of the correspondence as a whole, give the scholar a lead on which patterns might benefit from further, in-depth analysis. More detailed information can be gathered by zooming in at the correspondence between two people, as has been done in the following bar graph.

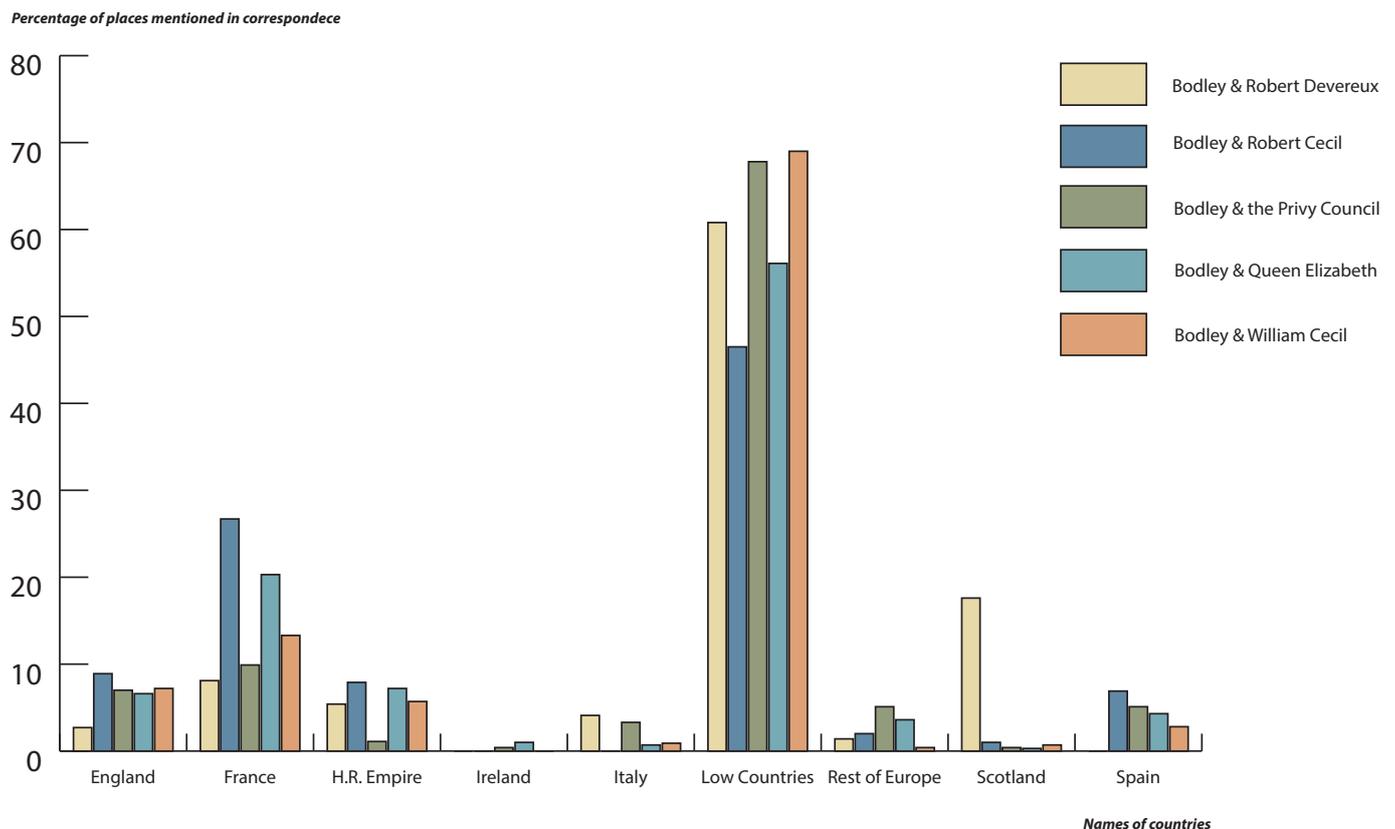


Figure 6. Places: mentions in individual conversations

By adding information to the place mentioned (the country in which these places were located), this visualization quickly shows which countries the various conversations centered. Compared to his correspondence with other people, Bodley and Sir Robert Devereux devoted a relatively large part of their letters discussing places in Scotland (which was not a surprise considering the patterns visible in other visualizations), while Bodley and Sir Robert Cecil often spoke about places in France. These are just a couple of examples of how augmenting the existing data with additional information and using a combination of different visualizations, each based on its own subset of data, can work together to create new patterns and relationship networks, that are hard to detect without the assistance of IT-tools.

