Appendix: Additional information about the datasets and surveys

“Structure in personal networks: Constructing and comparing typologies”
Raffaele Vacca
Published in *Network Science*

Table 1 in the main article reports basic information about the six datasets, including the name generators and alter tie questions (edge interpreters) used to collect them. This appendix provides more details about the studies in which the data were collected.

The *MBC* data were collected with two surveys using identical name generators and edge interpreters. These were conducted in 2004–2006 (Barcelona, Spain) and 2012 (Milan, Italy) to study the cultural and economic incorporation of non-European immigrants in Spain and Italy. Respondents were immigrants from African, Asian, and Latin American countries, recruited using a combination of venue-based and link-tracing sampling (Vacca et al., 2018; Lubbers et al., 2007). The 385 ego-networks in this dataset have a fixed size of 45 alters, including any type of family, friend, or acquaintance with whom ego has had contacts in the past 2 years (and whom ego could still contact). The name generator aimed to capture respondents’ *total* personal networks (McCarty et al., 1997).

The *FRA* dataset was obtained in 2012 as part of a study on community resilience following the 2010 Deepwater Horizon oil spill in the Gulf of Mexico, and the ensuing collapse of oyster fishery in the Florida Apalachicola Bay area (Puetz & Mayer, 2018). A survey was administered to 303 residents of Franklin county, in the Florida’s Gulf coast, to gain information on their personal relationships that could provide support in case of disaster. Following community-based participatory research principles, respondents were recruited by community organizations and complete personal network data were obtained for 293 individuals.

The *ECU* data include 264 personal networks with a fixed size of 25 alters. As part of a research on community resilience in case of natural disaster, a personal network survey was administered in 2009 to a random sample of residents of five Ecuadorian villages exposed to natural hazards on the stratovolcano Tungurahua (Jones et al., 2013). Similar to those in the MBC data, the resulting egocentric networks include any family, friends, and acquaintances with whom ego has had recent contact.

The *PAR* data were collected in 2014–2015 in Paris and other large metropolitan areas of France for a study on Romanian Roma migration and incorporation in the French society. The 119 respondents were Romanian Roma immigrants in France, and they were asked to name 30 family members, friends, or acquaintances "on whom they could rely." While tie data were collected for all the 30 alters, I used this dataset to extract realistic personal networks of variable size between 10 and 25 alters. The goal was to simulate the common situation in which the data include ego-networks of different sizes within a given minimum and maximum number of alters. The data were simulated by extracting one random integer number (n_i) in the [10, 25] range for each respondent, and retaining only the subgraph of the first n_i alters nominated by the respondent. When a survey asks to list a free number of alters between a minimum and a maximum value, respondent fatigue leads to lower numbers of alters being more likely. To simulate this characteristic of the data, the random integers were extracted assigning decreasing probabilities to the integers between 10 and 25. Figure S1 in the Supplementary Information shows the resulting distribution of network size.

The *TAL* dataset is the result of a 2011 epidemiological and social survey conducted in Tallahassee, Florida. This was part of a research on racial health disparities with a focus on cardio-vascular diseases among African Americans (Boston et al., 2015). Respondents were selected from African American households using a multistage probability sampling design, which started with the identification of neighborhoods with higher proportions of African

American residents, and then chose random block groups within neighborhoods and random residential addresses within block groups. A network of 30 alters was elicited from each respondent using a generic name generator based on acquaintance and contact. Similar to the PAR data, edge interpreters were asked for all the 30 alters, but these data were used to simulate a dataset with variable network size (in this case in the smaller network size range of 5–20 alters).

