ABSTRACT


Ji-Ping Lin
Research Center for Humanities and Social Sciences, Academia Sinica
115 Taipei, Taiwan
jplin@sinica.edu.tw

Keywords: complexity, micro genealogy, record linkage, simplicity, Taiwan indigenous peoples (TIPs)

Background: Taiwan Indigenous Peoples (TIPs) are a branch of Polynesian-Malaysian (or Austronesian) ethnic groups in genetic and linguistic context (Li et al., 2015). Since early 17th Century, TIPs had been playing a crucial role during the Great Marine Times of East Asia trades. There was a rich body of ethnographic, official and academic records on TIPs before 1940. However, the period of 1940-2000 marks as data “Dark Ages” for TIPs due to 1941-45 Pacific War and 1946-1990 KMT authoritarian rule (Lin,2012). Persistent lack of TIPs data led TIPs to become isolated and marginalized and thus underdeveloped. Taiwan resumed TIPs population census in 2000 and began recording TIPs individual records in household registration system since 2003. The research is based on a four-year Joint Research Agreement between Academia Sinica and Council of Taiwan Indigenous Peoples. Similar to other countries’ ethnic minority population, contemporary TIPs in Taiwan are associated with lower SES, much shorter life expectancy, more disadvantaged
labor market outcomes such as income gains and employment opportunities, etc. (Lin, 2013a). Although TIPs share to total Taiwan population is of only 2.3%, the importance of research on TIPs lies in the following facts. Based on the author previous co-authored studies on the internal migration of TIPs, TIPs are characterized by four features in terms of population distribution and migration: (1) geographically segregated population distribution, (2) very migratory and mostly rural-to-urban migration, (3) periphery of metropolitan areas serving as main destination choice for TIPs rural-to-urban migrants; (4) weak ability of TIPs migrants to make onward migration and mostly choose return migration; once repeat migration occurs, TIPs are more likely to choose return migration (Lin, 2013b).

**Objectives:** The research question is: if migration serves as an effective means in promoting individual SES and thus in enhancing integration in the long run, then how can we explain a phenomenon that TIPs have persistently been associated with much lower SES than non-TIPs which apparently counters a widely accepted wisdom by migration scholars? It is not easy to answer the above-mentioned question. The author suggests that construction of individual genealogy, which is termed as “Social DNA Sequencing” in the paper, by taking advantage of micro data sets & computing technique & programming hacking skills as the first step to answer the research question and to unveil the level of assimilation and integration. This research aims to answer a question: in light of long history of Chinese immigration into Taiwan, to what extent are contemporary Taiwan indigenous peoples (TIPs) integrated with Taiwan population system since the onset of ethnic Han population immigration from southeastern China to Taiwan four centuries ago?

**Methods:** The research at first constructs contemporary TIPs Data (or TIPD) by integrating the micro data of 2000 Taiwan population census and household registration system, using state-of-the-art data science technology like geocoding of primary statistical areas, record linkage to distinguish natural & social increases, using high-performance cluster computers (HPC) to construct micro genealogy to study pattern of inter- & intra-ethnic marriage and level of integration, etc. The most challenging in the research is developing effective computing method to construct individual genealogy. Such work requires both exact and probabilistic record linkage techniques that involve thousands of billions matching of given names & family names by gender and ethnicity. To overcome computing barriers, the research takes advantage of in-memory high performance computing (HPC) techniques that comprise three kernel skills of manipulating digital hardware: (1) overclocking CPUs, (2) widening I/O bus bandwidth, and (3) accelerating internal memory I/O (Lin,2014). Based on constructed micro genealogy, the research constructs a macro “ethnic marriage similarity index (EMSI)” by
aggregating values from individual “intra/inter-ethnic marriage indicators (IEMIs)”. EMSI serves as a composite index of overall social integration. Measurement of identity is based on two composite indices that are aggregated from individual MIs (Matriarchy Indicators) & PIs (Patriarchy Indicators).

**Results:** Research outcomes are twofold: (1) construction of TIPS family trees & marriage structure; (2) to assess the extent of integration of TIPS and Non-TIPS. Main findings are: (1) EMSI computing result suggests that in terms of the extent of integration of TIPS with Taiwan population system, ethnic population size is negatively associated with integration; (2) Aggregation of MIs & PIs suggest that in terms of ethnic identity, TIPSs’ matriarchy identity tends to outweigh patriarchy identity a little bit. This finding fit general wisdom and TIPSs cultural tradition.

**Conclusions:** With gradual availability of massive micro data & substantial drop of digital hardware costs, computation for social complexity based on simplicity like the construction of micro genealogy becomes feasible. Computing issues remain challenging & total costs of computing are still time expensive. The emerging data science that integrates multi-disciplinary skills & knowledge of “hacking skills”, “advanced math/stat”, and “domain knowledge” is crucial to overcome such constraint.

**Main references**