A final example of some of the visualizations that have been created are the two SDL-diagrams below. The acronym SDL stands for Specification and Description Language and these diagrams are normally used to visualize aspects (e.g. actions) of a particular process taking place within a system (e.g. a computer program). We have, however, used these diagrams for creating visualizations which incorporate a large subset of the information that is to be found in the dataset, including actions the author of the letter required from the recipient, and the relationships that were established as a result of those requirements. The diagrams are based on two case studies, namely the sieges of Geertruidenberg (1589-93) and Groningen (1594) and their immediate aftermath, which generated a select sequence of letters. The diagrams make clear that writing and receiving letters was only a part of the process, as often recipients were asked to pass on to third parties (part of) the information included in the letters or documents enclosed with the letters. The visualizations show that letters did not only create a relationship between the author and the recipient, but rather forged a number of links, thus expanding the network beyond the standard binary epistolary structure, while also providing understanding of the way the information flowed through this network. For useful as they are, the network visualizations which depict the connections between the authors and the recipients only show a part of the organic and brittle process of gathering and disseminating early modern information, and omit the people who were closely related to these networks as transmission agents.

The diagrams, developed to highlight the complexity of the network (and the way information was disseminated through the network and beyond), require some explanation, and the meaning of the various icons are given in this overview.

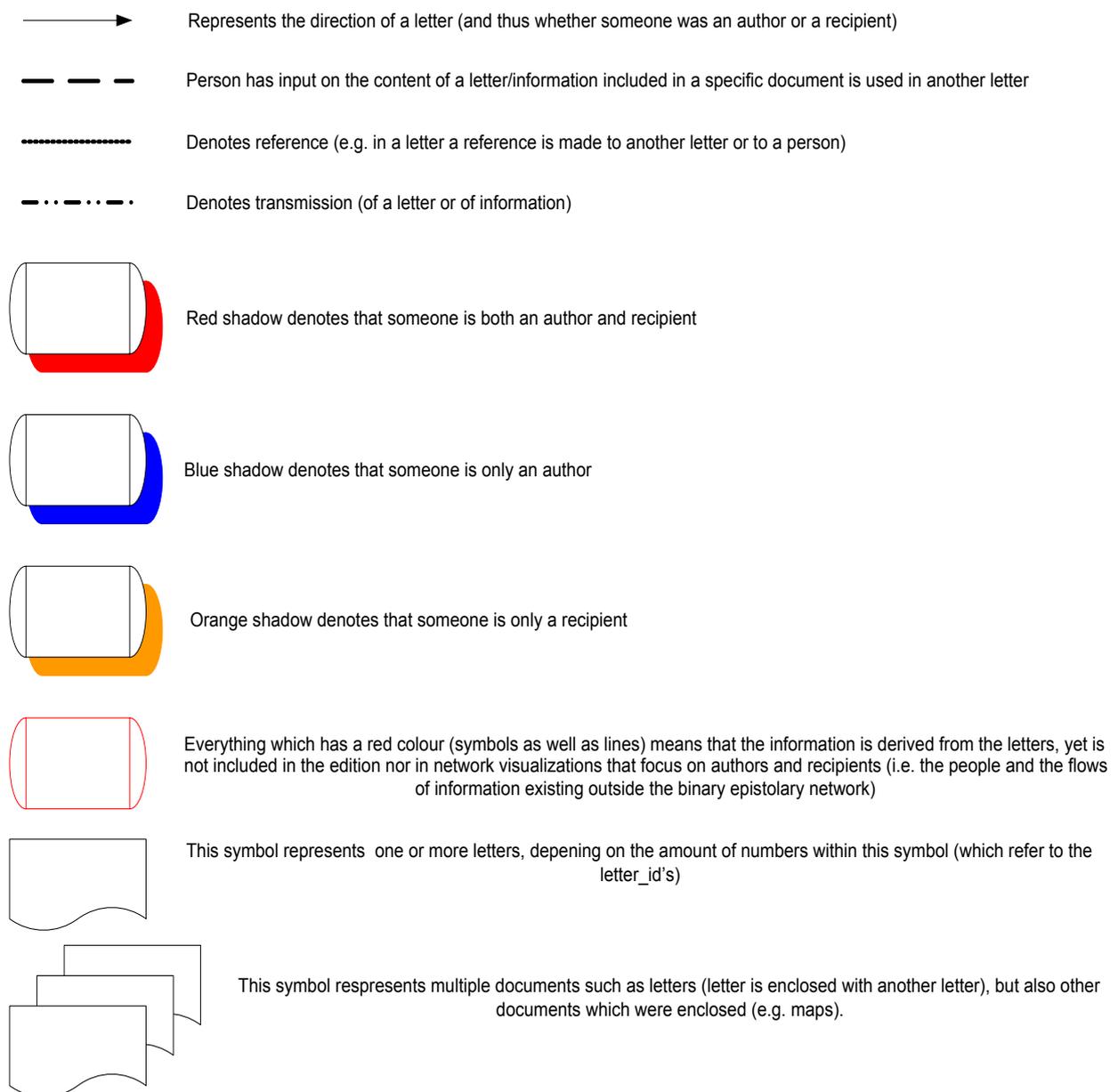


Figure 7. Key to SDL diagram

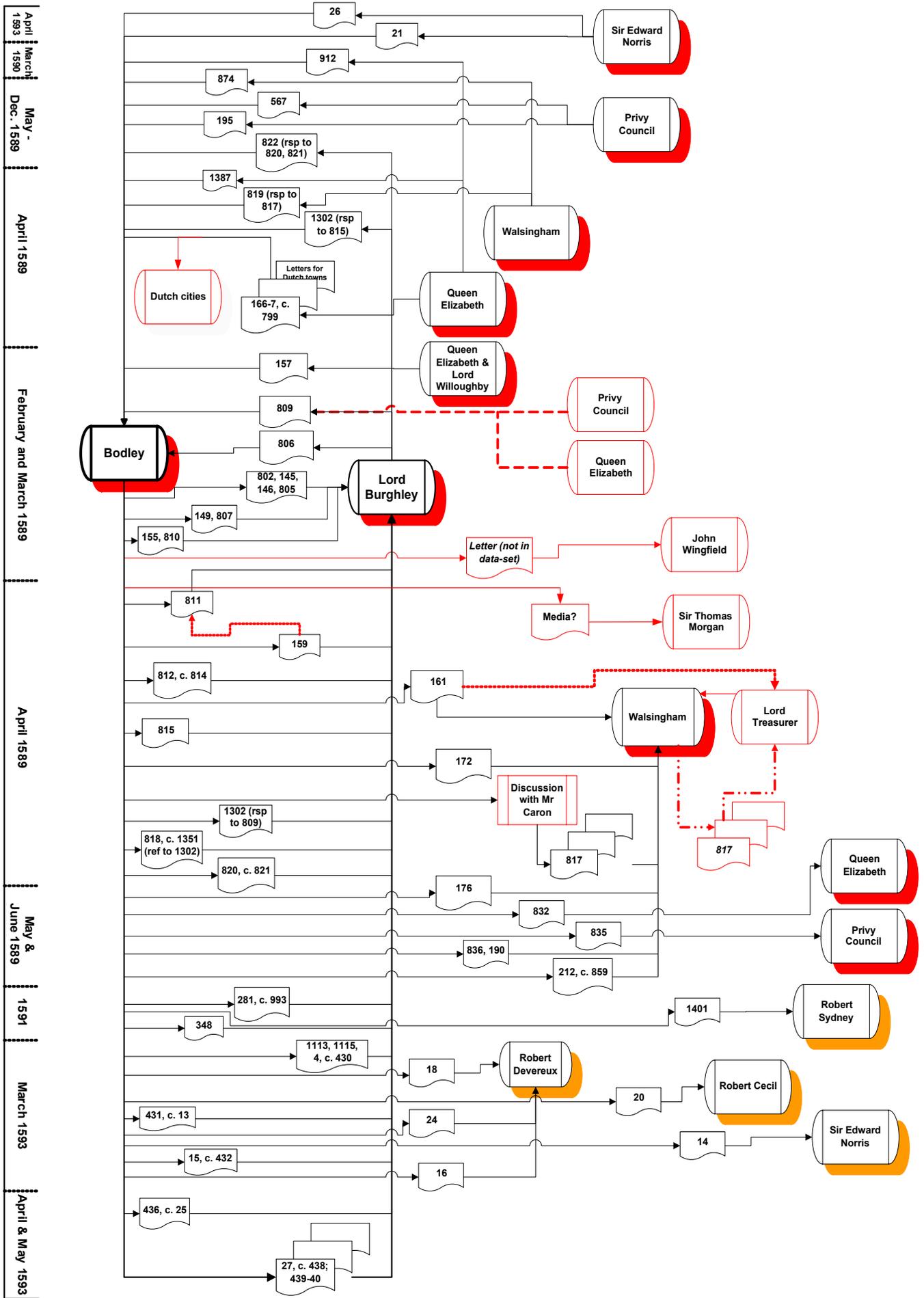


Figure 8. Case study I: Geertruidenberg (SDL)

