Appendices to

Beyond the Core: Who Has Larger Social Networks?

by

Bas Hofstra, Rense Corten, and Frank van Tubergen

in

Social Forces

Contents:

Appendix A. Sample Selections, Sensitivity Checks, and Robustness Analyses

Sample selection. At wave 1 (2010-2011), two classes were randomly chosen in schools, resulting in a sample size of 118 schools, 252 classes, and 4,963 (N_{WAVE1}) pupils participating in the Dutch survey. Classroom composition changes are common in the Netherlands. Therefore, respondents from wave 1 could be distributed among different classes at the time of the second wave. To ensure that many wave 1 respondents participated in wave 2 (2011-2012) as well, schools were asked to participate with all of the classes that wave 1 respondents were attending. Therefore, 2,118 new students were interviewed (attending the same class as original wave 1 respondents), and 3,803 students who participated in wave 1 participated in wave 2 as well (76.6%; total N_{WAVE2} = 5,921). In wave 4 (2013-2014) of the CILSNL, 4,073 (N_{WAVE4}) respondents participated, of which 3,611 had also participated in wave 2 (88.7%).

Six hundred respondents in wave 1 were sampled who were not part of the random sampling frame because some schools participated with more than the two randomly drawn classrooms. Therefore, 4,363 pupils were randomly sampled in wave 1. Because of attrition rates between waves 1, 2, 3, and 4, the representativeness of our sample is not guaranteed. We include as many respondents as possible in the sample for analyses, including newcomers (nonrandom) and the nonrandom sample of wave 1 to ensure a large sample size.

Straightlining. Because the network scale-up method is sensitive to outliers we examined the "seriousness" of respondents' answers to the questionnaire. Specifically, we investigate whether there are "straightlining" respondents – i.e., respondents that tick the same box for each of a set of items belonging together. We calculated for items that belong to a battery of questions (e.g., several

items measuring “health behavior”) whether there were respondents that had a standard deviation of zero on their answers to these items. If a respondent has a standard deviation of zero it means that he/she ticked the boxes of all items in a measurement similarly. In total there were seven batteries of questions, containing 36 items that were not part of different routing options in the questionnaire. Of the 4,073 respondents in wave 4, four respondents (about 0.1%) had a row standard deviation of zero on all seven batteries of questions. In a less rigid sample selection of scoring a standard deviation of zero over, for instance, three batteries of questions, we end up with the same four respondents. A visual inspection of these respondents’ answers to other survey questions also showed careless responses. We did not consider these four respondents in the analyses, as they likely provided inaccurate answers that may disproportionately affect our results.

Facebook sample selection. Facebook networks may capture fictive accounts and inflate the Facebook network size. Additionally, Facebook friends are not limited to the Dutch, thus complicating comparisons with the scale-up method. The network size of this study correlates highly with the network size filtered for realistic Dutch names (r = .99). Hence, if there are fictive accounts or non-Dutch friends on Facebook, they are relatively randomly distributed among respondents.

Facebook network size transformations. We regressed the scale-up network size on several transformations of the number of Facebook contacts to find out whether they relate linearly or not: we regressed the scale-up network size on quadratic, cubic, logarithmic, and exponential transformations of the Facebook network size and on Facebook networks with more than 50, 100, or 200 contacts. The raw count of the Facebook contacts fits the data best.

Romantic partners as a confounding factor. The relation between romantic partners and extended network sizes may be confounded by adolescent popularity – e.g., more popular adolescents are more likely to enter romantic relationship and have more friends. We investigated whether our patterns of results were robust to "Adolescent popularity" (Mean = 1.755; SD = 2.760) as the number of incoming nominations an adolescent receives from other classmates on the question "Who are the most popular students in your class?" Our results on romantic partners are robust to either in- or excluding this covariate.

Appendix B. Populations and Their Sizes in the Scale-up Categories in the Netherlands

Table A1 provides an overview of the total population numbers in the Netherlands that are used as the X's in our scale-up method questions in the survey.

Table A1. *The used scale-up populations in 2014 (Population in the Netherlands in 2014=16,829,289)*

X's prompted in the survey	X's in the population[a,b]
1. Thomas	40538
2. Kevin	23162
3. Anne	29720
4. Melissa	11706
5. Moham(m)ed	13443
6. Arrested	251900
7. Jailed	10365
8. Groningen	198317
9. Utrecht	328164
10. Maastricht	122488
11. Den Haag	508940
12. Zwolle	123159
13. MBO (tertiary lower-vocational)	497319
14. HBO (tertiary higher-vocational)	440293
15. University	250186

[a] Firstname population estimates are from Meertens Institute (2016); [b] City, police, and tertiary education estimations are from Statistics Netherlands (2015).

