# Supplementary Material for Communication Now and Then: Analyzing the Republic of Letters as a Communication Network

## Contents



# 1 Analysis of the Static Graph

## Degree distribution

We first examine the degree distribution, where the degree is the number of neighbours for each node. Degree distributions in most empirical networks are heavy-tailed, which means that the majority of the nodes have a relatively small degree, whereas a minority of the nodes concentrate a vast amount of connections. If both axes are in logarithmic scales, heavy-tailed distributions might resemble downward-facing straight lines. Much literature and debate about the exact nature of this empirical behaviour [1–3], which we will not cover in this paper. Figure 1 depicts the degree distribution of correspSearch including in consecutive fifty-year periods. The degree distribution displays some of the characteristic inhomogeneities of social networks, with a broad distribution. This behaviour is relatively in line with our reference social networks, although decay seems to be faster for small degree nodes. Overall, large-degree nodes might be over-represented in the static graph. More importantly, the amount of low-degree nodes is larger than from other social networks. For instance, 78% of nodes are of unitary degree, meaning that a substantial majority of the data corresponds to single epistolary contacts (Platform 24%, Forum 19 %, Email 65 % and Mobile 55 %). These features of large and low degree nodes, not present in the reference datasets, is likely an effect of constructing datasets from ego-networks.

## Centrality

Next, we focus on network centrality measures. In network science, a node centrality informs of the relative *importance* of a node in a network [4]. This importance might tell of different measures: how easy it is for a node to reach others, how many shortest paths pass by this node. Indeed, the degree of a node is a form of a centrality measure, since it informs how many other nodes a single node can directly reach. For social networks, the distribution of node centralities might differ depending on the context. While we would not expect single nodes to be highly important for large-scale social networks that include multilayered connections, this might not be the case for specific social foci [5]. As an example, in hierarchical settings such as workplaces, people with administrative positions might have higher network centralities, while other individuals would interact mostly with their immediate work team.

Figure 2 depicts the betweenness centrality distribution for correspSearch and our reference datasets. This centrality measure can be interpreted as a node's importance in connecting the network, as high betweenness implies that that a node is on large fraction of shortest-paths between other nodes pairs. In the context of epistolary networks, betweenness has been used to assess the role of introduction letters to gatekeepers who controlled access to particular communities within the Republic of Letters [6]. Most of the reference datasets exhibit fast decay, with most nodes in reference networks having scores below 0.07. CorrespSearch contains a small number of central nodes ($> 0.1$ and up to 0.6). As with degrees, this effect is possibly due to the sampling of the network, based on the making of letter editions of famous historical figures and successive digitization practices.

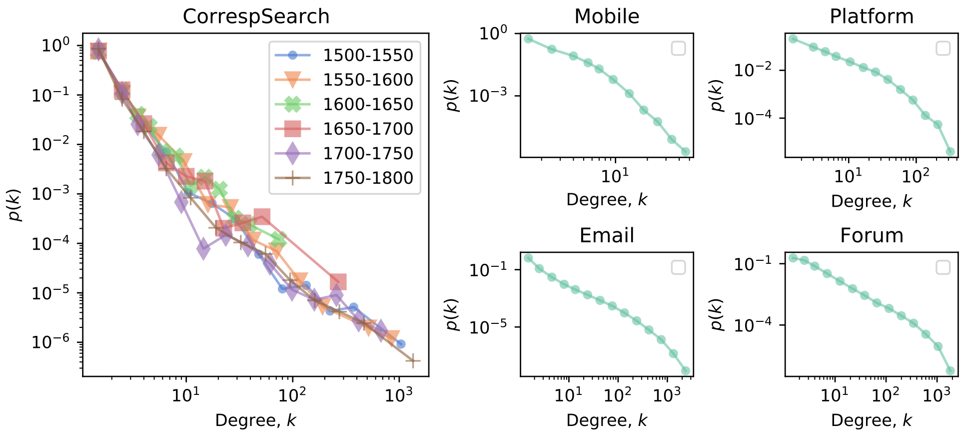


Figure 1: Degree distribution for 50-year aggregation windows of (*left*) correspSearch. Degree distribution of static nets over all observations for (*top-center*) Platform, (*bottom-center*) Forum, (*top-right*) Email, (*bottom-right*) Mobile. All aggregation windows follow a similar trend, where decay slows down faster than in the reference dataset with a higher number of "popular" nodes than in reference datasets.

