Online Supplement A: ERG models for Facebook friends

To check the robustness of our substantive findings using a different measure of friendship, we replicated all ERG models in this paper on the network of "Facebook friendships" among the same population of students. Following Mayer and Puller (2008), we have argued elsewhere that—while the meaning of each tie and the strength of the underlying relationship it represents certainly vary across students—Facebook friendships more closely approximate "weak tie" or "acquaintance" relationships on the aggregate (Lewis et al. 2008b; see also Ellison et al. 2007). To become Facebook friends, one student must simply click a link on another student's profile to "request friendship," and the other student is then given the option to either accept or reject the tie. The consequence of accepting is that an undirected tie is added between the two students, and the consequence of rejecting is that no tie appears, such that it is impossible for an outside observer to determine who initiated the friendship.

Table A1 displays the results of running identical models to those presented in Table 1 on the network of Facebook friends, except that the reciprocity term is omitted and an undirected form of GWESP is used. Because a model estimated over the entire cohort of 1,640 students was computationally intractable, we restricted attention to the same subpopulation of 736 students as in our previous analyses (minus 6 students for whom picture albums could be viewed but not Facebook friends due to privacy settings).

These models suggest that there are indeed some differences between the picture friend and Facebook friend networks.[1] First, the edge parameter is expectedly higher in almost all models, indicating a greater baseline tendency for ties to form (i.e. the network is more "dense"). The exception is Model 3, where the much lower edge coefficient combined with the high GWESP coefficient indicates that much of this density results from a preponderance of triangles in the Facebook friend network.

[1] It is important to note that among networks with different quantities of nodes, ERG parameter estimates are not necessarily comparable. We therefore focus attention in this section on the significance and direction of coefficients.

Table A1

Replication of Models 1-5 for Facebook friendships

	Model 1	Model 2	Model 3	Model 4	Model 5
Edges	-2.53 (0.01)***	-2.53 (0.01)***	-7.72 (0.01)***	-2.90 (0.03)***	-2.58 (0.01)***
Racial homophily					
Whites	0.19 (0.01)***	0.22 (0.02)***	0.16 (0.02)***	0.56 (0.03)***	0.20 (0.02)***
Blacks	2.34 (0.05)***	2.17 (0.07)***	2.48 (0.06)***	1.79 (0.06)***	2.29 (0.05)***
Asians	0.81 (0.03)***	0.50 (0.04)***	0.81 (0.03)***	0.90 (0.04)***	0.78 (0.03)***
"Mixed"	0.76 (0.15)***	0.76 (0.15)***	0.68 (0.15)***	-0.09 (0.16)	0.71 (0.15)***
Hispanics	1.16 (0.08)***	1.10 (0.12)***	1.07 (0.10)***	0.90 (0.09)***	1.13 (0.08)***
Ethnic homophily					
Mainstream whites		-0.06 (0.02)*			
Ethnic whites		0.11 (0.07)			
Mainstream blacks		0.30 (0.09)**			
Ethnic blacks		0.56 (0.32)			
South Asians		1.39 (0.13)***			
East Asians		0.43 (0.06)***			
Middle East/North Africans		0.93 (1.16)			
South-East Asians		0.66 (0.28)*			
Mainstream Hispanics		0.14 (0.16)			
Ethnic Hispanics		-0.21 (0.32)			
Micro-ethnic homophily					
Chinese		1.19 (0.28)***			
Cubans		0.79 (0.75)			
Indians		1.80 (0.53)***			
Irish		0.18 (0.26)			
Koreans		1.11 (0.45)*			
Arabs		0.41 (1.44)			
Scandinavians		1.35 (0.69)			
British		2.89 (1.23)*			
Jews		0.75 (0.15)***			
Russians		2.29 (0.44)***			
Vietnamese		1.86 (0.53)***			
Africans		-0.41 (0.36)			
Mexicans		0.25 (0.71)			
Caribbean		8.37 (43.96)			
Nigerians		0.61 (0.35)			
Balancing mechanisms					
Reciprocity			NA		
Triadic closure (GWESP)			3.85 (0.01)***		
Sociality[a]					
Blacks				0.46 (0.03)***	
Asians				0.14 (0.03)***	
"Mixed"				0.61 (0.03)***	
Hispanics				0.32 (0.03)***	
Homophily based on regional origin					
Foreign-born					0.64 (0.07)***
New Englanders					-0.03 (0.07)
Students from Pacific states					0.19 (0.08)*
Californians					0.45 (0.09)***