The final dataset, *SFB*, is a subset of the first-wave data from the UC Berkeley Social Networks Study (UCNets), a research on personal network change and health among young and older adults (Offer & Fischer, 2018). In its first wave (2015–2016), the survey was completed by 1,159 residents of the San Francisco Bay Area, aged 21–30 or 50–70 years. Respondents were selected using random address-based sampling and were randomly assigned to face-to-face or Internet survey modes. Personal networks were elicited using six role and activity-based name generators focused on social activities (going out to concerts, plays, clubs, sports, etc.); confiding about personal matters; advice to make important decisions (e.g., about taking a job or family issues); practical help in everyday life (e.g., with moving furniture, looking after a child, getting a ride, etc.); help in case of serious injury or sickness; and people whom the respondent provides support to. A seventh name generator asked respondents to list people whom they "sometimes find demanding or difficult", but the alter tie question was not asked for contacts elicited by this name generator. The total number of alters nominated by respondents ranged between 1 and 30. Up to five nominated contacts were then selected during the interview to ask alter tie questions. The algorithm used to select these contacts is described in detail at http://ucnets.berkeley.edu. For this article, the data were subset to personal networks with alter tie information on five contacts. These are 366 personal networks, of which 64% were obtained in face-to-face interviews and 36% through Internet surveys.

Network size is an important dimension of variation between the six datasets. The data were deliberately selected to represent different designs in terms of network size, including large, fixed-size personal networks (MBC, FRA, ECU), variable network size in different ranges (PAR and TAL), and small networks (SFB) similar to those analyzed in studies of core discussion partners (e.g., Small, 2017) or elicited by large-scale ego-network surveys such as the U.S. General Social Survey (Burt, 1984).

Alter tie data are collected in the six surveys using slightly different edge interpreters about acquaintance between alters (see Table 1 in the main article). While the surveys include different response categories to alter tie questions, this article analyzed binary ego-networks by retaining alter–alter ties whose value is at least equal to a certain threshold. Lower or higher thresholds were set in different datasets, corresponding to weaker or stricter definitions of alter-alter tie, to reproduce different possible scenarios in personal network analysis.

Supplementary information

“Structure in personal networks: Constructing and comparing typologies”
Raffaele Vacca
Published in *Network Science*

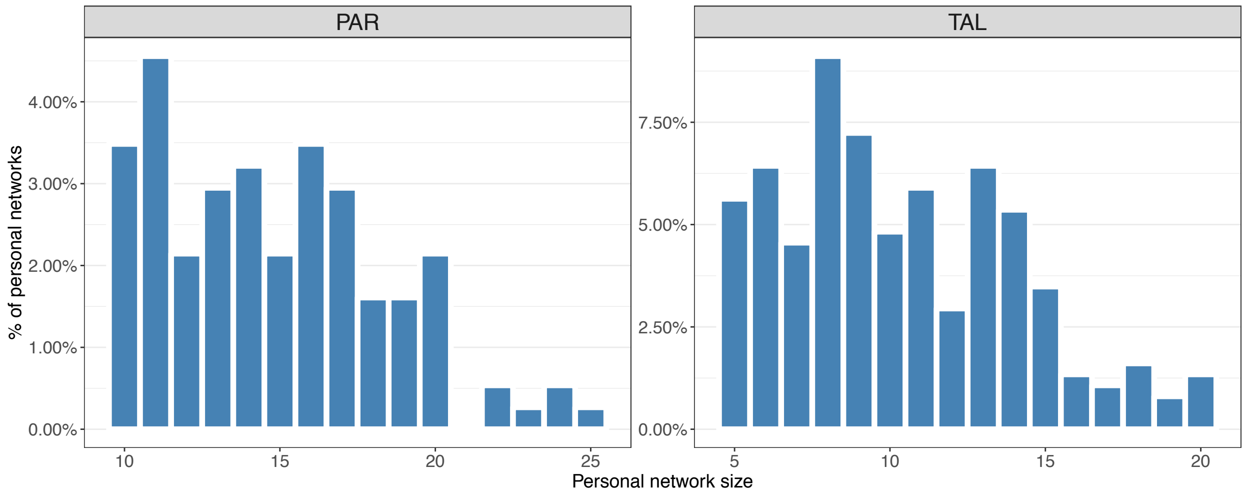

Fig. S1. Distribution of personal network size in the PAR and TAL datasets.