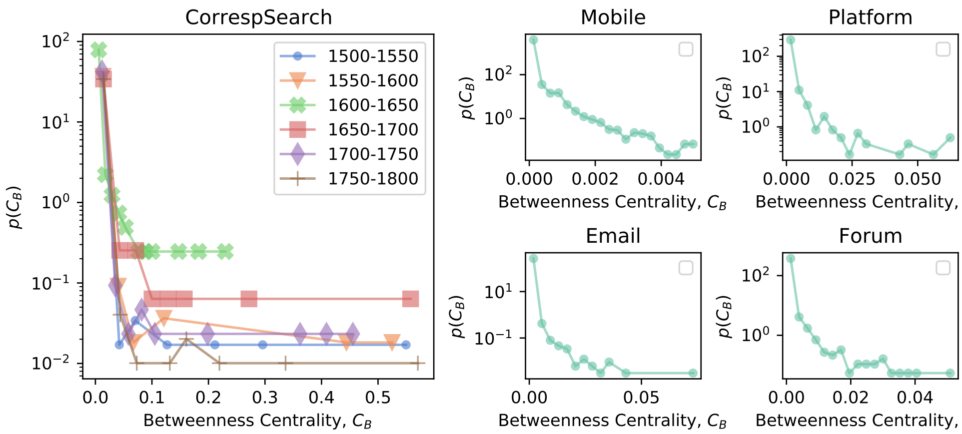


Figure 2: Distribution of betweenness centrality for 50-year aggregation windows of (*left*) correspSearch. Degree distribution of static nets over all observations for (*top-center*) Platform, (*bottom-center*) Forum, (*top-right*) Email, (*bottom-right*) Mobile.

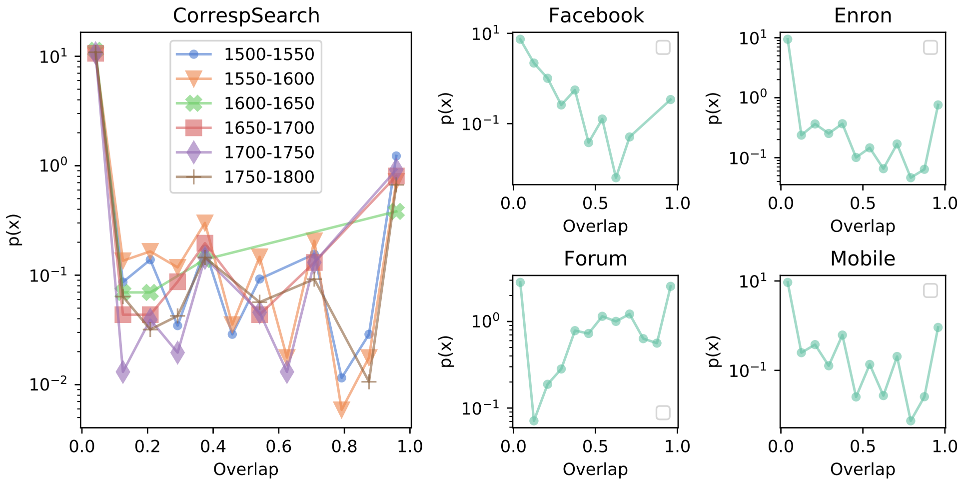


Figure 3: Distribution of topological overlap for (*left*)50-year aggregation windows of correspSearch. Overlap of static nets over all observations for (*top-center*) Platform, (*bottom-center*) Forum, (*top-right*) Email, (*bottom-right*) Mobile.