Students from Massachusetts					0.28 (0.10)**
Homophily based on socioeconomic status					
Graduates of "select 16" boarding schools					0.70 (0.13)***
Homophily based on shared cultural taste					
Fans of Pirates of the Caribbean					-1.63 (1.01)
Fans of The Beatles					0.30 (0.04)***
Fans of country					0.89 (0.22)***
Fans of R&B, hip hop, and rap					0.78 (0.08)***
Fans of The Bible					0.10 (0.27)
Fans of Kurt Vonnegut					0.90 (0.35)*
Shared foci based on academic major					
Economics					0.39 (0.04)***
History					0.44 (0.12)***
Applied Mathematics					1.23 (0.21)***
English Literature					0.20 (0.11)
Sociology					0.34 (0.44)
Physics					-0.09 (0.40)
Neurobiology					-0.27 (0.60)
Micro-biology					0.93 (0.17)***
AIC	152762	152478	150281	152232	152172

Note: Numbers in parentheses are standard errors. Reciprocity term omitted because network is undirected.
[a] "White" is reference category.
* p<.05. ** p<.01. *** p<.001.

Second, there are some instances in which coefficients for ethnic or micro-ethnic homophily changed in sign and/or direction—indicating that the very existence of in-group preference among members of certain ethnic categories depends on the nature of the tie involved. In particular, we note that Indian, Korean, Russian, South-East Asian, and mainstream black students all display significant homophily in Model 2 where they did not before; and the combination of a significant positive coefficient for white (racial) homophily and a significant negative coefficient for mainstream white (ethnic) homophily (though smaller in absolute value) suggests that white students who signal no particular ethnic identity still prefer friendships with each other compared to non-whites—but not as much as they prefer ties to their "ethnic" white counterparts.[2]

In Model 4, we again see significant racial differences in sociality, although here black students tend to have significantly *larger* networks than white students—almost as large as those of

[2] Specifically, controlling for other mechanisms, the increase in log-odds for the formation of a mainstream white/ethnic white tie is 0.22 (because only a "racially homophilous" tie is at stake), while this number is significantly lower (0.16 = 0.22-0.06) for a mainstream white/mainstream white tie.

"mixed" students—and white students tend to have the smallest networks. Because the network is undirected, however, it is impossible to determine whether these differences in sociality result from differences in "expansiveness" (the tendency to initiate friendships) or "popularity" (the tendency to receive friendship requests from others) or both. Finally, we see several instances in which homophily on other attributes becomes significant where it was not before: among foreign-born students as well as those from Massachusetts, California, and the Pacific region, and among fans of country as well as fans of Kurt Vonnegut. Again, however, this produces only modest decreases (and in the case of whites, a slight increase) in the coefficients for racial homophily.

Overall, then, we find confirmation of our primary results that the role of racial homophily in a network is at least distorted (due to differences in sociality) and most likely exaggerated (due to nested ethnic homophily and balancing mechanisms) if alternative mechanisms of tie formation are not disentangled from racial homophily proper. Due to their relatively small networks of Facebook friends, however (Model 4)—combined with the relative "aversion" we found among mainstream whites in Model 2—the role of racial homophily among whites may actually be *underestimated* as a contributor to racial homogeneity if controls for other mechanisms are not included. Interestingly, there is also an increase in the estimates of racial homophily among black students between Model 1 and Model 3, and no difference in the estimates for Asian homophily. The implications of this finding are unclear. It may be that in the case of Facebook friends, the strength—or even presence—of balancing mechanisms varies among racial groups; or it may be that the undirected nature of the data causes a number of possible triadic configurations in an underlying, "directed" social reality to become conflated, washing out the effects of "triadic closure" proper (see recent work by Robins et al. 2009).