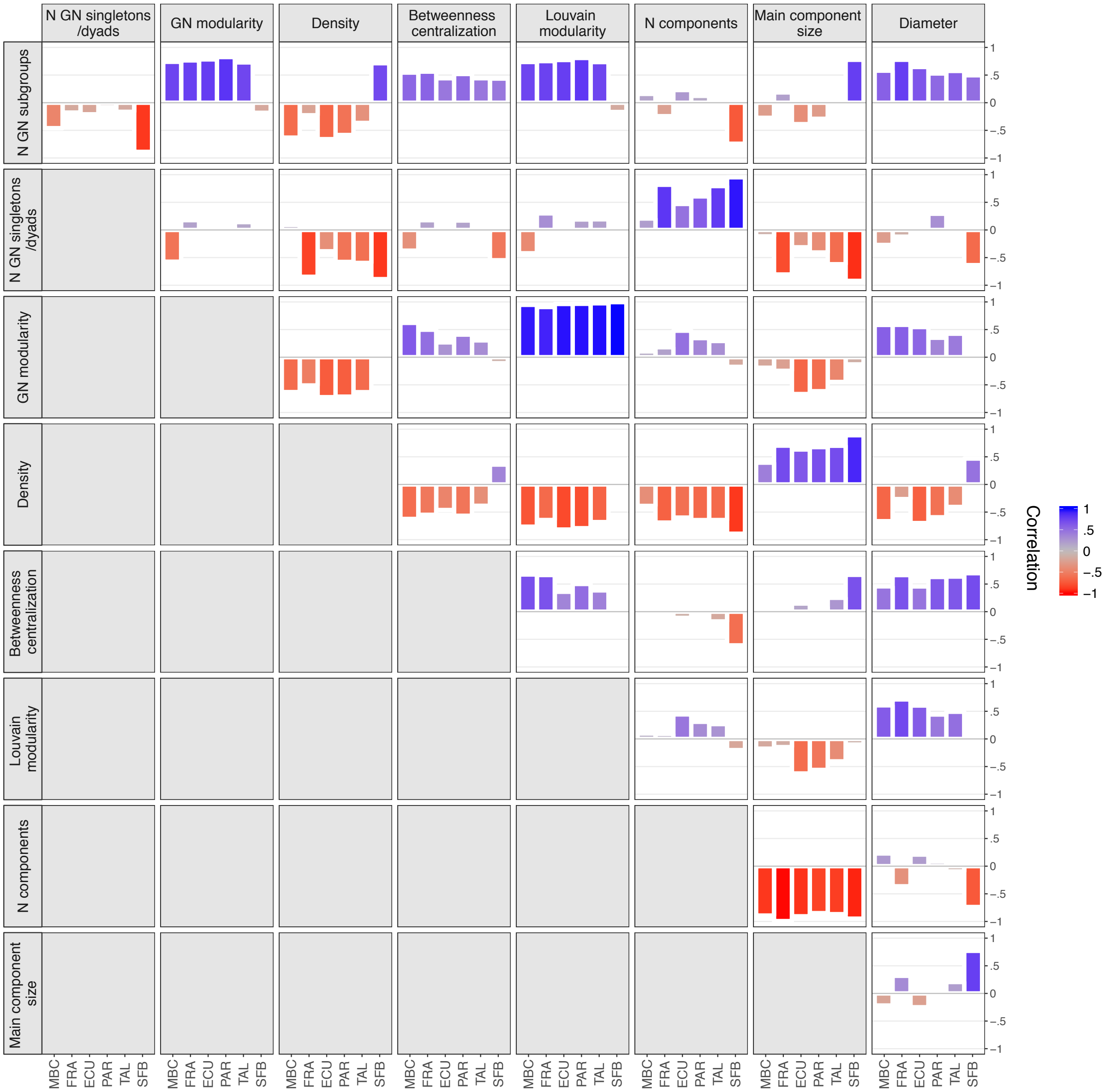

Fig. S2. Correlations between the T_1 and T_2 variables in the six datasets.