### Topological Overlap

In social networks, nodes tend to cluster together forming overlapping communities. The notion of community "bridges" characterizes links that lie between communities, i.e., links where the two nodes don't share many common friends. One way to characterize such links is via topological overlap [7, 8], or the ratio between common neighbours between two linked nodes and all their neighbours (for details see Materials and Methods). Figure 3 depicts the distribution of overlap values. In all cases zero overlap dominates, followed by unitary overlap, both cases likely due to (a) lower propensity of sampling triads (b) overall low sampling of neighbours. Notably, for the Forum dataset overlap has a mode at around 0.6, local pointers towards community structures in the network.

### Shortest Paths

Next, we focus on the distribution of shortest paths. A path is a collection of edges that you'd need to follow in order to reach another node. Since in networks it is usually possible to reach another node via several paths, the shortest path is a measure of the minimum number of steps between two people—the idea behind "degrees of separation". In this sense, the distribution of shortest paths takes pairs of people, although these people do not need to be directly connected.

Figure 4 displays the distribution of shortest paths. Strikingly, the most common shortest path length is between 3 and 4, which occurs for 31% to 52% of the pairs of nodes depending on the temporal aggregation. This is quite notable since the correspSearch contains data across large temporal and spatial dimensions. A reason for this could be related to the existence of highly central nodes that facilitate short paths. Albeit not directly comparable, precisely because the geographical and temporal scope is different, the most common shortest path for our reference networks is between 2 and 5. The overall largest shortest paths are found on the Mobile network, corresponding to a region in a large country.

## 2 Social Signatures

We claim that social signatures hold for several filter parameter ranges, where these parameters refer to the minimum number of neighbors at each temporal bin and the temporal length (in years) for each bin. The former parameter thus refers to a data-quality filter that attempts to ensure more complete samples of ego networks, while the latter controls the temporal resolution of the effect. Figure 5 depicts the results for different filter values, with an observable effect of smaller $d_{self}$ distribution values than $d_{ref}$ across different parameter combinations. On average, $d_{ref}$ has a mean value of 0.3, whereas $d_{self}$ of 0.2. For the top row (minimum one neighbor), the distributions have concentration points on zero, which is a bias effect induced by having social signatures with one person (100% of time is devoted to the top alter).

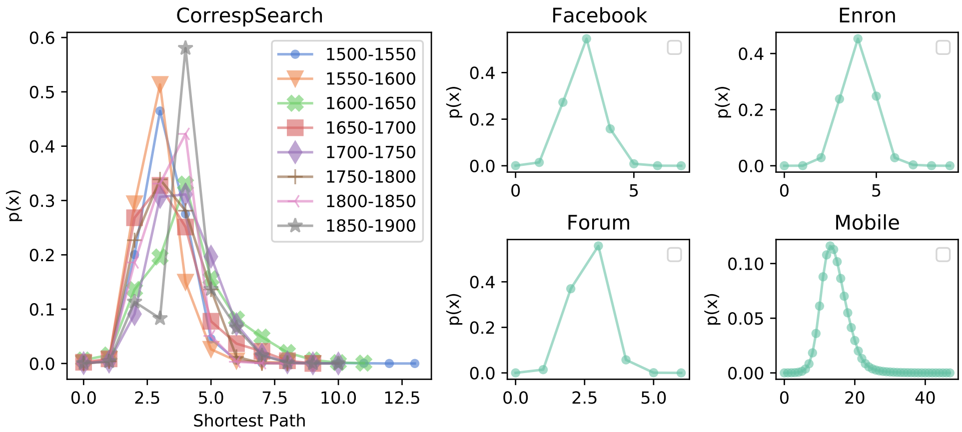


Figure 4: Distribution of shortest path length for (*left*) 50-year aggregation windows of correspSearch. Degree distribution of static nets over all observations for (*top-center*) Platform, (*bottom-center*) Forum, (*top-right*) Email, (*bottom-right*) Mobile.

