Note: To be published in *History of European Ideas, vol. 36 no 3 (2010) pp. 330-339*

Mapping the structure of the intellectual field using citation and co-citation analysis of correspondences*

Yves Gingras[a]
CIRST, Département d'histoire, Université du Québec à Montréal

Key words : Republic of letters, citation analysis, co-citation analysis, Mersenne, Oldenburg, Darwin, mapping, intellectual networks, correspondences

Abstract :
This article uses the methods of citation and network analysis to map the global structure of the intellectual field and its development over time. Through the case study of Mersenne's, Oldenburg's and Darwin's correspondences, we show how looking at letters as a corpus of data can provide a global representation of the evolving conversation going on in the Republic of Letters and in intellectual and scientific fields. Aggregating general correspondences in electronic format offers a global portrait of the evolving composition of the intellectual and scientific scene, its changing foci of interests and the fortune of the intellectual discussions as expressed in cited persons in the letters. Such tools help replace a purely metaphoric use of the term "network" by a visible *map* of the intellectual relations between people on which well defined calculations of the centrality of the positions of different actors can be made as well as their evolution over time. These techniques provide welcome additions to the tool kit of scholars in an age where the computer and the web offer new ways of mapping and mining the rich store of information contained in intellectual correspondences.

For more than a century now, scholars, and historians of ideas in particular, have been used to mine published and unpublished correspondences of major historical figures in order to document particular events, ideas, discoveries or debates[1]. Huge collective efforts have been devoted to publish critical editions of the extant letters of the major intellectual figures from (at least) the 16th to the 20th century, from Erasmus to Einstein[2]. Though extremely useful, erudite and handy, these printed documents could not offer their full potential until they began to be accessible in full text searchable format. This major transformation, of which the Oxford University e-enlightenment project with its more than 55,000 letters from more than six thousands actors of the "Republic of Letters" offers a prime example[3], provide a unique opportunity to go beyond the atomized study of a particular actor or set of letters and construct the whole intellectual field and its changing structure over time. As is well known, before the scientific journals made their first appearance in 1665 and until they became the primary means of diffusion of new scientific discoveries during the 19th century, letters played a central role in the circulation of information and the diffusion of knowledge. Their global analysis would offer a unique access to the ongoing conversations between scholars across the world[4].

Instead of seeing each letter as a unique document and collected editions as simply a convenient way to access them in libraries, one can look at these collected documents as a global corpus of

data to be treated as a representation of the evolving conversation going on in the Republic of Letters and in the intellectual and scientific fields. Though some work has been done in this direction of a structural analysis of correspondences, it has been limited to the study of the geographic distribution of correspondents[5] and, more recently, to the analysis of the time distribution of responses to received letters.[6] Much more can be done by using techniques developed for the citation analysis of scientific papers and for the analysis of social networks[7]. One can for instance follow the evolution of cited persons over time. Highly cited individuals give us a clue about the actors involved in the conversations of the times, their emergence and disappearance as recorded in these letters. In this way, one also gets an idea of the number of people involved in these exchanges not only through writing letters but as persons worthy of being discussed. Even more interesting than citations are co-citation networks, based on the fact that two different names mentioned together in many different letters strongly suggest the existence of a connection (social or intellectual) between the two[8]. As the co-citations of authors in scientific papers provide an entry into the conceptual map of disciplines and specialties[9], so too the co-citations of persons in correspondences offers the possibility to really map the intellectual structure of the Republic of letters by providing measures of proximity between authors, through their being cited frequently together in many different letters. Using these methods, the letter become the bearer of information on the actors of the intellectual and scientific fields and the frequency of their presence in different letters as well as their connectedness with others an index of their centrality in a given field (intellectual or scientific) at a given time.

In this essay, we would like to give some examples of the kind of results that can be obtained using bibliometric and social network techniques applied to a large corpus of letters. These techniques could easily become an integral part of the electronic editions of aggregated correspondences and serve as tools for mining and analyzing simultaneously several thousands letters covering many decades and even centuries. Aggregating general correspondences like those of Nicolas-Claude Fabri de Peiresc (1580-1637), Marin Mersenne (1588-1648), Henry Oldenburg (1615-1677) which were central nodes of intellectual exchanges, with more personal or specialized ones like those of René Descartes (1596-1650), Robert Boyle (1627-1691) Isaac Newton (1643-1727), Voltaire (1694-1778) and Lavoisier (1743-1794), to name a few major figures, would offer a global portrait of the evolving composition of the intellectual and scientific scene, its changing foci of interests and the fortune of the intellectual discussions as expressed in letters and indexed through proper names like Aristotle, Galileo, Newton or Lamarck. We could already cover a period from at least 1600 to about 1800 by using existing editions and thus get a dynamic view of the evolving discussion between philosophers, natural or not, and other actors of the intellectual field. For in addition to the thousands of scholars writing and receiving letters, analyzing cited persons in these letters give access to many more actors -- many being dead but still alive in the intellectual conversations. We could also make visible generational patterns as few actors remain central more than 15 or 20 years[10]. The more we would add letters, the more the analysis of a given year would be representative of the state of the field at that time. Many obstacles preclude the immediate realization of such a project, such as copyrights or formatting issues, but the already existing databases make it clear that such a goal is near in sight and giving some examples of the kind of global or structural analysis that could be made on such a large quantity of letters can also contribute to its development. For, as we will see, we now have the tools to replace a purely metaphoric use of the term “network” by a visible *map* of the intellectual relations between people on which well defined calculations of the centrality of the positions of different actors can be made as well as measures of the extension and density of the network

itself. Moreover, such global analysis could also resurrect figures that were, at least for a given period of time, locally central though they now appear minor to the historian.

Since a general database of letters covering a long time period does not yet exist in the form we suggest here, we will use the cases of Mersenne, Oldenburg and Darwin, to give concrete examples of how we can analyze the changing landscape of cited and co-cited authors and show how this approach, which is complementary to the usual micro-analysis of the detailed content of each letter in its context, can help to better describe and understand the global changes of the intellectual field as reflected in the correspondences.

The correspondences of Mersenne and Oldenburg: The Decline of Scholastic and the Rise of Galileo and Descartes

Covering a period of about thirty years (1617-1648), the Mersenne's correspondence is not yet available in electronic form. That period being central in the emergence of modern science and being already well-studied, it can serve as a test-bed for the validity of the methods that we will present in this paper. We have thus manually constructed a list of the most cited authors in each of the 1880 letters written by 328 individuals during the period covered. On average, the letters mention two persons living (i.e. Galileo) or dead (i.e. Aristotle)[11]. Given that the distribution of citations in letters is highly skewed and that many individuals are mentioned only once over the period, we have limited the analysis to those who are mentioned at least 10 times over the thirty-year period. This limitation could be dropped once a complete electronic version is accessible. In all cases, though, there will be a strong concentration of the citations onto a small proportion of the total number of persons mentioned in the letters. Though much more are cited at least once, there are only 86 individuals cited more than ten times and they of course include the usual figures of the intellectual field at the beginning of the 17th century namely, Descartes, Galileo, Gassendi, along with comparatively lesser known ones like Roberval or Saumaise. The limited number of central figures is also evident in that 50% of the total number of citations is concentrated among 15 individuals, that is less than 20% of the total. These results are in fact typical of Pareto distributions that are often summarized in the "20-80 rule" according to which roughly 20% of individuals account for 80% of a given variable (be it citations obtained or papers published)[12].