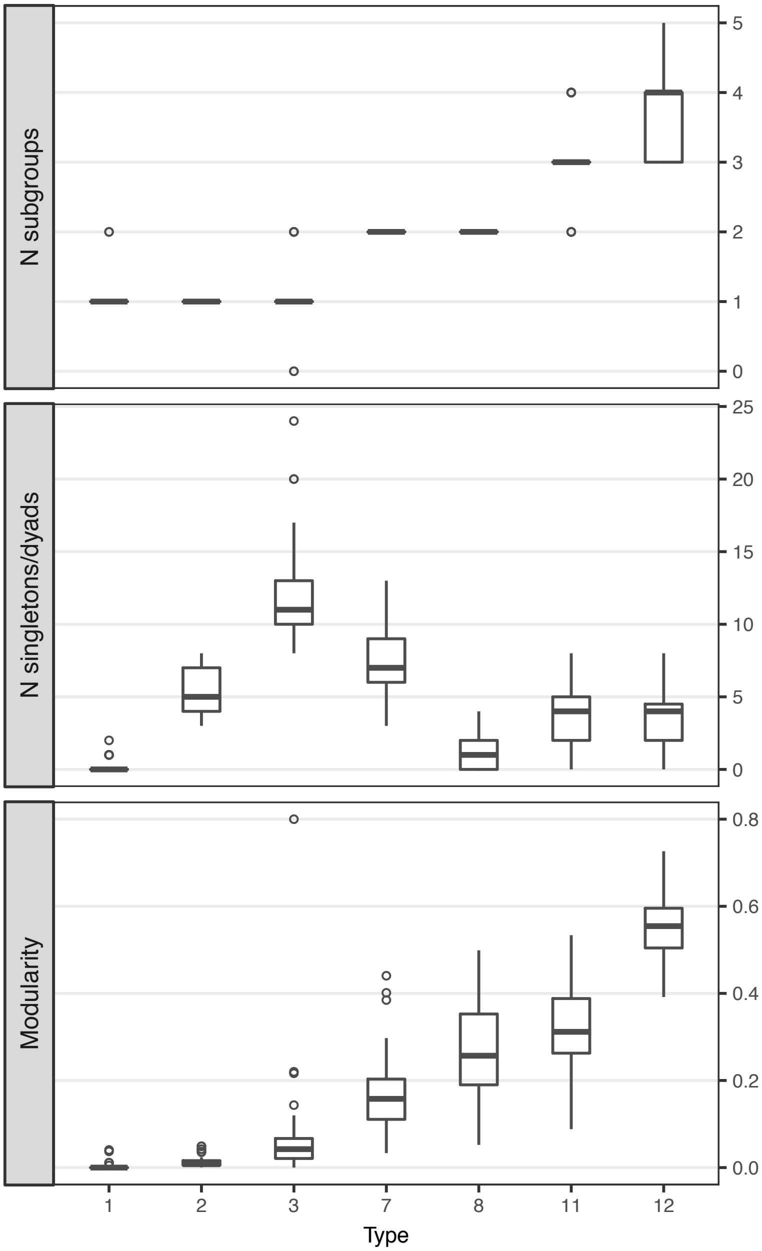

Fig. S3. Boxplots of the three T_1 variables in each type for the ECU dataset. See Table 3 for numeric IDs of T_1 types.

Table S4. Absolute frequencies (percentages) of BDG structural types (T_2) in the six datasets.

Type	MBC	FRA	ECU	PAR	TAL	SFB
1. Regular dense	79 (20.5)	189 (64.3)	66 (25)	28 (23.5)	48 (18.9)	
2. Centered dense				52 (43.7)		
3. Centered star	127 (33)	28 (9.5)		28 (23.5)		
4. Centered star/Pearl collar			29 (11)		66 (26)	
5. Pearl collar			40 (15.2)			
6. Dispersed	9 (2.3)	10 (3.4)	33 (12.5)	9 (7.6)	53 (20.9)	
7. Disconnected		20 (6.8)		2 (1.7)	9 (3.5)	
8. Small regular dense					31 (12.2)	42 (11.5)
9. Small centered star						31 (8.5)
10. Small segmented						60 (16.4)
11. Closed triad and isolates						42 (11.5)
12. Open triad and isolates						30 (8.2)
13. Connected dyad and isolates						88 (24)
14. Small disconnected						34 (9.3)
15. Hybrid	170 (44.2)	47 (16)	96 (36.4)		47 (18.5)	39 (10.7)
Total	385 (100)	294 (100)	264 (100)	119 (100)	254 (100)	366 (100)
N types	4	5	5	5	6	8

