Ureña-Carrion *et al. Applied Network Science* (2022) 7:26
https://doi.org/10.1007/s41109-022-00463-1

Applied Network Science

RESEARCH

Open Access

Communication now and then: analyzing the Republic of Letters as a communication network

Javier Ureña-Carrion[1*], Petri Leskinen[1], Jouni Tuominen[1,2], Charles van den Heuvel[3], Eero Hyvönen[1,2] and Mikko Kivelä[1]

*Correspondence: javier.urenacarrion@aalto.fi

[1] Department of Computer Science, Aalto University, Espoo, Finland
Full list of author information is available at the end of the article

Abstract
Huge advances in understanding patterns of human communication, and the underlying social networks where it takes place, have been made recently using massive automatically recorded data sets from digital communication, such as emails and phone calls. However, it is not clear to what extent these results on human behaviour are artefacts of contemporary communication technology and culture and if the fundamental patterns in communication have changed over history. This paper presents an analysis of historical epistolary metadata with the aim of comparing the underlying historical communication patterns with those of contemporary communication. Our work uses a new epistolary dataset containing metadata on over 150,000 letters sent between the 16th and 19th centuries. The analyses indicate striking resemblances between contemporary and epistolary communication network patterns, including dyadic interactions and ego-level behaviour. Certain aspects of the letter datasets are insufficient to corroborate other similarities or differences for these communication networks. Despite these drawbacks, our work helps confirm that several features of human communication are not artefacts of contemporary mediums or culture, but are likely elements of human behaviour.

Keywords: Communication networks, Granovetter effect, Social networks, Social signature

Introduction

New communication technologies and their digital traces have revolutionized the social sciences by providing new tools for understanding human behaviour (Lazer et al. 2020, 2021), and permitting to model society as a network of diverse and multifaceted relationships (Onnela et al. 2007a; Candia et al. 2008; Borgatti et al. 2009; Holme and Saramäki 2012; Miritello 2013). This has resulted in important insights regarding patterns of human communication (Miritello et al. 2013a, b) and the structure of social networks (Barabási et al. 2002; Onnela et al. 2007a). However, these analyses have been mainly described with data from contemporary human societies such as mobile phone calls, emails and online platforms. Although these phenomena have been observed in a myriad of contemporary communication media, the extent to which these phenomena are a product of current technologies and practices, or whether they are also present in other historical contexts and thus a characteristic of human behaviour is unclear as this question has mostly been analysed from theoretical perspectives (Innis 1951; McLuhan 1988). In this paper, we study a large historical epistolary dataset under a behavioural framework, and investigate whether letters—a major communication channel during the previous centuries—reflect patterns of human behaviour known to be present in contemporary communication.

Much of the progress in understanding the communication patterns and social behaviour has been made by conceptualising them as patterns in networks—a way of modelling social systems where individuals (or nodes) interact with each other (via links or edges) (Rombach et al. 2013)—. Communication metadata has provided crucial access for understating social networks as it allows us to construct exchange networks as proxies for the underlying social system that generated them. Figure 1 depicts an example of how interactions can be represented as networks: from a history of communication between two people or nodes, we create links or edges if there has been an interaction, and analyze both reconstructed networks and the time series of interactions.

This approach has confirmed several prominent sociological theories. Milgram's notion of "six degrees of separation" (Milgram 1967), where any two people are expected to be at a path of at most six people away from each other, was popularised in the 60s but only later studied with large-scale data, and is now known as small-world phenomena (Karsai et al. 2011). Similarly, in the 70s, Mark Granovetter hypothesised the "strength of weak ties", which states that the rarely-used weak social links keep networks connected by bridging separate social groups. This was later shown to hold at the level of a country using a communication network of millions of mobile phone users (Onnela et al. 2007a, b).

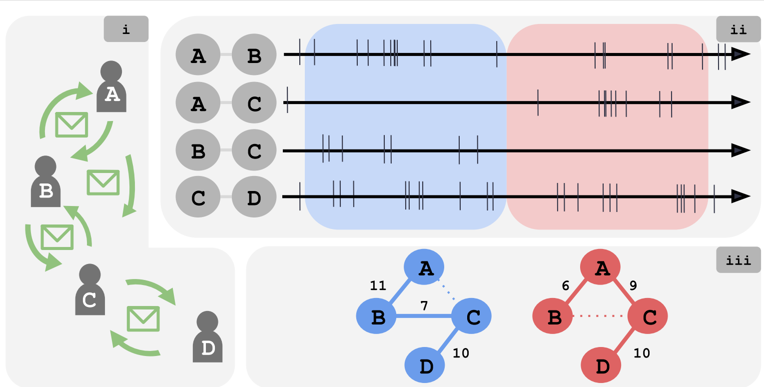

Fig. 1 Network reconstruction from recorded interactions. **(i)** In a simple social system, individuals A–D contact each other through directed letters, emails, phone calls, etc; **(ii)** we model undirected interactions as a time series of contact events, reconstructed from recorded contact metadata (sender, receiver, timestamp). Datasets are constrained to observation periods or snapshots (blue and red). **(iii)** Given an observation period, we can reconstruct a static network by creating links between nodes A–D if we observe at least one contact (solid lines). We do not account for a link if we do not observe data during the observation period (dashed lines). A common technique is to assign a link weight *w* (numbers) based on the number of contacts

These new approaches have also produced completely new theories of social behaviour and the system-level phenomena that emerges from it. One of the first key findings was large degree heterogeneity. That is, the number of neighbors varies considerably, where a majority of people have little connections yet a small number of individuals are highly connected. This has major implications on, e.g., disease and information spreading (Kivelä et al. 2012). The number of connections is of course limited as there are temporal, spatial and cognitive constraints into maintaining connections, which has lead to common strategies that people follow in managing their communication in the contemporary settings (Saramäki et al. 2014; Heydari et al. 2018; Miritello et al. 2013b). In this sense, people have robust social signatures: they tend to devote a similar amount of attention to their active circles of contacts, even if the circle of contacts changes (Saramäki et al. 2014; Heydari et al. 2018).

Large-scale data has also facilitated the analysis of different dimensions of temporality in human behaviour, including sequential events (Barabási 2005; Karsai et al. 2012a) and daily or regular cycles (Aledavood et al. 2015). Event sequences are often not regular or uniformly random, but the contacts are organised in bursts (Barabási 2005; Karsai et al. 2012a). This tells us about the possible processes of how the communication originates from the individuals, such as internal models of priority queues. Despite the small-world phenomena, information does not circulate as freely as burstiness can hinder information flow (Karsai et al. 2012a; Barabási 2005; Karsai et al. 2011).

The idea that communication mediums have an effect on societal structures and practices has been actively studied from historiographical perspectives. Harold Innis in ”The Bias of Communication” (Innis 1951) explored the notion that each historical epoch is distinguished by dominant forms of media consonant with the institutional power. Inspired by him, Marshall McLuhan stressed the impact of electronic technology on human behaviour and social organization (McLuhan 1988). Social theories that have informed social network analysis have also been used in historical contexts. Lux and Cook (1998) used the concept of weak ties in scientific exchange in the seventeenth century by contrasting the role of closed circles of academies in England and France and open networks of interactions that facilitated, e.g., the work of Dutch scientists. The small-world phenomena has also been studied on historical networks (Langmead et al. 2016) in the ”Six Degrees of Francis Bacon” project (Warren et al. 2016).