Dividing the results by periods, as shown in Table 1, makes more visible the rise and decline of various authors, the Greek sources in particular decline rapidly after 1630, as should be expected. Figure 1 shows the ups and down of the most cited authors over time and we see that Galileo rises rapidly after 1632, the publication year of his *Dialogue,* while Descartes becomes important only after the appearance of his *Discourse on Method* in 1637. It should be noted that most of the persons cited are of course not the ones who write or receive the letters and include long deceased intellectual figures like Aristotle, Archimedes or Augustine. While the analysis of the senders and recipients of letters help construct the *social* networks of the times, the analysis of the cited persons give access to the *intellectual* networks which can mix living and dead scholars.

Cited Person	1617-1631	1632-1636	1637-1641	1642-1648	Total
Descartes	10	4	92	147	253
Galileo	5	63	52	92	212
Gassendi(Pierre)	27	51	16	102	196
Roberval (Gilles Personnne de)		6	43	123	172
Aristotle	26	12	26	25	89
Huygens (Constantin)			43	46	89
Saumaise (Claude)	4	15	27	43	89
Doni(Jean-Baptiste)	1	43	12	30	86
Fermat		2	37	42	81
Mydorge(Claude)	20	16	15	24	75
Beaugrand (Jean de)		20	43	10	73
Archimedes	2	5	24	41	72
Morin(Jean-Baptiste)	5	26	27	11	69
Hobbes (Thomas)		2	12	50	64
Torricelli (Evangelista)			2	55	57
Grotius (Hugo de Groot)	1	1	8	45	55
Augustine	12	3	21	18	54
Apollonius	7	2	18	26	53
Viète (François)	1	4	18	30	53
Hardy (Claude)	4	21	11	15	51
Boulliau (Ismael)	1	13	8	28	50

Table 1: Most cited persons in Mersenne correspondence (1617-1648).

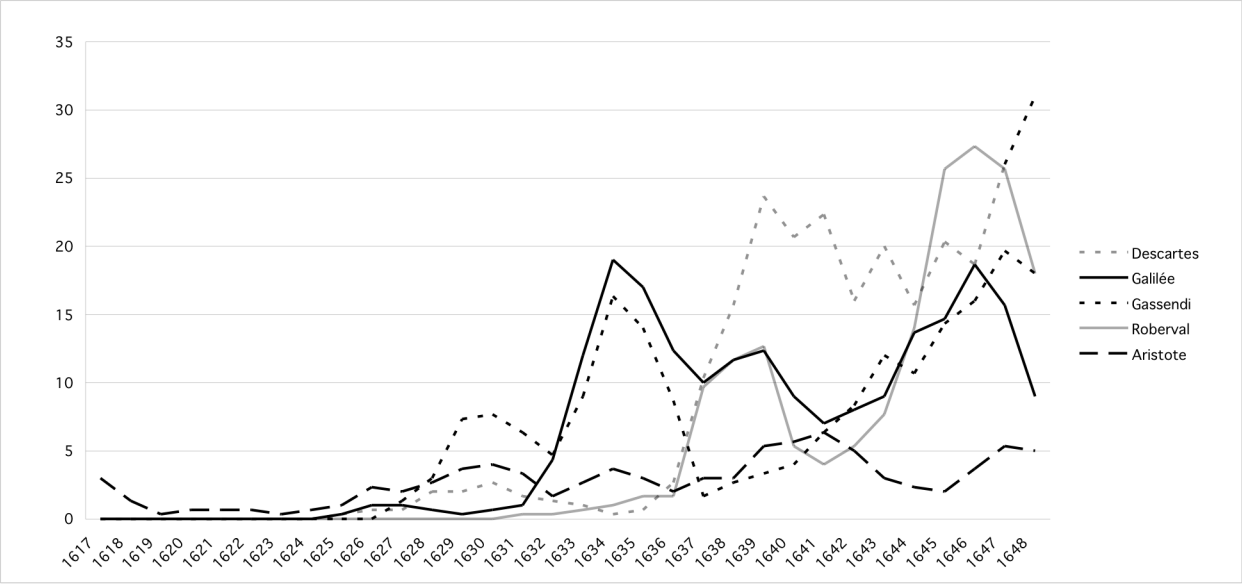

Figure 1: Most cited authors in Mersenne correspondence (3-year moving average).

Since the correspondence obviously stops with the death of Mersenne in 1648, the time line is broken but can be continued using Oldenburg's correspondence, which contains 3296 letters covering the years 1641-1677. We have thus also constructed a database of the 145 most cited persons, some of course being the same as in Mersenne but most are emerging figures[13]. As shown in Table 2, the focus of attention of the intellectual field of the mid-seventeenth century, as expressed in the letters, changed compared to the 1620s when classical authors were still quite

present in the conversations. While Descartes is still discussed in the 1650s he his displaced in the 1660s by figures like Robert Boyle and Robert Hooke until Newton appears on the scene in the 1670s, particularly after the publication of his 1672 paper on the nature of light and colors in the *Transactions of the Royal Society,* which are published by Oldenburg himself. As an element of continuity between the two correspondences, we see, in addition to Descartes, Christian Huygens (1629-1695) who became visible in the last period (1642-1648) of the Mersenne letters and is among the most cited natural philosophers in Oldenburg during the period 1662-1677. Though there is a probably a national bias since Oldenburg is in England and secretary of the Royal Society, it remains that aggregating many letters from different sources for the same period would compensate this bias and give a more complete view of state of the intellectual field. All the same, the figures emerging from Oldenburg's correspondence are those who are central to the new science of the period, which saw the move from a Cartesian to a Newtonian physics[14].

1653-1661		1662-1671		1672-1677	
Cited author	**N**	**Cited author**	**N**	**Cited author**	**N**
Jones, Richard	32	Boyle, Robert	211	Boyle, Robert	239
Boyle, Robert	27	Wallis, John	196	Hooke, Robert	145
Charles II, King of England	17	Hooke, Robert	148	Wallis, John	122
Cromwell, Olivier	15	Huygens, Christiaan	147	Huygens, Christiaan	114
Charles X, King of Sweden	11	Brouncker, William	120	Newton, Isaac	108
Descartes, René	11	Hevelius, Johannes	117	Brouncker, William	88
Louis XIV, King of France	11	Wren, Christopher	114	Grew, Nehemiah	78
Poleman, Joachim	10	Auzout, Adrien	103	Cassini, Giovanni	75
Borel, Pierre	9	Cassini, Giovanni	87	Hevelius, Johannes	72
Willis, Thomas	9	Sprat, Thomas	73	Collins, John	68
Digby, Sir Kenelm	8	Moray, Robert	72	Malpighi, Marcello	68
Bacon, Francis	7	Justel, Henri	57	Flamsteeed, John	67
Becher, Johann Joachim	7	Descartes, René	55	Sluse, René François de	56
Leopold I, Emperor	6	Collins, John	53	Descartes, René	45
Hartlib Samuel	5	Willis, Thomas	49	Willis, Thomas	42
Helmont, J. B. von	5	Barrow, Isaac	46	Picard, Jean	41
Huygens, Christiaan	5	Lower, Richard	46	Justel, Henri	36
Beale, John	4	Wilkins, John	46	Jessop, Francis	34
Boulliaud, Ismael	4	Gregory, James	41	Boulliaud, Ismael	33
Petit, Pierre	4	Ray, John	37	Charles II, King of England	33

Table 2: Most cited authors in Oldenburg correspondence (1653-1677).