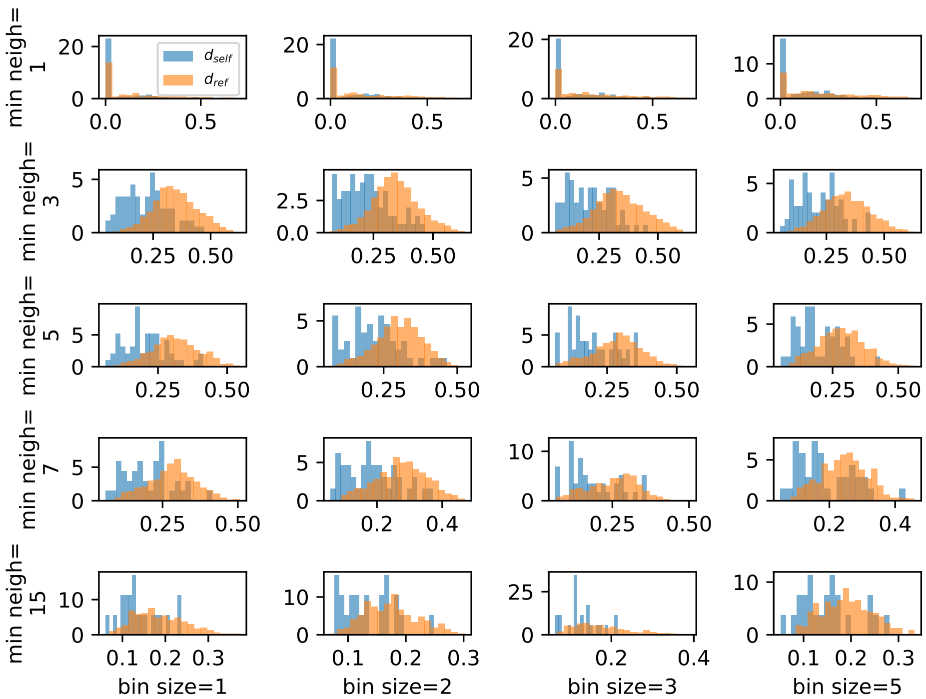


Figure 5: Comparison between self and reference distances for social signatures using different filter parameter values. On the rows, minimum number of neighbors for each time bin, and on the columns, the length of the temporal bin. Distribution values for self distances are smaller than reference distances.

# References


[1] Albert, R., Barabási, A.-L. (2002). *Statistical mechanics of complex networks*. Reviews of Modern Physics, 74(1), 47–97. https://doi.org/10.1103/revmodphys.74.47

[2] Clauset, A., Shalizi, C. R., Newman, M. E. J. (2009). *Power-Law Distributions in Empirical Data*. SIAM Review, 51(4), 661–703. https://doi.org/10.1137/070710111

[3] Holme, P. (2019). *Rare and everywhere: Perspectives on scale-free networks*. Nature Communications, 10(1). https://doi.org/10.1038/s41467-019-09038-8

[4] Newman, M. E. J. (2010). *Measures and metrics*. In Networks (pp. 168–234). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199206650.003.0007

[5] Feld, S. L. (1981). *The Focused Organization of Social Ties*. American Journal of Sociology, 86(5), 1015–1035. https://doi.org/10.1086/227352

[6] van Vugt, I. (2021). *Brockerage and closure in the Republi of Letters: the case of Antonio Magliabechi and the Dutch Republic*. Journal of Interdisciplinary History. Accepted.

[7] Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., Barabási, A.-L. (2007). *Structure and tie strengths in mobile communication networks*. Proceedings of the National Academy of Sciences, 104(18), 7332–7336. https://doi.org/10.1073/pnas.0610245104

[8] Ureña-Carrion, J., Saramäki, J., Kivelä, M. (2020). *Estimating tie strength in social networks using temporal communication data*. EPJ Data Science, 9(1). https://doi.org/10.1140/epjds/s13688-020-00256-5