While letters are not automatically collected into large databases similar to some forms of digital communication, recently there have been large and concentrated efforts to collect and unify such large historical epistolary data sets distributed across different countries and collections. Metadata about the letters have been aggregated and provided for the research community through web services, such as Europeana (Foundation 2021), Kalliope (Library 2021), ePistolarium (History of the Netherlands 2021), the Mapping the Republic of Letters project (University 2021), correspSearch (Dumont et al. 2021) and the Early Modern Letters Online (EMLO) catalogue (Hotson et al. 2021) with a particular focus on the Republic of Letters (RofL) (van den Heuvel 2015; van Miert 2016; Hotson et al. 2019; Ravenek et al. 2017).

We contextualize epistolary communication by examining the RofL, a self-proclaimed community of scholars between the fifteenth and eighteenth centuries that consisted of overlapping networks. In these networks scholars communicated by letters and exchanged knowledge, news, and material objects such as books (van Miert 2016). This group is often described in idealistic terms as a self-regulating open community of scholars not limited by political borders or religious convictions. Nevertheless, these scholars could face persecution on account of religion and politics and often had to choose their words carefully due to the risk that their letters were intercepted, duplicated and even published (Daston 1991). Openness, confidentiality and secrecy co-existed and had an impact on the topological nature and development of these networks of scholars in which knowledge exchange was not merely reciprocal, but hierarchical as well (van den Heuvel 2015; van den Heuvel et al. 2016). These overlapping networks were in continuous flux. New members were often introduced by central gatekeepers who played a strong regulating and mediating role to keep these networks in balance, but that could also fall apart, resulting in structural holes (van Vugt 2019).

Some characteristics of epistolary communication might be analogous to their contemporary counterparts. The RofL was a community organized around specific social foci that included scholars and the clergy (Feld 1981). In current settings, such focused communication occurs in emails sent within organizations or in specialized internet forums. In contrast, a major difference between epistolary and contemporary channels is related to immediacy. In epistolary communication, the time between contacts had a considerable time gap constrained by exogenous factors such as distance and logistics, which contrasts with digital channels where waiting times depend mostly on users' behaviour (McLuhan 1988). Now, both historical and contemporary datasets have biases. In the former case, a major source of bias stems from the data-collection process itself (Ryan et al. 2020; Ryan and Ahnert 2021), with researchers slowly building an epistolary archive from different sources; in the latter, auto-recorded data is usually one of many communication channels and temporal windows are shorter.

Our goal is to investigate how several of the prominent results on contemporary communication hold for letters sent between 1510 and 1900. We analyse correspSearch (Dumont 2016), a large-scale epistolary data set representative of the RofL and compare it to four reference data sets on contemporary communication using emails, two digital platforms, and mobile phones. We compare historical and contemporary datasets under four analytical frameworks: (i) aggregate static networks, (ii) features of pairwise communication (iii) Granovetter effect, or the relationship between communication and local topology, and (iv) social signatures of ego networks. We find striking resemblance between historical and contemporary communication patterns, including in the distribution of dyadic temporal features, the robustness of social signatures, and certain network-wide characteristics. We note, however, that there are also considerable structural differences. For these negative results, we can't assure whether they are caused by social phenomena or the biases from the collection process of historical data.

Data

We analyze four datasets of contemporary human communication that encompass different channels and environments, including email data, mobile calls and online postings. Details of the datasets are available on Table 1. Our historical dataset, *Letters*, comes from the aggregated data service correspSearch (Dumont 2016) provided

Table 1 Description of the four reference datasets, including the observation span and source

Dataset	Content	Span	Source
Platform	Wall-posting on Facebook in New Orleans, where a directed link is created if a user posts on another user's wall. Data obtained from crawling public profiles in January 2009, each link tracks activity for one year after the first communication event. Total of ∼ 41K nodes, ∼ 183.5K edges and ∼ 854.6K wall posts.	1 year	Viswanath et al. (2009)
Email	Dataset of emails between employees of Enron Corporation, obtained as part of an investigation into Enron's collapse. Sampling might focus on 150 users. Total of ∼ 20K users, 297.5K edges and ∼ 1.1M emails.	6 years	Rossi and Ahmed (2015)
Forum	Activity on a movie forum, where a contact is created if a user comments on a thread posted by another user. Total of ∼ 6K users, ∼ 138.1K edges and ∼ 1.4M thread comments.	7 years	Karimi et al. (2014)
Mobile	Call Detail Records of mobile phone calls from an operator in a European country in 2007, where the operator had approximately 20% market share. We used a reduced version of the full dataset, where we focused on a region of the country. The small version of the dataset has ∼ 174K nodes, ∼ 190K edges and ∼ 6.8M phone calls.	4 months	Onnela et al. (2007a)

by the Berlin Brandenburg Academy of Sciences and Humanities. This dataset includes metadata about ca. 135,000 letters sent in 1510–1900.

All reference data is anonymized from source, and for the purpose of our research, we also use an anonymized version of *Letters*. The contemporary datasets provide fruitful insights into modern human communication. *Mobile* and *Platform* are both large user databases, allowing us to examine communication patterns at regional scales in both European and North American contexts. The sampling of both datasets is foci-independent in the sense that it does not follow specific societal units within their region. *Platform*, however, still represents a sample of people with public Facebook profiles. On the other hand, *Email* and *Forum* contain information from specific contexts (a company and an online community), and thus their interactions might follow certain setting-specific patterns. At least in the case of *Email*, we expect that the sampling methodology favors certain users.

Letters constitutes a unique sample of human communication encompassing five centuries of contact data across different countries and time periods. This presents both opportunities and challenges, as some of the common assumptions and techniques used for communication networks might not hold true. Notably, these datasets include letters from high-ranking personalities of religion, politics and science from several European countries, and constitutes a non-random sample from a communication network spanning five centuries.

The epistolary datasets have been transformed into Linked Data and published on the Linked Data Finland platform (Hyvönen et al. 2021, 2014) according to the Linked Data publishing principles and other best practices of W3C (Heath and Bizer 2011), including, e.g., content negotiation and provision of a SPARQL endpoint. However, the data are at the moment not publicly available on the Web due to copyright issues under negotiation with the owners of the primary metadata.

Results

Analysis of the static graph

We first model communication as aggregate static networks (Saramäki and Moro 2015), where the temporal dimension is projected out by disregarding the contact times. This approach allows us to determine whether there are any major structural similarities between the historical and contemporary networks.

We construct static networks from communication data following the commonly-used process described by Onnela et al. (2007a) and Saramäki and Moro (2015). We process data in the form of logs, which in its most basic form include information of a sender, a receiver and a timestamp. This holds both for the historical and modern reference datasets. We create an undirected edge if we have observed at least one letter between two nodes, but we keep track of the time series of interactions for further analysis, as well as basic statistics such as the number of contacts w.