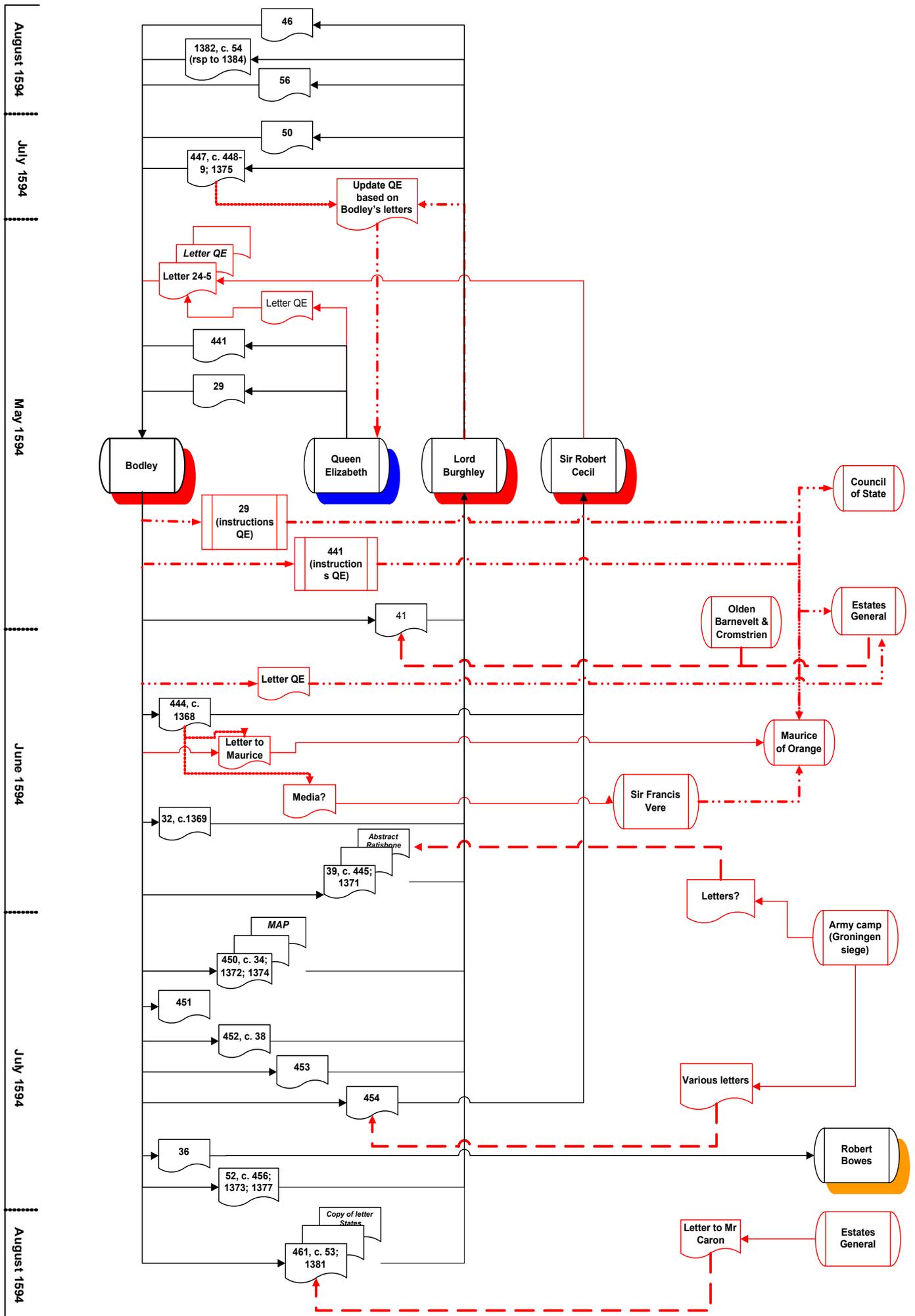


Figure 9. Case study 2: Groningen (SDL)