Charles Darwin's correspondence: from family to science

The correspondence of Charles Darwin offers a different case from the preceding ones since his letters are more personal and includes a lot about his own family. Also his scientific network is more specialized. Whereas Mersenne and Oldenburg were still generalist who were contacted essentially for academic reasons in order to make known some research results to the larger intellectual community, one wrote to Darwin because he was Darwin and naturalist. A large portion of his correspondence is now available in sixteen volumes covering period from 1821 to 1868 and many tomes are still to come from the dedicated group of editors of the Darwin Correspondence Project at Cambridge University to cover the years until his death in 1882[15].

Here we had the chance to use the complete electronic version of the index graciously given to us by the team of the Darwin Project so that all cited persons are included and not only the most cited ones. So, instead of having only a subset of the citations, we have the complete set and our database thus contains 3,758 cited individuals for a total of 18,691 citations, over the period 1821-1867 – the latest year available to us. This set confirms the that the distribution of cited persons is very skewed: 50 individuals (1.3% of the persons mentioned) account for a third of the total number of citations and 157 (or 4.1 % of the persons cited) account for 50% of the total citations. In fact, only 50% of the individuals appear more than once in the letters. If we limit the core network to the persons mentioned more than 45 times over the period, we get only 50 persons. Contrary to what was the case for the correspondences of Mersenne and Oldenburg, we see in Table 3 that the family occupies an important place in the letters. This is of course to be expected but again, it also serves to confirm the usefulness of the methods used. The focus of the conversation changes with time and the geologist Charles Lyell becomes an important part of it in the mid 1830s. Before that time, family members were the most present in his letters and naturalists take the central position after his return from his voyage on the Beagle with a mix of well-known naturalists who are also friends (Hooker, Gray, Owen, Huxley) occupying an important place immediately after Lyell.

1821-1831		1832-1836	
Cited Author	**N**	**Cited Author**	**N**
Darwin, Robert Waring	40	Darwin, Robert Waring	74
FitzRoy, Robert	22	FitzRoy, Robert	56
Henslow, John Stevens	22	Darwin, Susan Elizabeth	45
Darwin, Susan Elizabeth	18	Henslow, John Stevens	41
Hope, Frederick William	16	Owen, William Mostyn, Sr	34
Owen, Fanny Mostyn	15	Langton, Charlotte	28
Jenyns, Leonard	14	Williams, Sarah Harriet	28
Owen, William Mostyn, Sr	14	Biddulph, Fanny Myddelton	26
Sedgwick, Adam	11	Sedgwick, Adam	24
Parker, Marianne	10	Wedgwood, S. E.	24

1837-1847		1848-1858		1859-1867	
Cited Author	**N**	**Cited Author**	**N**	**Cited Author**	**N**
Lyell, Charles	87	Hooker, Joseph Dalton	107	Lyell, Charles	243
Darwin, Robert Waring	78	Lyell, Charles	98	Hooker, Joseph Dalton	233
Darwin, Emma	70	Darwin, Emma	97	Gray, Asa	197
Henslow, John Stevens	58	Hooker, Frances Harriet	52	Huxley, Thomas Henry	195
FitzRoy, Robert	40	Forbes, Edward	51	Owen, Richard	162
Humboldt, A. von	39	Darwin, Henrietta Emma	46	Darwin, Henrietta Emma	160
Wedgwood, S. E.	37	Candolle, Alphonse de	45	Darwin, Emma	136
Forbes, Edward	36	Gray, Asa	44	Hooker, Frances Harriet	125
Ehrenberg, Christian G.	30	Owen, Richard	41	Murray, John	120
Brown, Robert	29	Lindley, John	39	Bentham, George	113

Table 3: 10 Most cited persons in Darwin's correspondence (1821-1867).

Following the citations of a given person over time can help identify peaks of activity on which one could then focus more closely by reading the letters. Thus, we see a sharp peak of references to her daughter Henrietta in 1860. Looking up at the entries we see that she was sick and the bulk of the conversation turned around her health problems. On the scientific side, we can follow the

curve (not shown here) of Jean-Baptiste Lamarck, to find a first peak (6 citations) in 1860 and a second, larger one (12 citations), in 1863. Before that, he appears only 12 times over a period of 40 years. This trend can be useful for the discussion of the role of Lamarck in Darwin's thoughts and discussions before and after the publication of The Origins of Species in 1859.

Instead of following particular individuals over time, we can focus on a particular year. If we take the year 1831 for example, when the decision to go or not on the *Beagle* was taken, we find that by far the three persons most often mentioned in the letters are Robert FitzRoy, (22) Darwin's father (15) and his former professor of Botany, the Reverend John Stevens Henslow (13), who was pivotal in securing Drawin a place on the *Beagle*. So, starting from citation trends can provide a useful entry into the correspondence and an electronic database of all letters could easily incorporate that function by producing a list of cited persons with "hot links" directing to the content of the letters. One could limit the analysis to more homogeneous correspondences pertaining to a given discipline or a given geographical space, like France or England and compare the relative visibility of individuals, some being limited to the network of their field of specialization, while others, like Darwin probably, going well beyond it.

Intellectual networks in correspondences

Though useful as a first index of centers of discussions going on at a given time or to follow the fortune of a particular person over time, citations remain a limited indicator. More useful as an indicator of the intellectual content of the conversation contained in the letters is the analysis of the co-citations of authors. Two authors, say Galileo and Descartes, who are both mentioned *in the same letter* are said to be co-cited. If the number of these co-citations is high, that is if they are cited together in many letters, it suggests that there is a strong link between these persons (as perceived by the writer of the letter), since the chance that two persons having nothing in common are *frequently* cited together in many different letters is low[16]. The network of co-citations can thus be interpreted as a measure of closeness between the persons co-cited. Co-citation analysis of scientific papers is a well-developed tool in bibliometrics. What we do here is simply applying the idea to letters. Of course, to a larger extent than for citations and co-citations of scientific papers, there are many reasons for a person to cite and co-cite authors in a given letter. In addition to naming in the context of a scientific discussion, one can also discuss family questions or mere gossips about relatives or enemies. Nonetheless, significant patterns emerge when mapped with standard social networks software[17]. Moreover, the average number of persons cited in letters is between 2 and 3 for the cases analyzed here so that high level of co-citation really indicate close links between the co-cited individuals.

Figures 2 and 3 show the structure of the co-citations networks in Mersenne's correspondence in a graphical manner that makes visible the radical transformation of the dominant authors present in the conversations: we clearly see an important network of related classical figures in the period 1617-1631 that essentially disappears in the later period (1642-1648) except, significantly for the Greek mathematicians Appolonius, Euclide and Archimedes, the latter being Galileo's model. The intensity of the connecting lines (edges) is proportional to the number of times the two persons are co-cited. That figure also shows a central "clique", that is a group of persons strongly linked to each other: Descartes, Gassendi Galileo and Roberval. Note that whereas the strong

links between the first three were to be expected, the presence of mathematician Gilles Personne de Roberval on the same level is more surprising, since he is not usually closely associated with the central figures of the period. Other smaller cliques are visible linking secondary persons like Grotius, Huygens and Somaise to Descartes. These data remind us that the usual histories of intellectual life tend to retain only the names of the major actors (here Descartes and Galileo) while in fact many other now forgotten or minor figures were in fact important in the conversations of the times and the mapping of co-citations makes them visible again. Note that close relations between two persons may be positive or negative since they may in fact be debating both sides of an issue like, for example, Descartes and Gassendi on atomism, the former negating their existence while the latter promotes it. Though that would require a manual codification, one can imagine that a richer version of the database could indicate whether the citation and the relation is positive or negative. In the meantime, such an information can be obtained from the existing contextual histories on the cited or co-cited persons.