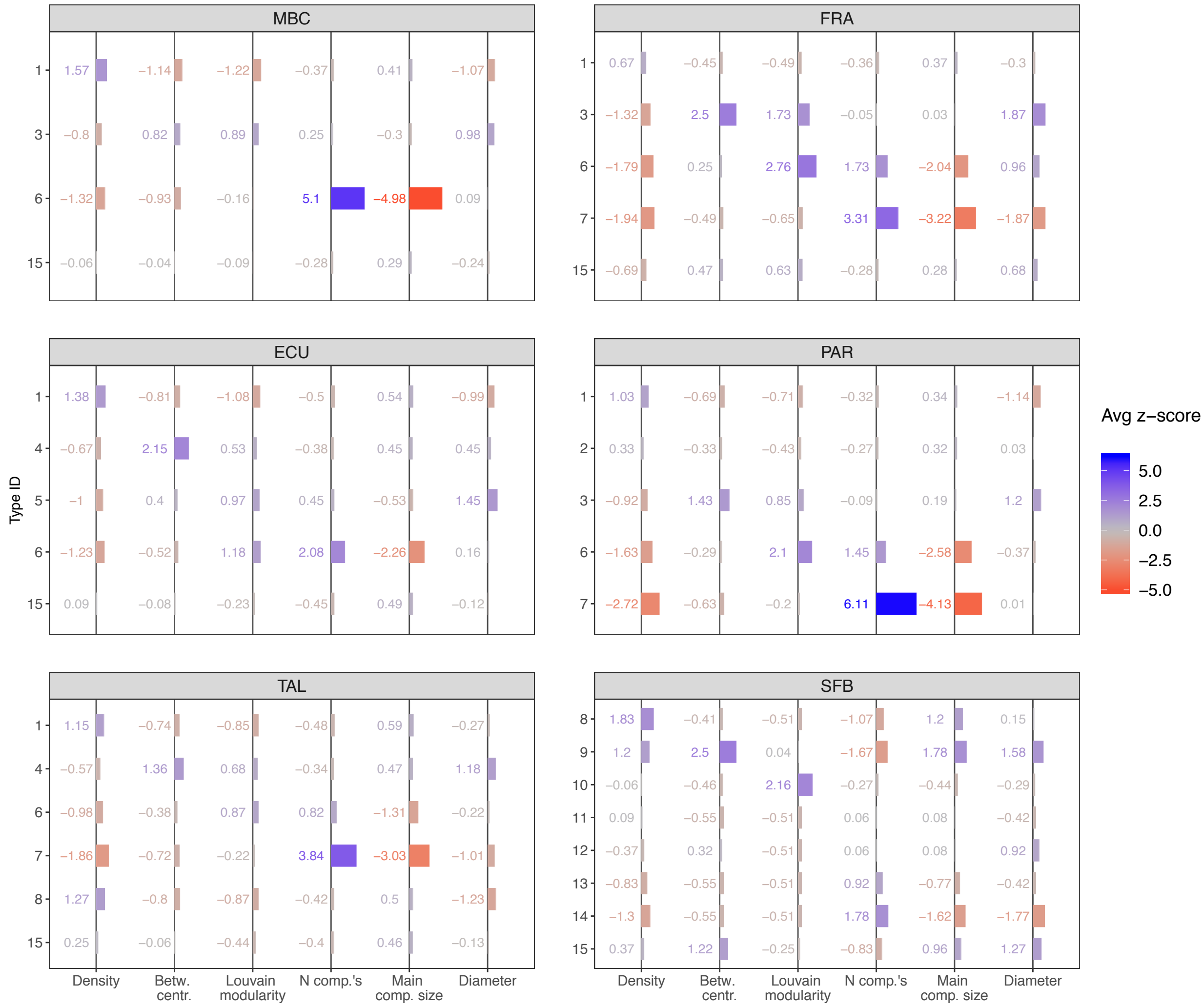

Fig. S5. Average z-scores of the six T_2 variables in each type in the six datasets. See Table A4 for numeric IDs of T_2 types.