We also find strikingly similar results for the replication of Model 6 on the Facebook friend network, presented in Table A2. Some caution should be exercised in parameter interpretation, as the MCMCMLE process did not converge after repeated attempts (likely due to the density of the network and complexity of the model), and so maximum pseudo-likelihood was used instead. Nonetheless, almost all parameter estimates are consistent with the original model in terms of significance and direction, except that sharing a major in general social science is no longer a statistically significant basis for tie formation (but sharing a major in applied mathematics is). We also observe that the propinquity effect of sharing a "neighborhood" is now significant, and the undirected higher-order term for accumulations of stars (GWD) is not. While the precise parameter estimates are not necessarily comparable with Table 2, we again note the important role of sharing a room or dorm building for tie formation (though this is not terribly surprising if Facebook friends are indeed capturing real-life “acquaintances”) and the enduring significance of racial homophily among black and Hispanic students in particular, even controlling for a variety of other mechanisms of tie formation.

Table A2

Replication of Model 6 for Facebook friendships

	Model 6
Edges	-3.56 (0.05)***
Racial homophily	
Whites	0.30 (0.02)***
Blacks	2.14 (0.05)***
Asians	0.48 (0.05)***
Hispanics	1.04 (0.08)***
Ethnic and micro-ethnic homophily	
South Asians	1.61 (0.13)***
East Asians	0.58 (0.06)***
Jews	0.70 (0.15)***
Vietnamese	2.45 (0.47)***
Homophily based on regional origin	
Hawaiians	13.71 (77.65)
Illinoisans	1.19 (0.14)***
Homophily based on socioeconomic status	
Graduates of “select 16” boarding schools	1.28 (0.15)***
Homophily based on shared cultural taste[a]	
Fans of Coldplay and Dave Matthews Band	0.14 (0.03)***
Fans of R&B, hip hop, and rap	0.61 (0.08)***
Shared foci based on academic major	
Economics	0.27 (0.04)***
General social science	0.18 (0.11)
Applied mathematics	0.98 (0.23)***
Micro-biology	0.94 (0.18)***
Propinquity due to co-residence	
Shared neighborhood	0.06 (0.02)**
Shared residence	1.58 (0.02)***
Shared room	5.63 (0.50)***
Sociality effects	
(20 sociality terms for various ethno-racial and other categories, not shown here)	
Balancing mechanisms and other higher-order terms	
Reciprocity	NA
Degree (GWD)	0.45 (1.70)
Triadic closure (GWESP)	1.08 (0.04)***
Two-paths (GWDSP)	-0.14 (0.00)***
AIC	132164

Note: Numbers in parentheses are standard errors. Reciprocity term omitted because network is undirected, and GWD (undirected) replaces GWOD.

[a] Listed tastes refer to the predominant favorites among students in a given subgroup.

* p<.05. ** p<.01. *** p<.001.

With respect to all of the above replication models, it is worthwhile to note in closing that there may be important and potentially undesirable consequences of taking a social process that is in fact dynamic—i.e. one student requests a friendship, and the second must confirm in order for the tie to appear at all—and analyzing only those "successful" cases of tie formation. This practice is also common in network studies based on Add Health data in which only mutual friendships are considered. While this may be theoretically desirable in some cases (e.g. to examine only the "strongest" ties) or practically necessary in others (e.g. data on unreciprocated ties are unavailable, as with Facebook friends, or methods have not yet been extended to directed networks, as was the case until recently with higher-order ERG specifications), future research should further explore the benefits and limitations of inducing undirected relational data from explicitly dynamic underlying processes—especially given the demonstrated importance of reciprocity as a basic principle of tie formation.