Figure 4 shows the co-citation network in the Oldenburg correspondence for the period 1656-1677 with a clique formed by the well known figures of Huygens, Hooke, Boyle, Wallis and the much lesser known Brouncker, with Descartes weakly connected to Boyle. Here we have made the diameter of the node proportional to the total number of citations received by that person for the period studied. One could analyze in more detail such structures but here we leave it at that since our main objective is to show how such analysis can be conducted. Going into more details would transform the analysis into a case study of a particular network at a particular time. Obviously, with the use of quick time or similar software, these mapping techniques can be used to create “movies” showing the successive state of the field over time by showing connected periods like 1650-1660, 1655-1665, etc. As was the case for the citations, one could incorporate a “hot-link” to co-citations in the electronic correspondences and “click” on the interesting ones to see the letters in which they are discussed together. In this way, this structural and global analysis of correspondence could be directly related to the micro-analysis of the content of specific sets of letters defined by the corresponding co-citation networks.

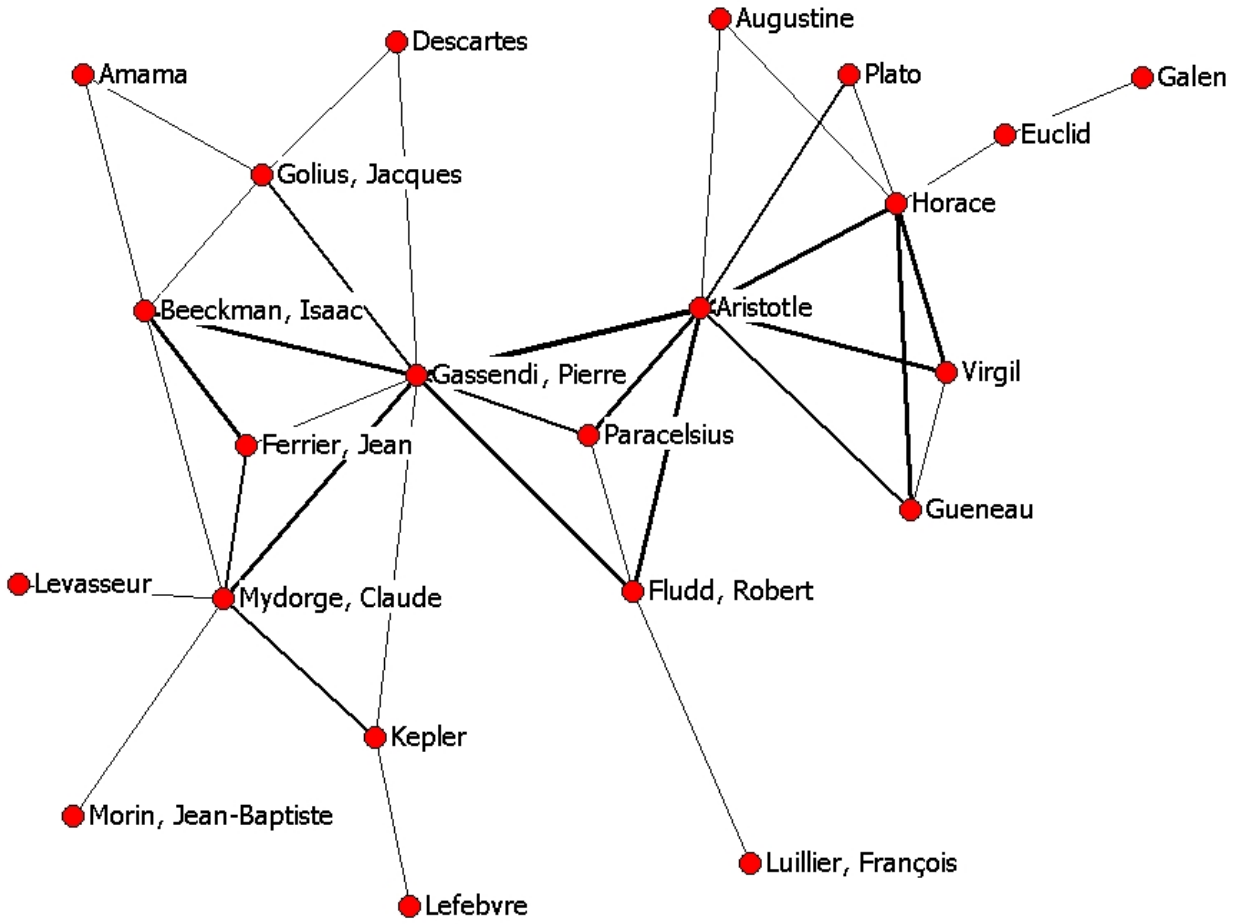

Figure 2: Co-citation network in Mersenne correspondence (1617-1631), 3 co-citations or more). The thickness of the links is proportional to the number of co-citations.