As several of the static network features have been studied in historical correspondence datasets (Ryan et al. 2020; Ryan and Ahnert 2021; Ahnert and Ahnert 2015), we describe these results only briefly, and include a more detailed analysis on Additional file 1. We find that the epistolary network snapshots are similar to other communication networks only to some extent, with the main difference stemming from over-represented nodes.

The interpretation of static networks differs over large observation periods. For our correspondence data, dyads are constrained to the participants' active life, but the overall network can be thought of as the overlapping activity from dyads, where larger structures might not represent, e.g., communities, but temporally overlapped communities, which are more difficult to interpret. We ameliorate some of these issues by aggregating data over different 50-year periods, resulting in a series of static graphs.

We present a selection of common network statistics: centrality measures, tie overlap and the distribution of shortest paths. Centrality measures are informative of the relative importance and roles of a node in the network. In this sense, *degree* k measures the number of connections, while *betweenness* C_B is related to the number of shortest paths that pass through a node. Tie overlap O is link-level statistic used to assess whether an edge is contained within overlapping circles of friends, or serves as a "bridge" (see Granovetter effect), and the distribution of shortest paths l helps characterize the small-world phenomena.

On Fig. 2 we compare static network statistics for the correspondence and reference datasets. We find similarities in terms of the heavy-tailed degree distribution (Holme 2019), decaying centrality measures and average shortest-path distribution, with epistolary networks exhibiting small-world phenomena even at 50-year aggregation periods. Nevertheless, in correspondence networks some nodes have an outsized impact as compared to most important nodes of the reference networks. This is reflected in (i) slower decay for the degree distribution, where the tail reflects an upwards flattening curve in the log-log axis; (ii) overall larger centrality measures, e.g., some nodes score $C_B > 0.5$ betweenness, meaning that more than half of all shortest paths go through them as opposed to all other networks having $C_B < 0.08$. Notably, the overlap distribution might be the most affected from sampling, as it requires sampling dyads as opposed to nodes. *Mobile*, the data set with most individuals, displays the largest differences when

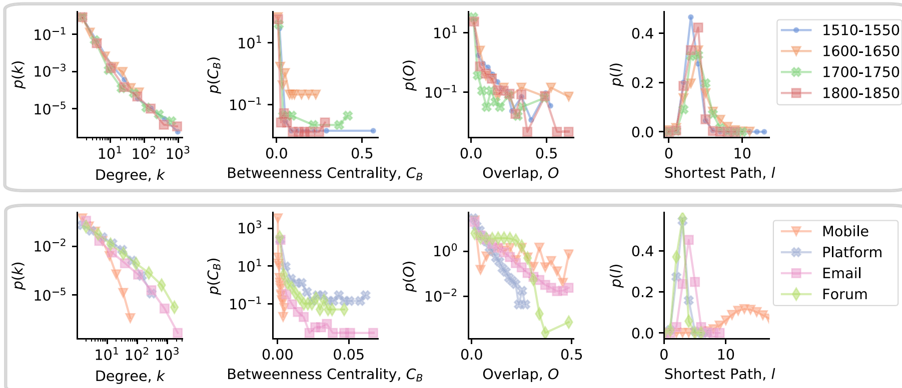

Fig. 2 Static features of communication networks, where the top (bottom) rows contains features of the correspondence (reference) networks. Epistolary datasets are aggregated in 50 year periods. Columns represent distributions of (left to right) degree *k*, betweenness centrality C_B, topological overlap *O* and average shortest path length *l*. For ease of comparison, we do not include individual data points, but present averaged binned values

compared to the other references, with fast degree and betweenness centrality decay, and longer shortest paths.

Features of human communication

We now explore *how* communication occurs between pairs of individuals, where the focus is on different tie-level temporal dynamics, and not only on the network structure.

We derive different statistics based on the time series of interactions, with some emphasis on the distribution of inter-event times (IETs). Figure 3(i) depicts a visual example of approaches to modelling communication sequences. For a sequence of interaction times $\{t_0, t_1, \ldots, t_n\}$ we measure temporal stability in terms of the first and last observed interactions, $TS = t_n - t_0$. We define the kth inter-event time $\tau_k = t_{k+1} - t_k$ for sequences of at least lenght three, and estimate moments from the IET distribution from the empirical observations $\{\tau_k\}$. We use the Kaplan–Meier estimator (Kivelä and Porter 2015), which considers *censored* data, as a means to correct for the lack of data in epistolary contexts. We thus obtain $\mu(\tau)$ (in the text μ) and τ_σ , the mean and standard deviation of our IETs. We use the burstiness coefficient as proposed by Goh and Barabási (2008), which is defined as $B = (\tau_\sigma - \tau)/(\tau_\sigma + \tau)$.

The total contact count w has been used in communication networks as a proxy for tie strength. Figure 3(ii–iii) shows statistics of the number of contacts w and temporal stability *TS*, respectively. Aggregate contacts w are highly inhomogeneous for all datasets, displaying heavy-tailed decay. Note that w directly impacts many temporal features. For this reason, we show distributions conditioned on the number of contacts w categorized by quartiles. Given the heterogeneity of w it is not straightforward to partition data into bins of equal size. In Table 2 we show the composition of quartiles for each dataset. Notably, only *Mobile* is partitioned into bins of roughly equal sizes (as percentage), for all other datasets, q_1 is disproportionately large.

On the other hand, *TS* displays major inter- and intra-dataset variability. Chiefly, we identify a mix of quartile-independent sharp decay (*Letters*, *Email*), heavy-tailed

