

The anatomy of a population-scale social network

Keywords: social networks, population-scale, register data, formal tie, clustering

The analysis of large-scale societal networks has recently seen tremendous growth, in part because of the relative abundance of digital data sources such as online social networks or mobile communication datasets¹⁻³. However, most of these data sources lack demographic data on users or are uncertain with respect to the representativity of the user sample. Moreover, it is often not clear what exact social relations these online or communication ties represent, thus, it is difficult to interpret findings⁴.

We overcome a number of these drawbacks by presenting a thorough overview of the structure of a 17M node *population-scale social network* of a European country containing roughly 1.6B edges. This network is derived from highly curated official data sources of the country's national statistics institute. As such, it includes every resident registered on a certain day in 2018. In addition, rich individual-level demographic and socio-economic attributes on the nodes are available alongside the network structure, as well as the precise type of each social relationship we observe: family, household, work, school, or neighbor relationship, each extracted from country-level register data. Just as a typical (online) social network data may suffer from missing connections, the studied population-scale social network data may miss informal friendship connections not captured in the formal ties in this network. However, we know that we have precisely all nodes (people), and we know that for the types of connections that we have, data is very complete, which is a unique setting in social network analysis research. In this work, we present first results of how such a high quality population scale social network is markedly different from many of the large scale social networks we typically study. Below, we in particular do so by revisiting the well-known concept of *closure* in a population-scale social network context.

First, we show how the degree distribution of this network is a composition of the degree distributions of the different types of edges. In the overall degree distribution, we find a characteristic value that is in sharp contrast to the scale-free or other fat-tailed distributions found in online social networks or communication networks⁵. Second, we discuss different types of clustering in this multilayer network, and show how closed or open network structures emerge for people of certain ages. In particular, we introduce a normalized multilayer clustering coefficient that we call *excess closure*, that captures the fraction of triangles in people's social circles that span across multiple types of relationships.

Figure 1 shows how degree and excess closure change with age (a demographic attribute) in the population. Young children have low degrees and very high excess closure since they are only part of family, neighborhood, and household structures. Subsequent levels of education paired with working opportunities come with both an increasing median degree, and decreasing excess closure, reaching its minimum around the university age. Working years are characterized by a slight increase in closure, and gradually decreasing degree, giving place to low degrees and increased closure in retirement years. Finally, we find that long-range ties that span large distances are very scarce in this network, only 0.02% of all edges not being part of any triangles, which is in contrast to findings in online social networks, and does not promote fast and efficient diffusion processes over this structure.

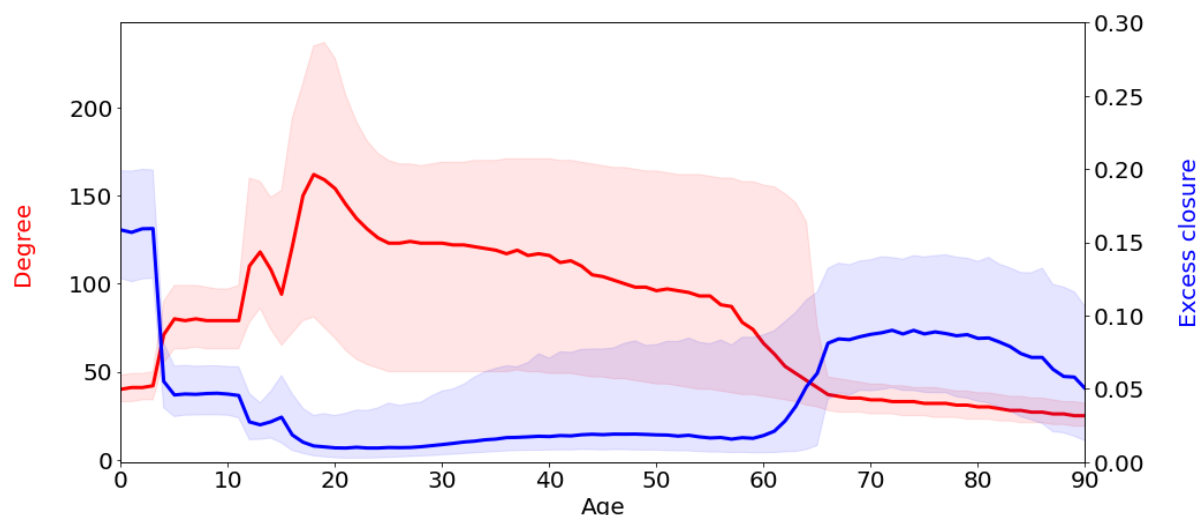


Figure 1. Median degree (red) and median excess closure (blue) in ego networks of people of a certain age. Shaded areas correspond to the 25th and 75th percentiles for each age year.

Concluding, our results show a sharp transition from closed to open network structures as young adults engage in higher levels of education, and a reverse process as people retire. The findings empirically confirm using large-scale data that individuals have very different resource structures throughout their lives, which affects their access to opportunities and information⁶. Our measurements are first steps in building both methods and universal insights on the rich network structure of highly curated population-level network datasets.

References

1. Eagle, N., Pentland, A. S. & Lazer, D. Inferring friendship network structure by using mobile phone data. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 15274–15278 (2009).
2. Park, P. S., Blumenstock, J. E. & Macy, M. W. The strength of long-range ties in population-scale social networks. *Science* **362**, 1410–1413 (2018).
3. Bailey, M., Cao, R., Kuchler, T., Stroebel, J. & Wong, A. Social Connectedness: Measurement, Determinants, and Effects. *J. Econ. Perspect.* **32**, 259–280 (2018).
4. Lazer, D. *et al.* Meaningful measures of human society in the twenty-first century. *Nature* **595**, 189–196 (2021).
5. Onnela, J.-P. *et al.* Structure and tie strengths in mobile communication networks. *Proc. Natl. Acad. Sci.* **104**, 7332–7336 (2007).
6. Tóth, G. *et al.* Inequality is rising where social network segregation interacts with urban topology. *Nat. Commun.* **12**, 1143 (2021).