

**Scuola estiva in Demografia Storica
SIDeS**



Utilizzo degli alberi genealogici digitali per l'analisi demografica: il database FamiLinx

Riccardo Omenti

07/10/2022

*Dottorando presso Dipartimento di Scienze Statistiche P.
Fortunati, Università di Bologna.*

Che cosa sono gli alberi genealogici digitali?



Fig. 1 – Pagina principale del sito **Geni.com**.

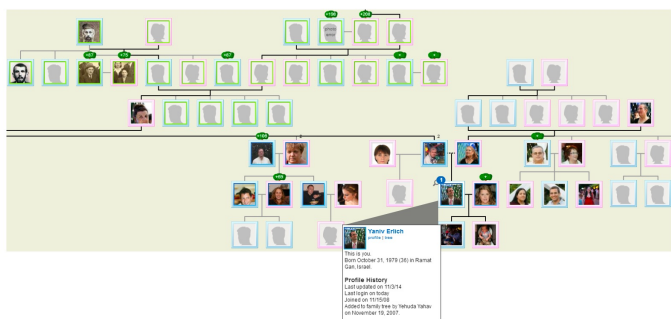
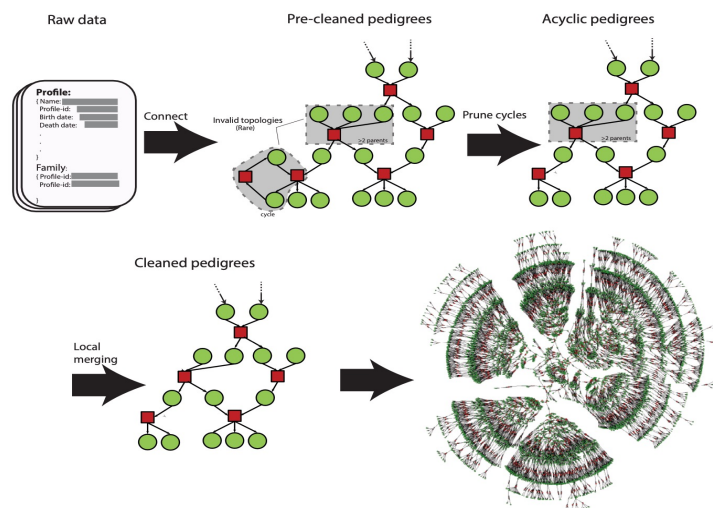


Fig. 2 – Esempio di albero genealogico digitale presente nel sito **Geni.com**

- Negli ultimi sono stati creati diversi siti internet nei quali gli utenti hanno la possibilità di inserire il proprio albero genealogico includendo, se possibile, informazioni anagrafiche per ogni familiare.
- Tra questi configurano i seguenti siti: **Geni.com**, **Ancestry.com**, **MyHeritage.com**, **Antenati.com**.
- Si tratta di una risorsa digitale creata da una **rete decentralizzata di genealogisti**.
- Tale fonte costituisce un **archivio storico digitale e transnazionale di informazioni di tipo demografico e familiare** (Alburez-Gutierrez et al., 2022).

Il database *FamiLinx*



- Database costruito e curato da un gruppo di ricercatori estraendo i dati dal sito **Geni.com** (Kaplanis et al., 2018).
- Il database contiene oltre **86 milioni di profili** creati da una rete decentralizzata di ca. **3 milioni genealogisti**.

Fig. 3 - Diagramma informativo per la pulizia del database.

Informazioni anagrafiche principali:

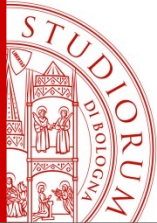
- ID (chiave identificativa e univoca per ogni profilo)
- Data e luogo di nascita, di battesimo, di morte e di sepoltura (in campi separati)
- ID del padre e della madre

Caratteristiche di *FamiLinx*



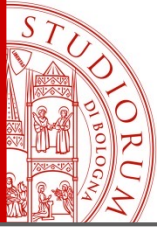
Fig. 4 – Distribuzione geografica del luogo di nascita.