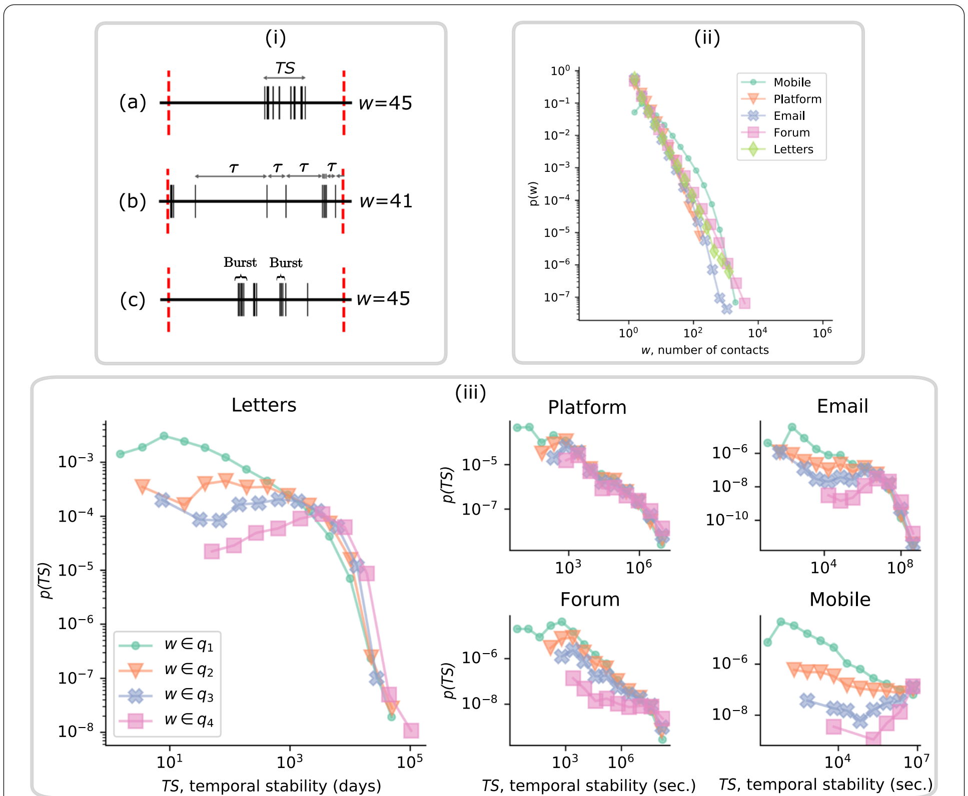

Fig. 3 Temporal features of human communication. (**i**) Visual representation of different temporal features for similar number of contacts, *w*. (**i**, a) Temporal stability *TS*, where we only consider the first and last times during the observation period. (**i**, b) Estimation of the inter-event time (IET) distribution τ, the elapsed time between consecutive events. (**i**, c) Example of bursty behaviour, where bursts of consecutive calls are followed by longer IETs. The burstiness coefficient is defined as $B = \frac{\sigma-\mu}{\sigma+\mu}$ for μ, σ the mean and standard deviation of τ. (**ii**) Empirical distribution of number of contacts *w* for the epistolary dataset and four reference datasets. (**iii**) Empirical distribution of temporal stability *TS* for the epistolary and reference datasets. For each case, we plot the distributions conditioned on link weight quartiles. For ease of comparison, we do not include individual data points, but present averaged binned values

Table 2 Quartile composition for each dataset

Dataset	%%$w \leq 2$	q_1(%)	q_2(%)	q_3(%)	q_4(%)
Letters	70.9	3–4 (47.7)	5–7 (16.2)	8–16 (17)	> 16 (19)
Platform	57.8	3 (42.5)	4–5 (20.8)	6–9 (16.9)	> 9 (20)
Email	73.4	3 (43.4)	4–5 (18.2)	6–10 (18.5)	> 10 (19.8)
Forum	64.9	3 (35.8)	4–5 (15.4)	6–14 (24.3)	14 (24.5)
Mobile	14.6	3–5 (28)	6–13 (23.5)	14–39 (23.7)	> 40 (24.8)

The first numerical column contains the fraction of links that have either one or two contacts, and which were thus not used to compute temporal features. The latter columns contain the composition of each quartile, including their ranges of contacts and in brackets the percentage of links in that category

decay (*Platform*, *Forum*) or quartile-dependent growth and decay (*Letters*, *Email*, *Forum* and *Mobile*). These different patterns might arise due to an interplay of the observation period and tie decay. In *Mobile* we do not expect to observe tie decay since the observation window is short (Miritello 2013; Navarro et al. 2017), and thus the *TS* differs conditioned on w quartiles. *Email* and *Letters*, on the other hand, display a mixture of the latter behaviour and sharp decay, constrained by people's active life (within a company, in the case of *Email*).

The IETs, or distribution of times between consecutive contacts among two individuals, are an important feature of human communication that has undergone extensive scrutiny in the contemporary communication patterns (Goh and Barabási 2008; Karsai et al. 2011, 2012a, b; Malmgren et al. 2008). In epistolary data, IETs have been analysed for only a handful of exceptionally prominent individuals, and these finding indicate similarities between contemporary and correspondence data (Oliveira and Barabási 2005; Malmgren et al. 2009). We expand on this work by focusing on large-scale behaviour and in other statistics. Just as with the degree distribution, we find that IETs tend to be very heavy-tailed and bursty (Goh and Barabási 2008; Karsai et al. 2011). This means that the time between events shows high variability, with long resting times followed by short *bursts* of consecutive contacts.

On Fig. 4 we focus on two statistics of the IET distribution—mean μ and burstiness B—. We uncouple μ from the number of contacts by scaling the distribution by μ_{w_i}, the mean

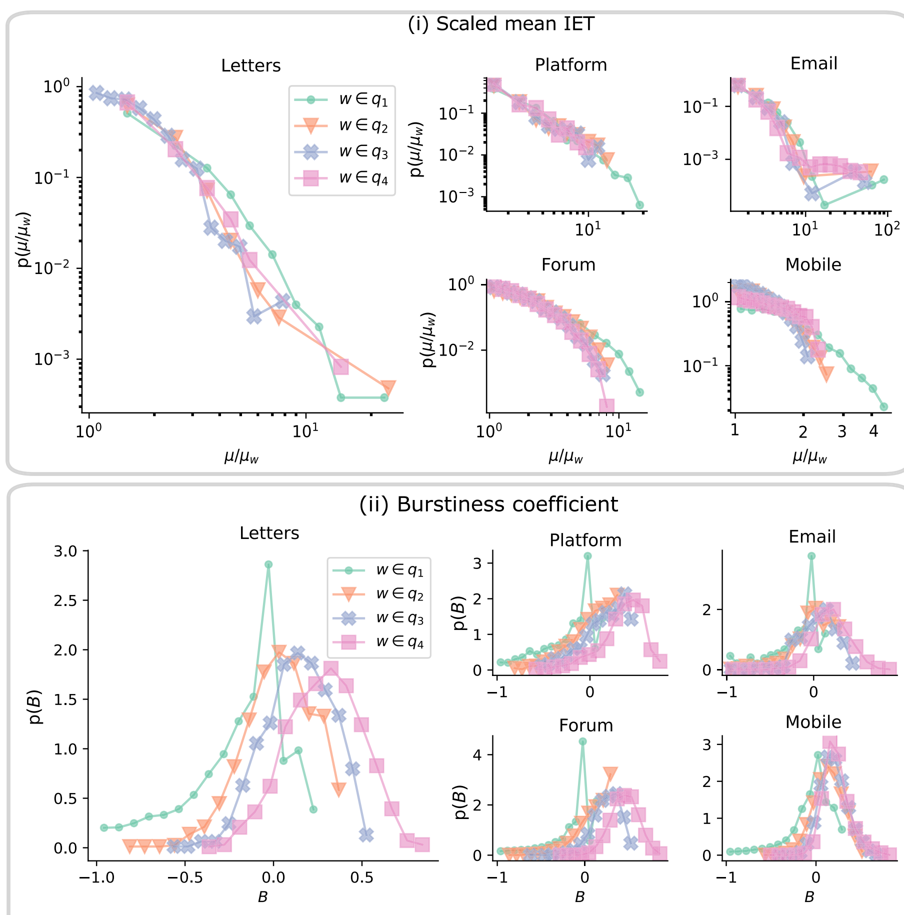

Fig. 4 Distribution of different values derived from the IET distribution for correspSearch and our four reference datasets. The scale is days for correspSearch, and seconds for the rest of the datasets. (**i**) Scaled mean IET μ/μ_w, (**ii**) Burstiness coefficient B

distribution value for the quartile category i. Previous research has shown that the distributions of μ/μ_{w_i} collapse onto a single distribution for all different w_i (Goh and Barabási 2008; Karsai et al. 2011; Saramäki and Moro 2015). We replicate this result for most of our reference datasets and for *Letters*, suggesting that the IET distribution for epistolary data follows a pattern observed before for multiple large contemporary communication data sets (Goh and Barabási 2008; Karsai et al. 2011; Candia et al. 2008), and for the set of prominent individuals in historical correspondence (Oliveira and Barabási 2005; Malmgren et al. 2009). We would expect people in close contact to have short average IETs (Ureña-Carrion et al. 2020); nevertheless, given the large timespan of the epistolary data, the closeness of contacts could change throughout a person's lifetime.