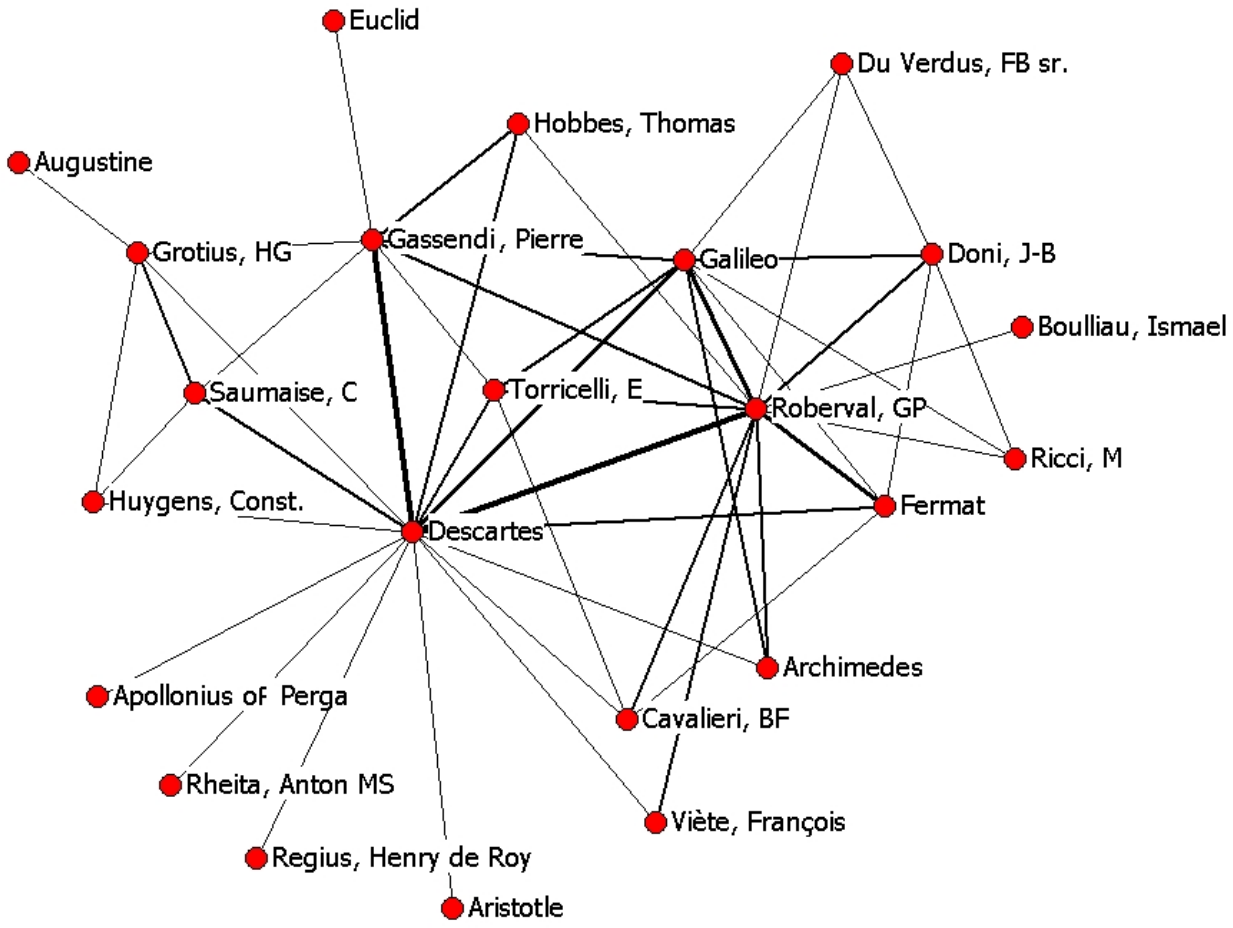

Figure 3: Co-citation network in Mersenne correspondence (1642-1648), 10 co-citations or more. The thickness of the links is proportional to the number of co-citations.

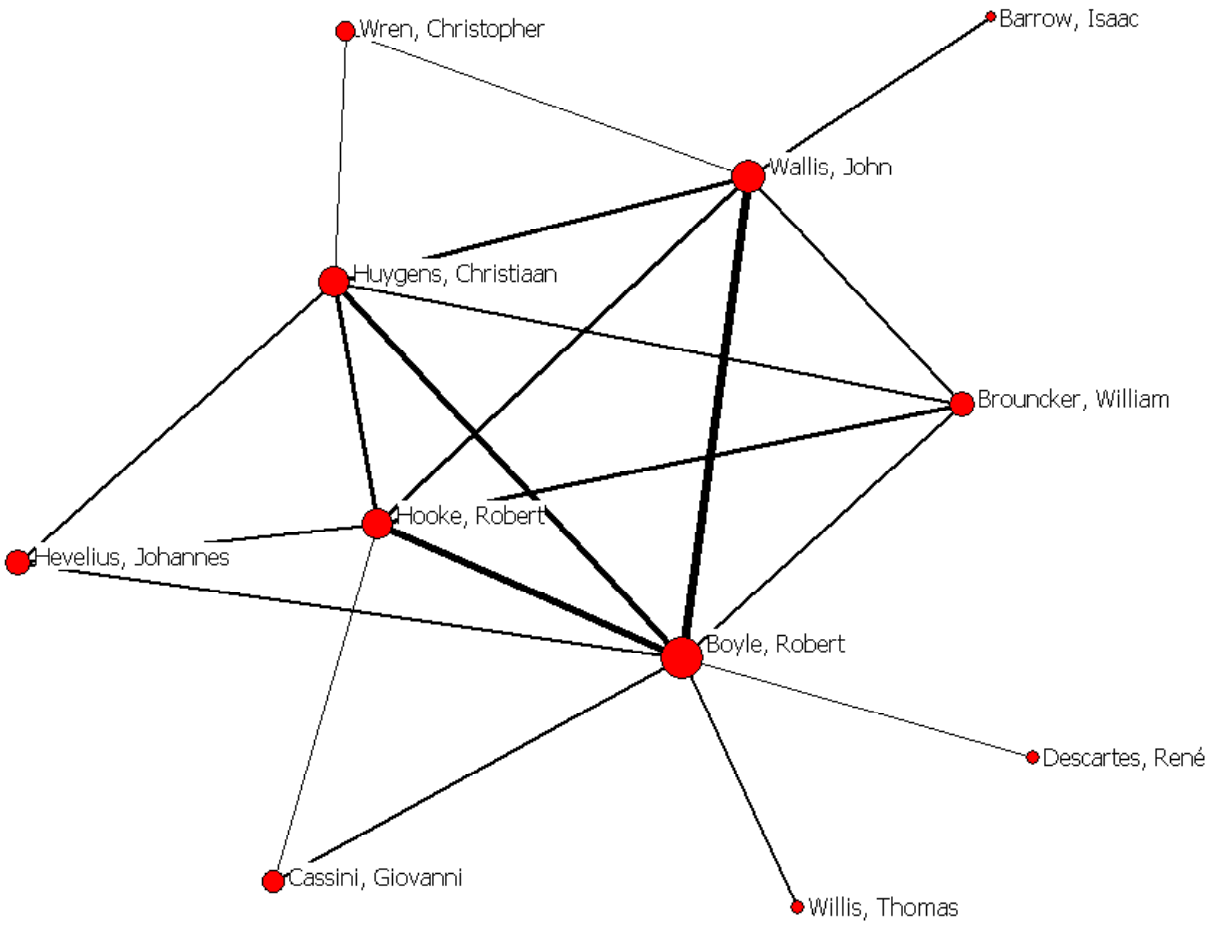

Figure 4: Co-citation network in Oldenburg correspondence (1656-1677), 40 co-citations or more. The size of the nodes is proportional to the number of citations and the thickness of the links is proportional to the number of co-citations.

Figures 5 and 6 shows the network of the most important links found in Darwin Correspondence for two time slices: 1821-1829 and 1859-1867. In the first period, the central persons are members of the family who are linked together (circles) while scientists (squares) are also linked together in a relatively distinct sub-structure. As we should expect, the period 1859-1867 gives a more central place to scientists and there is an even clearer separation between them and the family, which is at the periphery of the main network. Note that the mapping method used makes it possible to show, in addition to names, the professions of the co-cited persons (using different colors or geometry for the nodes: circle, triangle, etc) thus providing a richer vision of the properties of the network. One could also construct these maps by separating letters *to* and letters *from* a given individual and thus identify for instance the particular network constructed by Darwin himself in his letters. Finally isolated couples in lower and upper right corners of Figure 6 suggest that some people or topics are not at all connected to the general network. In interpreting these data, we must always keep in mind the possibility that many letters were lost and that as an individual become more famous he/she may have better husbanded his/her letters more carefully, thus creating a possible bias in the network. But, again, the contingent aspect of the remnant letters could be compensated by adding more letters form other sources. In the case of Darwin for

example, adding the letters of Charles Lyell or those of the American naturalist Asa Gray, would give a better idea of the structure of the field of natural history in the 1860s as reflected in their conversations.