- Profili riguardano individui vissuti a partire dal **1400 in poi**.
- Ca. 16 milioni di profili con luogo e data di nascita e/o di morte noto.
- Oltre l'80% di tali profili riguarda soggetti nati in Europa e/o in Nord America.
- Numerosi profili con informazioni mancanti.
- Numerosi alberi genealogici incompleti.
- Preferenza per alberi genealogici patrilineari.



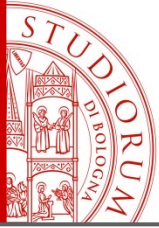
Fonti di distorsione

- **Bias di selezione:** sottorappresentazione di donne, di matrimoni senza figli, di famiglie poco numerose e di soggetti di classe economica-sociale bassa, di soggetti deceduti in età precoce, maggiore probabilità di inclusione di individui con più figli.
- **Bias di conferma:** alcuni antenati hanno una probabilità maggiore di inclusione rispetto ad altri in quanto ritenuti più degni di nota dal genealogista (es. mestiere più rilevante come cardinale etc.) .
- La **costruzione di tipo bottom-up** dell'albero genealogico determina un calo crescente nell'accuratezza delle informazioni tanto più si va indietro di generazione.



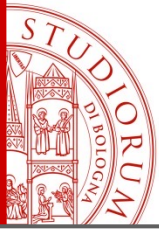
Studi demografici precedenti

- **Kaplanis et al. (2018)** : utilizzo dei dati di famiLinx per studiare la longevità e la sua ereditabilità attraverso l'età media alla morte nel periodo 1820-1940 e la correlazione di tale indicatore tra genitori e figli.
- **Blanc (2020)**: studio delle potenzialità di questa fonte per testare l'ipotesi di calo precoce della fecondità in Francia nel secolo XVIII.
- **Stelter & Alburez-Gutierrez (2022)** : analisi della mortalità dei maschi nell'Impero Germanico tra il 1600 e il 1900 che mette in evidenza come tale campione tenda a sovra-rappresentare le classi elitarie.
- **Chong et al. (2022)**: sviluppo di un modello Bayesiano gerarchico per correggere le distorsioni dei tassi di mortalità età-specifici in Svezia, Finlandia, Danimarca, Norvegia e Francia utilizzando come informazioni a priori i tassi di mortalità estratti dallo Human Mortality Database.



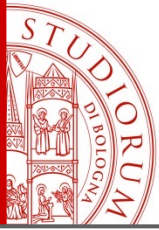
Obbiettivi della mia tesi

- Calibrazione dei tassi di fecondità età-specifici stimati tramite il database FamiLinx in alcuni paesi Europei e Nord Americani utilizzando modelli Bayesiani gerarchici.
- Calibrazione del numero medio di parenti di soggetti appartenenti a diverse coorti di nascita per diverse età calcolato dai dati famiLinx utilizzando il numero atteso di parenti stimato tramite un opportuno modello di Demografia Formale (si veda Caswell, 2019).



Bibliografia

- Kaplanis, J., Gordon, A., Shor, T., Weissbrod, O., Geiger, D., Wahl, M., ... & Erlich, Y. (2018). Quantitative analysis of population-scale family trees with millions of relatives. *Science*, 360(6385), 171-175.
- Alburez-Gutierrez, D., Barban, N., Caswell, H., Kolk, M., Margolis, R., Smith-Greenaway, E., Song, X., Verdery, A. M., and Zagheni, E. (2022). Kinship, demography, and inequality: Review and key areas for future development.
- Stelter, R., & Alburez-Gutierrez, D. (2022). Representativeness is crucial for inferring demographic processes from online genealogies: Evidence from lifespan dynamics. *Proceedings of the National Academy of Sciences*, 119(10), e2120455119.
- Zhao, Zhongwei. "Chinese genealogies as a source for demographic research: A further assessment of their reliability and biases." *Population Studies* 55.2 (2001): 181-193.
- Hollingsworth, Thomas H., and T-T. Hollingsworth. "Genealogy and historical demography." *Annales de demographie historique*. Editions Belin, 1976.



Grazie per l'attenzione

email: riccardo.omenti2@unibo.it