All datasets display larger burstiness values for larger w, with the epistolary data following similar average values to the *Platform* and *Email* datasets (B around -0.19, 0.02, 0.14 for the first three quartiles of *Letters*, *Platform* and *Email*; the last quartiles are 0.28 for *Letters*, 0.34 for *Platform* and 0.20 for *Email*). Although burstiness has been explained by behavioural patterns, such as mentally prioritising tasks (Barabási 2005) or cyclic activity periods (Malmgren et al. 2008), some of the bursty character of epistolary data could also be explained by the fact that news sharing and copying played an important role in scholarly correspondences (Colavizza 2014; Dooley 2010).

Granovetter effect

We now examine the Granovetter effect, a theory that establishes a connection between network topology and the qualitative relationship between two people, stating that strong links are surrounded by overlapping circles of friends, whereas weak links serve as bridges between communities (Granovetter 1973). Theoretically, Granovetter characterized four dimensions of tie strength: time, emotional intensity, intimacy and reciprocity. In practical terms, however, these dimensions are not straightforward to observe from interaction data, both because of incompleteness and a lack of a natural metric or measure (Zeller et al. 1982). In the Granovetter sense, however, the proxies for tie strength inform of topological roles of ties. This phenomena has been empirically observed by measuring the strength of ties via the number of contacts w, although it can also be observed for several communication features that are uncoupled from w (Ureña-Carrion et al. 2020).

In empirical social network analysis, the embeddedness of a tie can be measured via topological overlap (Onnela et al. 2007a; Ureña-Carrion et al. 2020). For i, j two nodes, and $\mathcal{N}(i)$ the set of neighbors of node i, overlap is the ratio of common neighbors over all neighbors, or the Jaccard similarity between the sets of neighbors.

$$O_{ij} = \frac{|\mathcal{N}(i) \cap \mathcal{N}(j)|}{|\mathcal{N}(i) \cup \mathcal{N}(j)|}$$

We explore the Granovetter effect on Fig. 5 by comparing overlap with w, TS and μ. We find an association between some of the features and overlap, with a visible effect for all datasets using w. This indicates that correspondence data exhibits a salient feature of social networks, where user behaviour tends to be found within local community structures. Notably, for TS we did not find a clear association for the epistolary data, as well as for μ in the period 1775–1825. This might be related to sampling biases in the network and in the time-series. In the first case, overlap is a measure based on triads, which tend

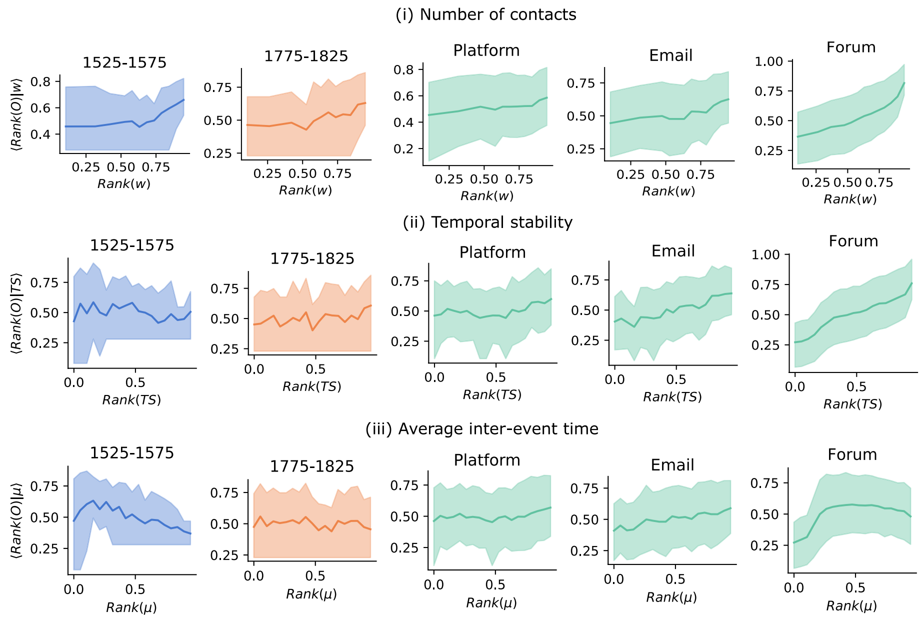

Fig. 5 Granovetter effect for different features of communication on (*left* to *right*) two snapshots of Letters (1525–1575) and reference datasets Platform, Email and Forum (the effect has already been described in *Mobile* by Ureña-Carrion et al. 2020). The embeddedness of ties, as measured by topological overlap, changes in association with behavioural communication features, although not always in a linear fashion. Shaded area represents 80% of the distribution. Features of communication: (*i*) number of contacts w, (*ii*) temporal stability TS and (*iii*) Mean IET μ. The Granovetter effect is most visible for w. Using TS as a proxy for tie strength we do not observe the effect for Letters and we only observe it partly for Platform. For μ, the effect (decrease of overlap with larger average IET) is visible for the period 1525–1575 and partially for larger values of Forum

to be less likely to be sampled. Second, tie-level dynamics might be affected by irregular sampling at different time periods. This could explain why we observe the effect for μ in 1525–1575 but not in 1775–1825.

Social signatures

Ego networks, where the emphasis is on central nodes (egos) and their immediate neighbors (alters), serve as useful characterizations of user behaviour. We focus on social signatures, the idea that people divide their attention to their high-ranking contacts in a persistent fashion across time and uniquely to each ego (Saramäki et al. 2014; Heydari et al. 2018). This does not mean that the alters don't change —they are highly dynamic—, but the attention devoted to high-ranking alters is persistent. We inspect the ego networks of different people at consecutive snapshots, where on each snapshot we create an ego network with the number of outgoing contacts as link weights, where we focus only on outgoing contacts to characterize the behaviour of the ego. Afterwards, we build the social signature by ranking the alters and obtaining the fraction of outgoing contacts.