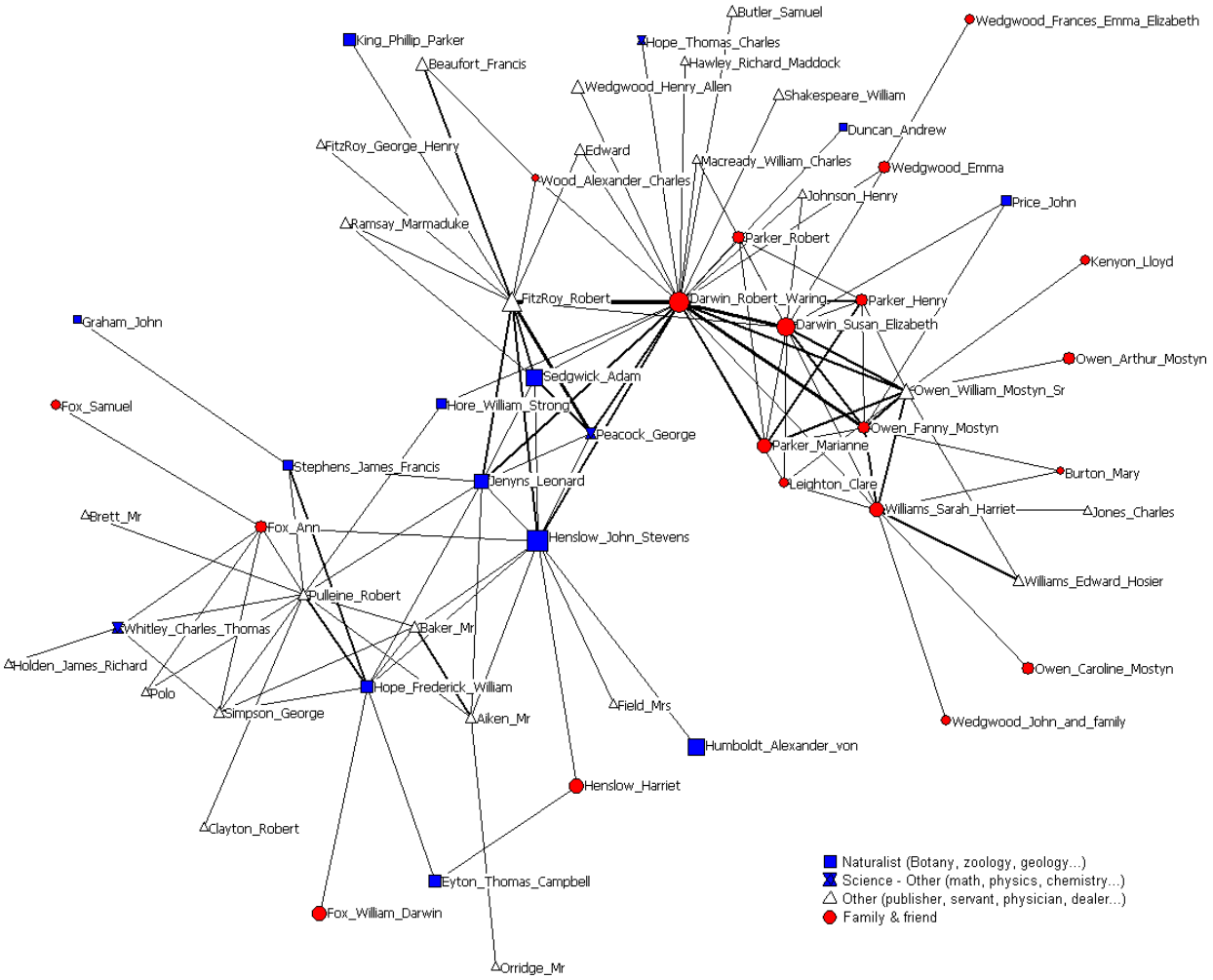

Figure 5:co-citation network in Darwin's correspondence (1821-1831). Only links stronger than 2 co-citations are shown. The size of the nodes is proportional to the number of citations and the thickness of the links is proportional to the number of co-citations.

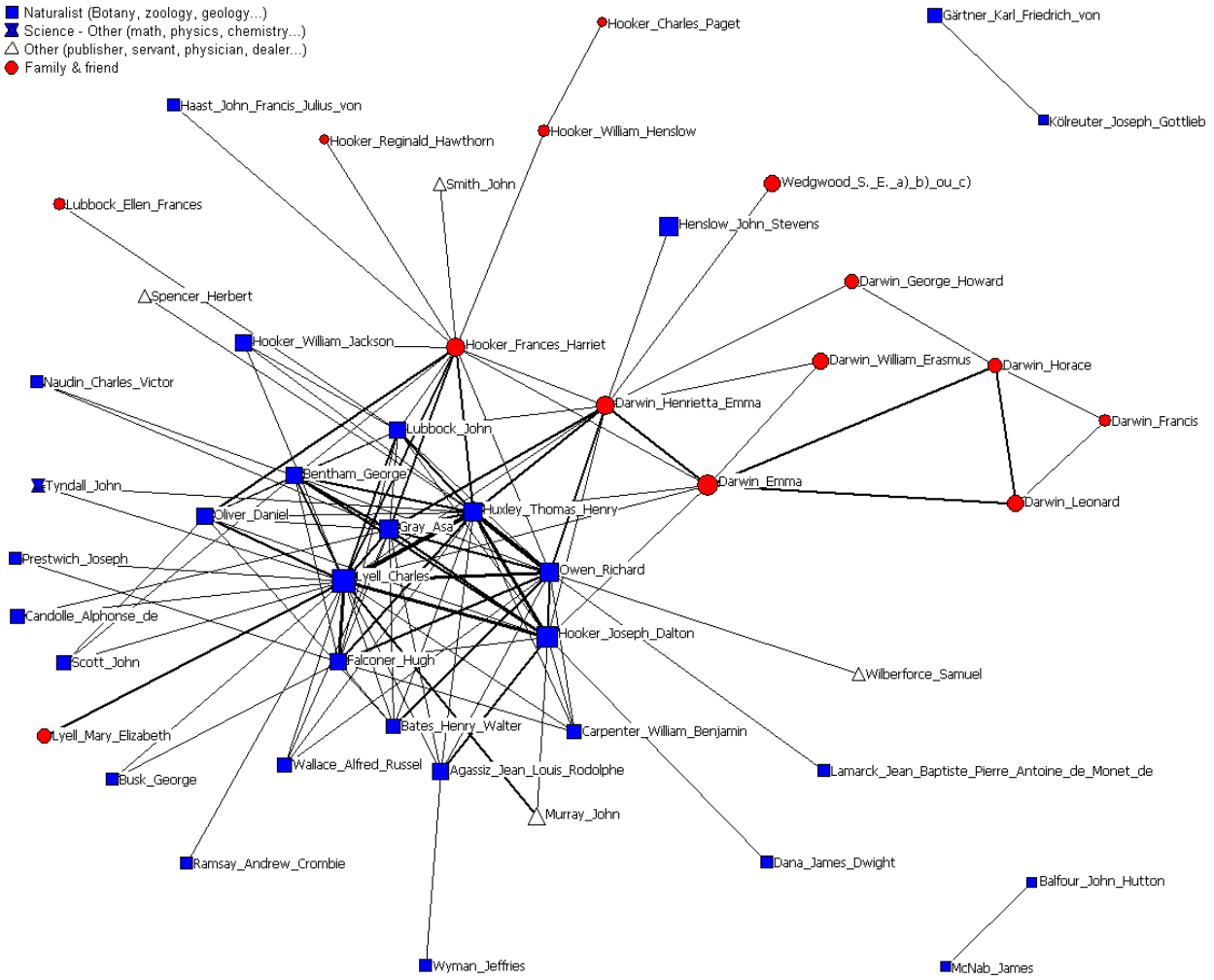

Figure 6: Co-citation network in Darwin's correspondence (1859-1867). Only links stronger than 10 are shown. The size of the nodes is proportional to the number of citations and the thickness of the links is proportional to the number of co-citations.