Let us survey a couple of examples included in the diagram which depicts the correspondence regarding the siege of Groningen. In his letter to Sir Robert Cecil (June, 6, 1594; letter id. 444), Bodley mentions that he received Cecil's letter of May 24, 1594, which is no longer extant and not a part of the dataset, (and therefore not included in the binary epistolary structure). Enclosed with Cecil's letter was a letter written by Queen Elizabeth, which Bodley duly brought to the States General 'the next day after [he had received Cecil's letter]', and they promised to answer, when according to their custome, they shall have taken some time to deliberat upon it'.<sup>11</sup> Bodley thus not only performed some actions, leading to links which are normally not included in the binary networks consisting of authors and recipients, but he also provided information about his actions and, in this case, the expected results. In the same letter Bodley tells that he tried to convince to Dutch authorities to 'arme as many shippes', and he mentions that he had written a letter to stadholder Maurice of Orange and also 'requested Sir Francis Vere, to urge him [=Maurice] to it very earnestly' (it is unknown whether Bodley did so by via a letter; hence the speculative note 'Media?'). On 14 July, 1594, Bodley wrote that he had 'lettres from the Campe at this very instant', and he included some information he derived from these letters in his letter to Burghley (letter id. 0454). By including such links in this diagram, we are able to show the larger scope of the network and, specifically, the **routes** on which this information flowed through the network.

### Outcomes

This project to analyze the dataset and to generate visualizations has proved enormously helpful in enabling us to detect patterns of communication and relationships in Bodley's correspondence network. It is clear that the process of using data visualization as an aid to historical research can identify additional and alternative routes of enquiry. For instance, only through close textual analysis would we have perceived that Bodley's correspondence with Sir Robert Devereux focused heavily on Scottish affairs, whereas it is immediately perceptible through Figure 2. People: Bodley and Devereux (zoom). Tracking the routes of transmission of specific sequences of letters – those concerning Geertruidenberg and Groningen – using the SDL diagram reveals with clarity the fluid and organic nature of a cross-channel early modern correspondence network. These examples demonstrate how valuable the visualization process can be to assist the scholar in interrogating a corpus of material with large or detailed questions.

However; there are some equally useful caveats to consider when deploying or using network visualizations in historical research which we have recognized during this learning process. It is essential that appropriate care and attention be paid to contextualizing the resource created. We have encountered numerous examples of visualizations where insufficient attention has been devoted to providing a suitable background and historical context to the results depicted. The lack of sufficient information given often results in decreased understanding of what exactly the visualization is trying to convey. Of course, too much information may have the result of data duplication; but if the context provided is robust enough (by means of an introductory summary, or running commentary with each visualization, for example), then swift perception of the data and patterns will occur. The confluence of interest in data visualization and infographics means that there is often positive overlap between well-constructed visualizations of networks and aesthetic presentation. However, problems can occur when the visualization is beautifully presented and artistically designed but the reader has little contextual information with which to understand the image.

It is also fundamentally important to take into account the individual context for each dataset. It is not enough to stumble across an open source dataset and begin generating visualizations; care must be paid to understanding the conditions surrounding the production of the data, the personnel and the (archival) material involved. In our case, certainly, the editorial questions raised by the material (what constitutes an early modern letter? what are the contemporary circumstances which prompt the writing of letters? how can we visualize the specific role of the transmission agents?) demanded that we approach the production of the visualizations with a sensitive eye to the historical context of the correspondence.

<sup>11</sup> 'The Diplomatic Correspondence of Thomas Bodley', CELL, <[http://www.livesandletters.ac.uk/cell/Bodley/transcript.php?fname=xml/1594/DCB\\_0444.xml](http://www.livesandletters.ac.uk/cell/Bodley/transcript.php?fname=xml/1594/DCB_0444.xml)>, accessed 21-3-2014.

Overall, this small project has provided us with some valuable insights. Of primary interest are the patterns and connections which were previously imperceptible without significant and time-consuming textual analysis of what is a substantial corpus of material. But it has been additionally positive to experience first-hand the value of having collected the metadata of people and places at the point of transcription. Without that extra subset of data these would have been a very bland set of network visualizations, and Bodley's small number of correspondents would have been perceptible without computational methods. The extra depth provided by this metadata has provided an alternative (and interesting!) route of research, and demonstrates that projects such as the *Diplomatic Correspondence* are value-added when the time is taken to take the data-collection stage to a higher level. Our next task will be to investigate in detail the fascinating patterns and routes of enquiry generated by the visualizations.