In the seminal work where social signatures were observed for the first time, Saramäki et al. (2014) divided the observation period into three snapshots, thus obtaining three social signatures for each individual. They compared social signatures for different intervals using the Jensen-Shannon Divergence (Signature Distance d), a measure for comparing probability distributions, and for reference they also compared social signatures to other individuals. To adapt their analysis to our dataset, we incorporate two filters regarding the number of alters during an interval, and the length of the interval. We determine social signatures on a interval of size b, measured in years for *Letters*, *Email* and *Forum*, and in months for *Platform* and *Mobile*. Following Saramäki et al. (2014), we compare social signatures for a given individual in $m = 3$ adjacent time intervals, and we compare signatures of other individuals during time intervals of the same length. We determine a limit on the minimum number of out-neighbors n_{o-ng} (alters contacted by the ego) that an ego must have during all intervals in order to be a valid social signature, and we vary this parameter for a range of values.

Figure 6 depicts the results for our analysis of social signatures for filter values $n_{o-ng} = 5$, and different bin sizes b depending on the dataset; we were, however, able to replicate the results of Saramäki et al. (2014), Heydari et al. (2018) for all datasets and for all filter values in *Letters* (see Additional file 1). This suggests that individualized patterns of communication that are consistent are not a feature of contemporary communication networks, but were also persistent in older and less immediate forms of communication as well. Even more, our results hold for observation windows ranging from 1 to 10 years, and totaling comparisons ranging 3–30 years considering three observation windows. This indicates that these individual patterns can be observed not only in shorter timescales, but also in periods spanning a large proportion of an adult life.

Discussion

We used common methods for analyzing social networks and applied them to a large epistolary dataset in an attempt to uncover whether major differences exist between contemporary and historical communication networks. For reference, we used four contemporary datasets of communication data. Our results strongly suggest that prominent characteristics of contemporary communication networks are also found in historical

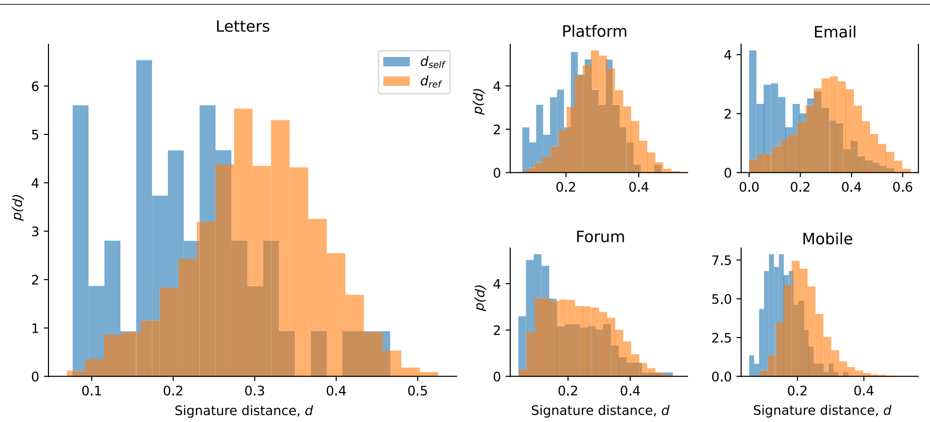

Fig. 6 Comparison between self-distances d_{self} and reference distances d_{ref} for ego networks of all datasets. The average self distances are smaller than average reference distances for all parameter combinations, implying that social signatures tend to be more similar for the same person during different times, than to other people's signatures. We performed a two-sided T-test for difference in means, finding that for all datasets the mean d_{self} differed from d_{ref} with all estimated p values $< 10^{-13}$. The estimated means for all datasets are: Letters $\hat{d}_{self} = 0.218 \pm 0.048$ and $\hat{d}_{ref} = 0.302 \pm 0.008$; Platform $\hat{d}_{self} = 0.241 \pm 0.025$ and $\hat{d}_{ref} = 0.292 \pm 0.002$; Email $\hat{d}_{self} = 0.191 \pm 0.019$ and $\hat{d}_{ref} = 0.317 \pm 0.0001$; Forum $\hat{d}_{self} = 0.189 \pm 0.019$ and $\hat{d}_{ref} = 0.235 \pm 0.001$; Mobile $\hat{d}_{self} = 0.162 \pm 0.014$ and $\hat{d}_{ref} = 0.227 \pm 0.002$. Confidence intervals for the mean reported at 95%

contexts, validating that these characteristics are related to fundamental human behavior, not only to specific communication mediums and practices. Despite these major positive results, we also found differences in the contemporary and historical communication patterns. We might attribute these negative results to differences in the completeness of the datasets, since we depend on letters that have survived and been included in collections. We note, however, that they might also stem from differences in the use of media and communication practices.

We divided our analyses in four main parts: global network structure and characteristics, dyadic communication patterns, their relationship to local topology, and analysis of ego networks. Regarding global structures, we found the epistolary datasets to be both highly-local and deeply connected—in the sense that most nodes have a very small number of connections, while still maintaining relatively short average shortest path distributions. However, we also found that a majority of nodes are under-represented in the network, having less-than-expected connections. Indeed, since epistolary datasets tend to focus on relevant historical figures, we centre the rest of our analysis on dyadic interactions and ego networks.

Dyadic interactions display major similarities both in terms of the number of contacts and in the distribution of inter-event times, even when considering that the latter follows naturally different scales. Our results suggest that major features of dyadic communication can be described with known mechanisms of average IET decay, and that the overall burstiness distribution is in line with contemporary communication channels. This is slightly striking, as one could expect that limitations in using letters, such as less immediacy, could imply greater regularity in communication. Further analyses, however, could reveal how burstiness varies across several years.

We also found some evidence that epistolary networks also reflect the Granovetter effect, where stronger ties are contained within overlapping circles of friends, while weaker serve as bridges. We found striking similarities in the characterization of social signatures, possibly due to more control in filtering more complete samples of ego networks. In this case, we found that individuals in previous centuries communicated in a similar manner than we do today: uniquely to individuals and persistent in time.

Abbreviations

EMLO:: Early Modern Letters Online; IET:: Inter-event time; RofL:: Republic of Letters.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s41109-022-00463-1.

Additional file 1. A PDF file with further analyses and tests. There are two main sections: first, we further explore and analyze different aspects of the static graph approach; second, we explore different parameter ranges for the construction of social signatures.

Acknowledgements

We thank Dirk van Miert, and Sebastiaan Derks for discussions and the possiblity of using ePistolarium data in our research. Thanks to Sebastian Dumont for providing the correspSearch data and discussions. Our work is continuation of the collaborations in the EU COST action Reassembling the Republic of Letters and with University of Oxford in the Cultures of Knowledge project and the Early Modern Letters Online service.

Author Contributions

J.U.C., M.K. and E.H. conceived the main ideas. J.U.C. and M.K. designed the analysis, and J.U.C. performed the analysis. P.L., J.T. and J.U.C. collected the data and contributed with analytic tools. C.V.D.H. contextualized historical data and designed analysis. J.U.C. took the lead in writing the manuscript, to which all authors contributed. All authors read and approved the final manuscript.

Funding
PL, JT, and EH acknowledge funding from the Academy of Finland and the EU project In/Tangible European Heritage (InTaVia). MK acknowledges the Academy of Finland Grant No. 320781, ECANET.