Online Supplement B: Some summary statistics

Table B1

Select descriptive statistics on picture-posting students

Variable	Value		%	N
Gender	Male		40.76	300
	Female		59.24	436
		Total	100.00	736
Race/ethnicity	White		60.87	448
	Black		8.97	66
	Asian		19.84	146
	“Mixed”		3.80	28
	Hispanic		6.39	47
	Non-identified		0.14	1
		Total	100.00	736
Socioeconomic status	Attended a “select 16” boarding school		4.21	31
	Did not attend a “select 16” boarding school		84.51	622
	Non-identified		11.28	83
		Total	100.00	736
Region of origin	New England		13.59	100
	Middle Atlantic		18.75	138
	East North Central		7.88	58
	West North Central		3.67	27
	South Atlantic		12.23	90
	East South Central		1.63	12
	West South Central		4.62	34
	Mountain		1.77	13
	Pacific		16.30	120
	International		7.61	56
	Non-identified		11.96	88
		Total	100.00	736

Note: The ethno-racial composition of the population is reproduced in greater detail in Figure 2 of the main text.

Online Supplement C: Quantitative tests of model fit

Following recent work on goodness-of-fit for ERG models (Hunter et al. 2008) and the example of other empirical studies (e.g. Goodreau 2007; Goodreau et al. 2009; see also Robins et al. 2009), we assess the fit of Model 6 by asking whether key global structural features of the observed network can be generated by the local mechanisms represented in our model. While the visual plots presented in the main text suggest an excellent fit of our model in most cases, it is helpful to also examine the more precise quantitative tests that complement these figures.

Table C1 presents summary statistics regarding our 100 simulations of Model 6 across the four dimensions selected above: out degree, in degree, edge-wise shared partners, and minimum geodesic distance. For each level of each distribution, the table presents the observed quantity of nodes (for in degree and out degree), friendships (for edge-wise shared partners), and dyads (for minimum geodesic distance) with the given value (note that the level of "Inf" for minimum geodesic distance signifies pairs of students who are "disconnected" in the network, i.e. they cannot be reached from each other). Next, the table presents the minimum, maximum, and average values of the same statistic over 100 simulations of Model 6. Finally, the table presents p-values comparing the observed statistic to the distribution generated by simulation. The idea behind these p-values is simply to calculate what proportion of the "null" distribution (in this case, represented by the 100 simulated values) is "at least as extreme" as the observed statistic. In other words, it is a test of the null hypothesis that says the observed statistic came from the same distribution as the simulated statistics.

Inspection of Table C1 shows that the p-values for the fit of Model 6 are consistent with our interpretation of the visual plots in the paper. This model does a particularly good job of capturing the in degree distribution, where in the entire distribution there is evidence that only *one* observed statistic (the number of students who "receive" exactly 15 friendships) does not come from the same distribution as the simulated statistics (i.e. all other p-values are greater than .05). P-values are also greater than .05 for almost all levels of the geodesic distribution (except for pairs of students separated by exactly 2, 3 or 5 friendships) and the out degree distribution (except for students who "send" exactly 1, 4, 6, 7, 35 or 39 friendships). Finally, as we note in the main text, the simulations of Model 6 substantially over-predict the number of friends with 0 or 1 friend in common and under-predict the number of friends with 2 to 5 friends in common, but there is no reason to reject the null hypothesis for the rest of the distribution.

Obviously, these are only four sets of statistics among the many possible with which our model could be compared against empirical reality. Further, we reiterate that even a perfect fit for all statistics in no way “proves” that the mechanisms represented in our model capture those actually at work among the students of this college. Nonetheless, given the undeniable complexity of the tie formation process as well as the demonstrated failure of past models to come anywhere close to reproducing actually observed network structures, the consistency of our model with the observed network on so many dimensions as well as the ongoing methodological development in this field shed promising light on the future of realistic models of social network formation.

