Ureña-Carrion, Javier; Leskinen, Petri; Tuominen, Jouni; van den Heuvel, Charles; Hyvönen, Eero; Kivelä, Mikko

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

Published in: Applied network science

DOI: 10.1007/s41109-022-00463-1

Published: 01/12/2022

Document Version
Publisher's PDF, also known as Version of record

Published under the following license: CC BY

Please cite the original version:
Communication now and then: analyzing the Republic of Letters as a communication network

Javier Ureña-Carrion1*, Petri Leskinen1, Jouni Tuominen1,2, Charles van den Heuvel3, Eero Hyvönen1,2 and Mikko Kivelä1

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; Miriello 2013). This has resulted in important insights regarding patterns of human communication (Miriello 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).
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
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.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Content</th>
<th>Span</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Platform</td>
<td>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 ~ 4K nodes, ~ 183.5K edges and ~ 854.6K wall posts.</td>
<td>1 year</td>
<td>Viswanath et al. (2009)</td>
</tr>
<tr>
<td>Email</td>
<td>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.</td>
<td>6 years</td>
<td>Rossi and Ahmed (2015)</td>
</tr>
<tr>
<td>Forum</td>
<td>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, ~ 1.6K edges and ~ 1.4M thread comments.</td>
<td>7 years</td>
<td>Karimi et al. (2014)</td>
</tr>
<tr>
<td>Mobile</td>
<td>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 ~ 1.74K nodes, ~ 190K edges and ~ 0.8M phone calls.</td>
<td>4 months</td>
<td>Onnela et al. (2007a)</td>
</tr>
</tbody>
</table>
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
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 \( k \)th inter-event time \( \tau_k = t_{k+1} - t_k \) for sequences of at least length 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 \( \mu \)) and \( \tau_\sigma \), 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
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 $\mu$ and burstiness $B$. We uncouple $\mu$ from the number of contacts by scaling the distribution by $\mu_{w_i}$, 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 $\mu/\mu_{w_i}$, (ii) Burstiness coefficient $B$.
distribution value for the quartile category $i$. Previous research has shown that the distributions of $\mu/\mu_w$ 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 $\mu$. 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 $\mu$ 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
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 $\mu$ 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 behavior. 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 behavior 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 \( n_{\text{no-ng}} \) (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 \( n_{\text{no-ng}} = 5 \), 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 time-scales, 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
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 Sebastian Derks for discussions and the possibility of using ePistolarium data in our research. Thanks to Sebastian Dumont for providing the correspondSearch 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
Heath T, Bizer C (2011) Linked data: evolving the web into a global data space. Theory and technology, synthesis lectures on the semantic web


Innis H (1951) The Bias of communication. Toronto: University of Toronto Press


Miritello G (2013). Temporal patterns of communication in social networks. https://doi.org/10.1007/978-3-319-00110-4


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