Availability of data and materials
All datasets except Mobile are available online. Mobile is not publicly available due to a signed non-disclosure agreement, as it contains sensitive information of the subscribers. The epistolary dataset is being integrated in the LetterSampo system (Hyvönen et al. 2020; Leskinen et al. 2022) underway for publishing and using epistolary linked data on the Semantic Web, a new member in the Sampo series of Linked Open Data services and semantic portals (Hyvönen et al. 2021), for an academic reference see Hyvönen (2020). A selection of the networks analytic tools discussed in this paper are being integrated in the LetterSampo system. Code is available at Ureña-Carrion and Leskinen (2021).

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Computer Science, Aalto University, Espoo, Finland. [2]HELDIG – Helsinki Centre for Digital Humanities, University of Helsinki, Helsinki, Finland. [3]Huygens Institute for the History of the Netherlands, Amsterdam, The Netherlands.

Received: 3 November 2021 Accepted: 17 April 2022
Published online: 10 May 2022

References

Ahnert R, Ahnert SE (2015) Protestant letter networks in the reign of Mary i: a quantitative approach. ELH 82(1):1–1. https://doi.org/10.1353/elh.2015.0000
Aledavood T, Lehmann S, Saramäki J (2015) Digital daily cycles of individuals. Front Phys 3:73. https://doi.org/10.3389/fphy.2015.00073
Barabási A-L (2005) The origin of bursts and heavy tails in human dynamics. Nature 435(7039):207–211. https://doi.org/10.1038/nature03459
Barabási AL, Jeong H, Néda Z, Ravasz E, Schubert A, Vicsek T (2002) Evolution of the social network of scientific collaborations. Phys A 311(3–4):590–614. https://doi.org/10.1016/s0378-4371(02)00736-7
Borgatti SP, Mehra A, Brass DJ, Labianca G (2009) Network analysis in the social sciences. Science 323(5916):892–895. https://doi.org/10.1126/science.1165821
Candia J, González MC, Wang P, Schoenharl T, Madey G, Barabási A-L (2008) Uncovering individual and collective human dynamics from mobile phone records. J Phys A Math Theor 41(22):224015. https://doi.org/10.1088/1751-8113/41/22/224015
Colavizza G (2014) Mapping early modern news networks: a digital humanities approach. Master's thesis, Università Ca' Foscari Venezia. https://doi.org/10.13140/RG.2.2.26344.88326.http://hdl.handle.net/10579/4893
Daston L (1991) The ideal and reality of the Republic of Letters in the enlightenment. Sci Context 4(2):367–386. https://doi.org/10.1017/S0269889700001010
Dooley B (ed.) (2010) The dissemination of news and the emergence of contemporaneity in early modern Europe. https://doi.org/10.4324/9781315240244
Dumont S (2016) correspSearch—connecting scholarly Editions of Letters. J Text Encod Initiat. https://doi.org/10.4000/jtei.1742
Dumont S, Grabsch S, Müller-Laackman J (2021) correspSearch–Connect Scholarly Editions of Correspondence (2.0.0). Berlin-Brandenburg Academy of Sciences and Humanities. Accessed 22 Oct 2021. https://correspSearch.net
Feld SL (1981) The focused organization of social ties. Am J Sociol 86(5):1015–1035. https://doi.org/10.1086/227352
for the History of the Netherlands HI (2021) ePistolarium. Accessed 12 Aug 2021. http://ckcc.huygens.knaw.nl/epistolarium/
Foundation E (2021) Europeana. Accessed 23 Jul 2021. https://www.europeana.eu
Goh K-I, Barabási A-L (2008) Burstiness and memory in complex systems. EPL (Europhysics Letters) 81(4):48002. https://doi.org/10.1209/0295-5075/81/48002
Granovetter MS (1973) The strength of weak ties. Am J Sociol 78(6):1360–1380. https://doi.org/10.1086/225469
Heath T, Bizer C (2011) Linked data: evolving the web into a global data space. Theory and technology, synthesis lectures on the semantic web
Heydari S, Roberts SG, Dunbar RIM, Saramäki J (2018) Multichannel social signatures and persistent features of ego networks. Appl Netw Sci. https://doi.org/10.1007/s41109-018-0065-4
Holme P (2019) Rare and everywhere: perspectives on scale-free networks. Nat Commun. https://doi.org/10.1038/s41467-019-09038-8
Holme P, Saramäki J (2012) Temporal networks. Phys Rep 519(3):97–125. https://doi.org/10.1016/j.physrep.2012.03.001
Hotson H, Lewis M (2021) Early modern letters online. Accessed 7 Aug 2021. http://emlo.bodleian.ox.ac.uk