Once co-citation networks are constructed, the quantitative methods associated with network analysis can be used to calculate the centrality of individuals in the network and see how it changes in time[18]. Though centrality is highly correlated with citations, it offers a different point of view by focusing on the *links* (relations) instead of on the *nodes* (individuals). Table 4 shows the changing centrality of the major individuals discussed in the correspondence of Mersenne, as measured by the total number of their links with all the other persons in the network. Obviously, peripheral individuals have few links and are thus at the periphery of the network. Comparing the periods, we see (and expect) the rising centrality of Galileo and Descartes over time as both become dominant in the period 1637-1641. Table 5 shows the same data for the correspondence of Oldenburg and we see that Descartes stays central in the 1650s as he was a decade before while Newton moves to a central position in the mid-1670s. In Oldenburg's correspondence, the discussion seems largely focused on living scientists and even Galileo is no more really central in the conversation of that period. Finally, Table 6, shows that in the case of the Darwin Correspondence, scientists become more and more central as time goes on while the family move to a more peripheral position, as should be expected as Darwin's reputation grows over time, making him more visible and also more active and focused on the scientific discussions of his time. For very large networks, which can contain relatively autonomous sub-groups, one could also apply community detection techniques to automatically identify coherent subsets of actors[19].

It should be noted that the methods used here to map co-citations networks could also be applied to map the structure of co-words, that is the analysis of co-occuring and important words or concepts in the letters. For once the letters are in full-text format, one could search for given words, concepts or expressions and how they are co-present or not in letters with other specific words[20]. This would produce maps of connected themes and their evolution in time.

Rank	1617-1631	1632-1636	1637-1641	1642-1648
1	Aristotle	Gassendi(Pierre)	Descartes	Descartes
2	Gassendi(Pierre)	Galileo	Galileo	Roberval (Gilles Personnne de)
3	Mydorge(Claude)	Campanella (Tomaso)	Roberval (Gilles Personnne de)	Gassendi(Pierre)
4	Beeckman (Isaac)	Doni(Jean-Baptiste)	Desargues (Girard)	Galileo
5	Fludd (Robert)	Hardy (Claude)	Beaugrand (Jean de)	Torricelli (Evangelista)
6	Horace	Gailhard (Noel)	Fermat	Hobbes (Thomas)
7	Kepler	Dupuy (Jacques)	Petit (Pierre)	Fermat
8	Descartes	Morin(Jean-Baptiste)	Huygens (Constantin)	Viète (François)
9	Virgile	Gaulmin (Gilbert)	Archimedes	Saumaise (Claude)
10	Golius (Jacques)	Gilles de Loches (le Père)	Morin(Jean-Baptiste)	Archimedes

Table 4: Most central persons in co-citation network of Mersenne correspondence (1617-1648)

Rank	1653-1661	1662-1671	1672-1677
1	Boyle, Robert	Wallis, John	Boyle, Robert
2	Jones, Richard	Boyle, Robert	Wallis, John
3	Descartes, René	Huygens, Christiaan	Hooke, Robert
4	Charles II, King of England	Hooke, Robert	Newton, Isaac
5	Louis XIV, King of France	Auzout, Adrien	Huygens, Christiaan
6	Cromwell, Olivier	Hevelius, Johannes	Collins, John
7	Borel, Pierre	Cassini, Giovanni	Cassini, Giovanni
8	Huygens, Christiaan	Brouncker, William	Descartes, René
8	Roberval, Gilles Personne de	Wren, Christopher	Brouncker, William
9	Poleman, Joachim	Moray, Robert	Malpighi, Marcello
10	Willis, Thomas	Gregory, James	Grew, Nehemiah

Table 5: Most central persons in co-citation network of Oldenburg correspondence (1653-1677)

Rank	1821-29	1830-39	1840-49	1850-58	1859-67
1	Darwin, Robert Waring	Darwin, Robert Waring	Darwin, Emma	Lyell, Charles	Huxley, Thomas Henry
2	Darwin, Susan Elizabeth	Darwin, Susan Elizabeth	Lyell, Charles	Hooker, Joseph Dalton	Lyell, Charles
3	Owen, Fanny Mostyn	FitzRoy, Robert	Darwin, Robert Waring	Darwin, Emma	Hooker, Joseph Dalton
4	Owen, William Mostyn, Sr	Owen, William Mostyn, Sr	Forbes, Edward	Owen, Richard	Gray, Asa
5	Parker, Marianne	Henslow, John Stevens	Henslow, John Stevens	Hooker, Frances Harriet	Owen, Richard

Table 6: Most central persons in co-citation network of Darwin correspondence (1821-1867).

Conclusion

The exchanges of letters between scholars being the central mean of circulation of knowledge for at least the period 1600-1800, a global analysis of intellectual change could profit by making use of the new software for texts analysis in order to mine the large body of data contained in these letters which are more and more digitized and made available through the world wide web. The brief examples of global and structural analysis of intellectual and scientific correspondences provided here using the case of Mersenne, Oldenburg and Darwin were only meant to show in a concrete manner how such a global database could be constructed and used in conjunction with techniques of bibliometrics and social network to visualize the evolving conversations involving the many thousands of persons mentioned in the letters. At this point, of course, the results obtain only confirm what historians of science already know about the history of ideas during the times of Mersenne, Oldenburg and Darwin, though the strong presence of Roberval along with Descartes and Galileo may be a more surprising result in light of the usual history of ideas focusing only on the latter two figures. But such a confirmation should be welcome: by first showing that the kind of results obtained from citations and co-citations analysis are consistent with what we know, one can be confident that applied to less well-known periods and corpus of letters, these methods will also produce robust and meaningful results. One could also create maps based on different groups of persons from different countries and see if their focus of attention are very different or not. One could then maps national republic of letters and see to what extent, for a given time period, their foci of interest is determined by those that are the current fads in centers like Paris or London. Likewise, the centrality of individuals could also be

analyzed in relation to the country of origins of letters to make visible the fact that one could be central figure at the regional scale but still peripheral at the international level. By merging many different sets of correspondence one could go beyond the analysis of a particular network (like that of Darwin) and see to what extent scholars were part of one or many conversation networks. Instead of the partly contingent collections of letters defining the networks, such merging could test whether the sub-networks are robust by using community detection techniques. Finally, letters having different functions than scientific papers and books, one could compare citations and co-citations obtained from these three sources to see if the individuals who are the most present in the private conversations of the letters are also the ones frequently mentioned in the more formal and coded expressions of the scientific paper and the book.

Though one should not expect that the techniques used here will lead one to discover a completely forgotten individual who had in fact been central for a decade or so in the intellectual field of the17th or 18th Century, it remains that such techniques can offer a unique view of the global structure of the intellectual field and its transformation over space and time. They could easily be integrated into the fabric of the database governing access to the existing electronic letters as found for example in the e-enlightenment project. Citation and co-citation analysis, as well as the more common geographical networks of connections between authors and receivers of letters, could nicely serve as a background to the usual micro-analysis of the specific content of the letters and the particular nature of the discussions concerning the co-cited persons. These techniques would provide welcome additions to the tool kit of scholars in an age where the computer and the web offer new ways of mapping and mining the rich store of information contained in intellectual correspondences.