Appendix C. The Scale-Up Model With Three Adjustments and Sensitivity Analyses

Equations 1 through 3 analytically summarize the scale-up model proposed by Maltiel et al. (2015). It adjusts the basic scale-up estimator of Equation (2) in the main text in three ways. Maltiel et al. (2015) specify this model in a more detail than we do here. But we can summarize the model as follows:

$$y_{ik} \sim \text{Binom}\left(d_i, \tau_k q_{ik}\right), \tag{1}$$

$$d_i \sim \text{Log-Normal}(\mu, \sigma^2), \tag{2}$$

$$q_{ik} \sim \text{Beta}\left(m_k, \rho_k\right), \tag{3}$$

where y_{ik} is the predicted number of individuals known by i in group k, d_i is the basic scale-up estimate from Equation (2), and q_{ik} is the probability that person i knows someone from group k, which is allowed to vary randomly across individuals. d_i follows a log-normal distribution (which is statistically suitable according to Maltiel et al., 2015), and the priors μ and σ follow from fitting the basic scale-up estimator in Equation (2) (see Equation (6) for σ), and q_{ik} follows a Beta distribution arriving at m_k and ρ_k using predefined priors or priors from Equation (2). Finally, τ_k is a multiplier for transmission bias, which is 1 for known populations and less than or equal to 1 for groups of unknown size. The 1 for known populations (no transmission error) is reasonable, as these groups are often less stigmatized than unknown populations. The "NSUM" package defaults to $\tau_k = 0.5$, but we experimented with two additional multipliers τ_k for the unknown population of choice (in the main analyses: individuals in prison), and in each trial, using either 0.5, 0.75, or 1 (while keeping other parameters equal) results were similar as those found in the main analysis. We keep τ_k at the 0.5 default for the main analyses presented in the paper.

Appendix D. Estimation of the Extended Social Network Size Using Linear Regressions

Table D1 shows linear regression models for the Facebook network size and the scale-up network size. It highlights the importance of accounting for sample (scale-up) and sample and privacy selection (Facebook), as some covariates change direction when one lacks to account for such biases (gender, scale-up; ethnic background, scale-up; education, both measures).

Table D1. Linear regression results of the extended network size measured via the number of Facebook friends and the scale-up method

	Facebook			Scale-up		
	Coef.	SE[a]	*p*-value	Coef.	Sea	*p*-value
Constant	359.134	106.422	0.001	170.697	316.051	0.590
Foci (H1)						
Going out	60.103	5.049	<0.001	76.816	11.988	<0.001
Associations	20.170	3.504	<0.001	40.613	7.988	<0.001
Concerts	15.325	6.558	0.021	50.573	15.595	0.002
Family	5.409	4.348	0.216	35.429	12.770	0.006
Religious meeting places	4.406	4.880	0.369	30.540	11.946	0.012
Job (yes/no)	26.826	7.697	0.001	-4.865	17.372	0.780
Similarity of cont. (H2+H3)						
Ethnic minority (ref. maj.)	24.784	17.654	0.163	71.626	43.525	0.103
# Co-ethnic Class	1.454	0.985	0.143	-0.843	3.116	0.787
# Co-ethnic School	-0.009	0.013	0.485	0.048	0.029	0.101
Rom. partner (yes/no) (H4)	37.887	8.419	<0.001	15.773	23.634	0.506
Education (H5)						
Vocational (ref.)	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Senior general	-2.426	17.877	0.892	-30.645	27.798	0.273
University prep.	-42.336	11.223	<0.001	-173.587	37.843	<0.001
Girl (ref. Boy) (H6)	26.152	8.740	0.003	-43.237	18.075	0.018
Confounders						
Early adopter of FB						
No (ref.)	Ref.	Ref.	Ref.	-	-	-
Yes	65.509	7.033	<0.001	-	-	-
Non-participation w2	48.813	32.749	0.139	-	-	-
Hours FB per day						
< 1 hour (ref.)	Ref.	Ref.	Ref.	-	-	-
1-2 hours	43.002	10.346	<0.001	-	-	-
2-3 hours	58.769	13.394	<0.001	-	-	-
3-4 hours	76.847	25.459	<0.001	-	-	-
> 4 hours	71.546	28.774	0.014	-	-	-
Non-participation w2	11.545	9.530	0.228	-	-	-
Age in months	-1.657	0.447	<0.001	0.909	1.367	0.508
Observations	2687			3120		
Adjusted R_2	0.192			0.077		

[a] Robust standard errors corrected for 112 school clusters.

Appendix E. Selection Equation Outcomes

Table E1 provides the selection equation results of our Heckman selection models. What we observe is that ethnic minorities are far less likely to be have an observed Facebook network size. Additionally, ethnic minorities, boys, and lower educated are less likely to have an observed scale-up network size.

Table E1. *Selection part of the Heckman selection model in Table 4*.

	Facebook			Scale-up		
	Coef.	SE	*p*-value	Coef.	SE	*p*-value
Constant	0.077	0.040	0.054	-0.046	0.048	0.347
Ethnic minorities (ref. maj.)	-0.557	0.049	<0.001	-0.293	0.044	<0.001
Girls (ref. Boys)	0.018	0.041	0.650	0.287	0.035	<0.001
Education						
Vocational (ref.)	Ref.	Ref.	Ref.	Ref.	Ref.	Ref.
Senior general	0.133	0.070	0.056	0.213	0.058	<0.001
University prep.	0.035	0.046	0.444	0.423	0.053	<0.001