Hotson H, Wallnig T (eds) (2019) Reassembling the Republic of Letters in the Digital Age. https://doi.org/10.17875/gup2019-1146
Hyvönen E "Sampo" model and semantic portals for digital humanities on the semantic web. In: DHN 2020 Digital Humanities in the Nordic Countries. In: Proceedings of the digital humanities in the Nordic Countries 5th conference, pp 373–378 (2020). http://ceur-ws.org/Vol-2612/poster1.pdf
Hyvönen E, Leskinen P, Tuominen J (2020) From the Republic of Letters to LetterSampo—Historical Letters on the Semantic Web. Aalto University. White paper, Aalto University, Semantic Computing Research Group (SeCo): https://seco.cs.aalto.fi/publications/2020/hyvonen-et-al-lettersampo-2020.pdf
Hyvönen E, Tuominen J, Alonen M, Mäkelä E (2021a) Linked data Finland. Accessed 4 Sept 2021. https://www.ldf.fi/
Hyvönen E, Leskinen P, Tuominen J (2021b) Sampo Model and Series of Semantic Portals. Accessed 1 Nov 2021. https://seco.cs.aalto.fi/applications/sampo/
Hyvönen E, Tuominen J, Alonen M, Mäkelä E (2014) Linked Data Finland: a 7-star model and platform for publishing and re-using linked datasets. In: The semantic web: ESWC 2014 Satellite Events, Revised Selected Papers, pp 226–230
Innis H (1951) The Bias of communication. Toronto: University of Toronto Press
Karimi F, Ramenzoni VC, Holme P (2014) Structural differences between open and direct communication in an online community. Phys A 414:263–273. https://doi.org/10.1016/j.physa.2014.07.037
Karsai M, Kivelä M, Pan RK, Kaski K, Kertész J, Barabási A-L, Saramäki J (2011) Small but slow world: how network topology and burstiness slow down spreading. Phys Rev E. https://doi.org/10.1103/physreve.83.025102
Karsai M, Kaski K, Barabási A-L, Kertész J (2012a) Universal features of correlated Bursty behaviour. Sci Rep. https://doi.org/10.1038/srep00397
Karsai M, Kaski K, Kertész J (2012b) Correlated dynamics in egocentric communication networks. PLoS ONE 7(7):40612. https://doi.org/10.1371/journal.pone.0040612
Kivelä M, Porter MA (2015) Estimating inter-event time distributions from finite observation periods in communication networks. Phys Rev E. https://doi.org/10.1103/physreve.92.052813
Kivelä M, Pan RK, Kaski K, Kertész J, Saramäki J, Karsai M (2012) Multiscale analysis of spreading in a large communication network. J Stat Mech Theory Exp 03:03005. https://doi.org/10.1088/1742-5468/2012/03/p03005
Langmead A, Otis J, Warren C, Weingart S, Zilinski L (2016) Towards interoperable network ontologies for the digital humanities. Int J Humanit Arts Comput. https://doi.org/10.3366/ijhac.2016.0157
Lazer DM, Pentland A, Watts DJ, Aral S, Athey S, Contractor N, Freelon D, Gonzalez-Bailon S, King G, Margetts H et al (2020) Computational social science: obstacles and opportunities. Science 369(6507):1060–1062
Lazer D, Hargittai E, Freelon D, Gonzalez-Bailon S, Munger K, Ognyanova K, Radford J (2021) Meaningful measures of human society in the twenty-first century. Nature 1–8
Leskinen P, Ureña-Carrion J, Tuominen J, Kivelä M, Hyvönen E (2022) Applying network science for analyzing historical communication networks on the semantic web--A linked data approach, knowledge graph and data service. Aalto University. White paper, Aalto University, Semantic Computing Research Group (SeCo). https://seco.cs.aalto.fi/publications/2022/leskinen-et-al-lettersampo-2022.pdf
Library, P.C.H.F.B.S. (2021) Kalliope-Verbund. Accessed 23 Jul 2021. https://www.kalliope-verbund.info
Lux DS, Cook HJ (1998) Closed circles or open networks? Communicating at a distance during the scientific revolution. Hist Sci 36(2):179–211. https://doi.org/10.1177/007327539803600203
Malmgren RD, Stouffer DB, Motter AE, Amaral LA (2008) A poissonian explanation for heavy tails in e-mail communication. Proc Natl Acad Sci 105(47):18153–18158
Malmgren RD, Stouffer DB, Campanharo ASLO, Amaral LAN (2009) On universality in human correspondence activity. Science 325(5948):1696–1700. https://doi.org/10.1126/science.1174562
McLuhan M (1988) The role of new media in social change. In: Sanderson G, McLuhan M, MacDonald F (eds) The man and his message, pp 34–35
Milgram S (1967) The small-world problem. Psychol Today 1(1)
Miritello G (2013). Temporal patterns of communication in social networks. https://doi.org/10.1007/978-3-319-00110-4
Miritello G, Lara R, Cebrian M, Moro E (2013a) Limited communication capacity unveils strategies for human interaction. Sci Rep. https://doi.org/10.1038/srep01950
Miritello G, Lara R, Moro E (2013b) Time allocation in social networks: correlation between social structure and human communication dynamics. In: Understanding complex systems, pp 175–190. https://doi.org/10.1007/978-3-642-36461-7_9
Navarro H, Miritello G, Canales A, Moro E (2017) Temporal patterns behind the strength of persistent ties. EPJ Data Sci. https://doi.org/10.1140/epjds/s13688-017-0127-3
Oliveira JG, Barabási A-L (2005) Darwin and Einstein correspondence patterns. Nature 437(7063):1251–1251. https://doi.org/10.1038/4371251a
Onnela J-P, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, Kertész J, Barabási A-L (2007a) Structure and tie strengths in mobile communication networks. Proc Natl Acad Sci 104(18):7332–7336. https://doi.org/10.1073/pnas.0610245104
Onnela J-P, Saramäki J, Hyvönen J, Szabó G, Argollo de Menezes M, Kaski K, Barabási A-L, Kertész J (2007b) Analysis of a large-scale weighted network of one-to-one human communications. New J Phys 9:179
Ravenek W, van den Heuvel CMJM, Gerritsen GH (2017) The ePistolarium: origins and techniques, pp. 309–316. Ubiquity Press Limited, London. https://doi.org/10.5334/bbi.26
Rombach MP, Porter MA, Fowler JH, Mucha PJ (2013) Core-periphery structure in networks. SSRN Electron J. https://doi.org/10.2139/ssrn.2002684
Rossi RA, Ahmed NK (2015) The network data repository with interactive graph analytics and visualization. In: AAAI. http://networkrepository.com
Ryan YC, Ahnert SE (2021) The measure of the archive: the robustness of network analysis in early modern correspondence. J Cult Anal. https://doi.org/10.22148/001c.25943
Ryan Y, Ahnert S, Ahnert R (2020) Networking archives: quantitative history and the contingent archive. In: Proceedings of the workshop on computational humanities research (CHR 2020), vol 2723

Saramäki J, Moro E (2015) From seconds to months: an overview of multi-scale dynamics of mobile telephone calls. Eur Phys J B. https://doi.org/10.1140/epjb/e2015-60106-6
Saramäki J, Leicht EA, Lopez E, Roberts SGB, Reed-Tsochas F, Dunbar RIM (2014) Persistence of social signatures in human communication. Proc Natl Acad Sci 111(3):942–947. https://doi.org/10.1073/pnas.1308540110
University S (2021) Mapping the Republic of Letters. Accessed 3 Sept 2021. http://republicofletters.stanford.edu/
Ureña-Carrion J, Saramäki J, Kivelä M (2020) Estimating tie strength in social networks using temporal communication data. EPJ Data Sci. https://doi.org/10.1140/epjds/s13688-020-00256-5
Ureña-Carrion J, Leskinen P (2021) Communication Patterns of the Republic of Letters. Accessed 1 Nov 2021. https://version.aalto.fi/gitlab/urenaj1/rofl_communication_patterns
van den Heuvel C (2015) Mapping knowledge exchange in Early Modern Europe: intellectual and technological geographies and network representations. Int J Humanit Arts Comput 9(1):95–114. https://doi.org/10.3366/ijhac.2015.0140
van den Heuvel C, Weingart SB, Spelt N, Nellen H (2016) Circles of confidence in correspondence. Nuncius 31(1):78–106. https://doi.org/10.1163/18253911-03101002
van Miert D (2016) What was the Republic of Letters? A brief introduction to a long history (1417–2008). Groniek 204(205):269–287
van Vugt I (2019) The structure and dynamics of scholarly networks between the Dutch Republic and the grand duchy of Tuscany in the 17th century. https://hdl.handle.net/11245.1/94502a28-e642-4ecc-81e2-100fda93ecba
Viswanath B, Mislove A, Cha M, Gummadi KP (2009) On the evolution of user interaction in Facebook. In: Proceedings of the 2nd ACM workshop on online social networks—WOSN 09. https://doi.org/10.1145/1592665.1592675
Warren CN, Shore D, Otis J, Wang L, Finegold M, Shalizi C (2016) Six Degrees of Francis Bacon: a statistical method for reconstructing large historical social networks. Dig Humanit Q 10(3)
Zeller RA, Nock SL, Carmines EG (1982) Measurement in the social sciences: the link between theory and data. Contemp Sociol 11(1):79. https://doi.org/10.2307/2066656

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.