NOTES

[*] I would like to thank Alain Couillard for the production of the figures and Vincent Larivière and the anonymous referees of the journal for useful comments and suggestions on an earlier version of this paper. Thanks also to the many comments received over the last ten years as I presented preliminary results of this research at the REHSEIS research center of Université of Paris VII in 1999, the Dibner Institute for the History of Science and Technology at MIT in 2000, at Pietro Corsi's seminar at Oxford University in 2007 and at the History Department of Ottawa University in 2010.

[a] e-mail : gingras.yves@uqam.ca.

[1] David A. Kronick, "The Commerce of Letters : Networks and 'invisible Colleges, in Seventheenth- and Eighteenth-Century Europe", *Library Quarlerly*, vol. 71 (2001), 28-43; Wacquet, françoise, "Les éditions de correspondance savantes et les idéaux de la République des letters", *XVIIe siècle*, no 178 (Jan.-March 1993), 99-118; Nellen, H.J.M., "La correspondance savante au XVIIe siècle", *XVII*e *siècle*, no 178 (Jan.-Mars 1993), 87-98; Ultee, Martin, "The Republic of letters : Learned Correspondence, 1680-1720", *Seventeenth century*, 2, (Jan. 1987), 95-112; Hans Bots, "Éditions de correspondences aux XIXe et XXe siècles. Méthodes et strategies", *XVII*e *siècle*, no 178 (Jan-March 1993), 119-129; Robert A. Hatch, "Peiresc as

Correspondent: The Republic of Letters & the 'Geography of Ideas'", in Brian P. Dolan (Ed.), *Science Unbound. Geography, Space and Discipline* (Umeå: Umeå Universitet, 1998), 19-61.

[2] Michael Hunter, "Whither editing?", *Studies in History and Philosophy of Science*, 24 (2003), 805-820; Nellen, H.J.M., "Editing 17th-Century Scholarly Correspondence : Grotius, Huygens and Mersenne", *Lias*, 17 (1990), 9-20. For a survey of many projects of edition of correspondences see the special issue on this topic in *Revue de synthèse*, third series, nos 81-82, Jan-June1976.

[3] See www.e-enlightenment.com.

[4] Pearl, J. L., "The Role of personal Correspondence in the Early Exchange of Scientific Information in Early Modern France", *Renaissance and Reformation*, 20 (1984), 106-113; René Taton, "Le rôle et l'importance des correspondances scientifiques aux XVII e et XVIIIe siècles", *Revue de synthèse*, 3rd séries, nos 81-82 (Jan.-June 1976), 7-22; Paul Dibon, "Les échanges épistolaires dans l'Europe savante du XVIIe siècle", *Revue de synthèse*, 3rd séries, nos 81-82 (Jan.-June 1976), 31-50; Andrea Rusnock, "Correspondence Networks and the Royal Society, 1700-1750", *British Journal of the History of Science*, 32 (1999), 155-169. Benoît Melançon (Ed), *Penser par lettre* (Montreal : Fides, 1998).

[5] For a recent survey, see Christiane Berkvens-Stelinck, Hans Bots and Jens Häseler, *Les grands intermédiares culturels de la République des Lettres. Études de réseaux de correspondances du XVIe au XVIIIe siècles* (Paris: Honoré Champion, 2005).

[6] João Gama Oliveira and Albert-László Barabási, "Human dynamics: Darwin and Einstein correspondence patterns", *Nature*, 437 (27 october 2005),1251.

[7] Eugene Garfield, *Citation Indexing. Its Theory and Application in Science, Technology and Humanities* (New York : John Wiley & Sons, 1979), 81-97.

[8] Small, H.G., "Co-citation model of a scientific specialty – Longitudinal study of collagen research", *Social Studies of Science*, 7 (1977), 139-166; Gmür, M., "Co-citation analysis and the search for invisible colleges: A methodological evaluation", *Scientometrics*, 57 (2003), 27-57.

[9] Börner, K., Chen, C., Boyack, K.W., "Visualizing knowledge domains", *Annual Review of Information Science and Technology* 37 (2003), 179-255; Gingras, Y., "Revisiting the 'Quiet Debut' of the Double Helix: A Bibliometric and Methodological note on the ''Impact'' of Scientific Publications", *Journal of the History of Biology*, vol. 43, no 1, 2010, 159-181.

[10] Gingras, Y., "Mapping the Changing Centrality of Physicists (1900-1944)" in Daniel Torres-Salinas and Henk F. Moed (Eds) *Proceedings of the 11th Conference of the International Society for Scientometrics and Informetrics (ISSI)*, (Madrid: Spain, 2007), 314-320.

[11] *Correspondance de Marin Mersenne,* de Waard, Cornelis (Ed), 18 vols, (Paris : Presses universitaires de France & Éditions du CNRS, 1945-1988).

[12] For more details on bibliometric distributions, see Leo Egghe, *Power Laws in the Information Production Process : Lotkaian Infometrics* (Amsterdam : Elsevier Academic Press, 2005).

[13] *The Correspondence of Henry Oldenbourg*, A. Rupert Hall & Marie Boas Hall (Eds) (Madison : University of Wisconsin Press1955-1973,), vols 1-9; (London : Mansell, 1975-1977), vols 10-11; (London : Taylor & Francis, 1986), vols 12-13 and Additions and Corrections; on Oldenburg's life see M. Boas Hall, *Henry Oldenburg. Shaping the Royal Society* (Oxford : Oxford University Press, 2002).

[14] Pierre Brunet*, L'Introduction des théories de Newton en France au XVIIIe siècle. Avant 1738* (Paris: Albert Blanchard, 1931).

[15] *The Correspondence of Charles Darwin* 16 vols (Cambridge: Cambridge Univ. Press, 1984-2008). See also the web site of the Darwin project: www.darwinproject.ac.uk.

[16] We count one co-citation per letter irrespective of the fact that a given name may appear more than once in the letter. The number of citations or co-citations is thus equal or less than the total number of letters in the database. The same method applies to citation counts.
[17] We used the sofware Ucinet and Netdraw developped by Steve Borgatti and very much in use among researchers in social networks. Another much used similar program is Pajek. See S.P. Borgatti, *NetDraw: Graph Visualization Software* (Harvard: Analytic Technologies, 2002); S.P. Borgatti,. M.G. Everett, and L.C. Freeman, *Ucinet for Windows: Software for Social Network Analysis* (Harvard: Analytic Technologies, 2002). On Pajek, see V. Batagelj, A. Mrvar, "Pajek - Analysis and Visualization of Large Networks", in Jünger, M., Mutzel, P., (Eds.) *Graph Drawing Software* (Berlin : Springer, 2003), 77-103.
[18] L.C. Freeman, "Centrality in Social Networks. Conceptual Clarification", *Social Networks*, 1 (1978/1979), 215-239. The centrality of a given individual in the network is obtained by simply adding the number of links he/she has with all the others.
[19] For an application these techniques to a large set of scientific papers in history of physics see Wallace, Matthew L., Yves Gingras, Russell Duhon, "A new approach for detecting scientific specialties from raw co-citation networks", *Journal of the American Society of Information Science and Technology*, 60 (2009), 240-246.
[20] Where spelling of words is not standard one could also use a table of variants spellings.