Table C1

Quantitative fit assessment for Model 6

out degree																	
level	obs	min	mean	max	p-val	level	obs	min	mean	max	p-val	level	obs	min	mean	max	p-val
0	38	24	34.64	47	0.58	16	7	3	8.99	19	0.80	32	0	0	0.32	3	1.00
1	94	20	35.04	48	0.00	17	7	2	7.05	13	1.00	33	1	0	0.17	2	0.32
2	53	31	45.24	58	0.28	18	6	0	6.18	12	1.00	34	1	0	0.12	1	0.24
3	53	44	57.52	77	0.48	19	8	0	6.01	12	0.56	35	2	0	0.09	1	0.00
4	51	43	64.65	87	0.02	20	8	0	4.77	14	0.36	36	0	0	0.03	1	1.00
5	57	53	67.35	95	0.28	21	8	0	5	12	0.56	37	1	0	0.04	1	0.08
6	48	44	65.4	83	0.04	22	5	0	4.06	10	0.94	38	0	0	0.04	1	1.00
7	40	42	60.98	84	0.00	23	8	0	3.4	10	0.16	39	1	0	0.02	1	0.04
8	44	38	56.24	81	0.10	24	2	0	3.14	10	0.82	40	1	0	0.03	1	0.06
9	43	27	47.27	61	0.54	25	2	0	2.5	9	1.00	41	0	0	0.02	1	1.00
10	33	28	39.32	60	0.42	26	5	0	2.2	7	0.16	42	0	0	0.03	1	1.00
11	27	17	31.64	47	0.54	27	5	0	1.47	5	0.08	46	0	0	0.01	1	1.00
12	17	14	24.53	38	0.16	28	2	0	1.1	4	0.60	48	0	0	0.02	1	1.00
13	20	10	19.61	30	0.98	29	1	0	1.03	5	1.00	52	0	0	0.01	1	1.00
14	20	5	15.38	25	0.20	30	2	0	0.55	3	0.26	53	0	0	0.01	1	1.00
15	14	5	12.26	21	0.74	31	1	0	0.51	3	0.68	54	0	0	0.01	1	1.00
in degree																	
level	obs	min	mean	max	p-value	level	obs	min	mean	max	p-value	level	obs	min	mean	max	p-value
0	6	2	7.42	14	0.84	12	31	12	22.12	33	0.08	24	1	0	2.48	8	0.86
1	22	10	20.93	32	1.00	13	18	6	15.12	27	0.58	25	2	0	2.06	9	1.00
2	41	23	38.61	54	0.78	14	15	3	10.22	20	0.16	26	1	0	1.25	5	1.00
3	43	41	57.86	75	0.06	15	17	2	8.38	20	0.04	27	0	0	1.18	6	0.84
4	66	42	70.89	100	0.64	16	10	0	6.47	13	0.20	28	1	0	0.62	4	0.80
5	74	56	81.26	100	0.32	17	12	0	6.51	15	0.14	29	0	0	0.38	3	1.00
6	76	60	82.44	105	0.54	18	3	0	5.46	14	0.58	30	0	0	0.24	2	1.00
7	80	56	78.22	104	0.88	19	4	0	5.3	11	0.76	31	0	0	0.19	2	1.00
8	71	48	66.71	94	0.68	20	3	0	5.07	13	0.68	32	0	0	0.07	1	1.00
9	43	36	53.95	76	0.20	21	1	0	4.08	11	0.54	33	0	0	0.05	1	1.00
10	50	27	42.12	61	0.30	22	2	0	3.94	12	0.76	34	0	0	0.02	1	1.00
11	42	20	31.44	45	0.08	23	1	0	2.94	10	0.74						
edge-wise shared partners												minimum geodesic distance					
level	obs	min	mean	max	p-value	level	obs	min	mean	max	p-value	level	obs	min	mean	max	p-value
0	1497	1860	2037.42	2237	0.00	13	0	0	5.29	24	0.58	1	5641	4938	5496.79	5928	0.84
1	1473	1588	1761.4	1976	0.00	14	1	0	2.16	10	1.00	2	32828	24759	26464.17	28581	0.00
2	1031	495	651.49	774	0.00	15	0	0	0.97	9	1.00	3	129666	89610	99867.3	112361	0.00
3	701	179	260.78	353	0.00	16	1	0	0.34	4	0.48	4	199184	191645	208298.9	221721	0.12
4	480	39	164.17	255	0.00	17	0	0	0.14	6	1.00	5	102225	119676	133956.6	147346	0.00
5	222	11	144.65	237	0.02	18	0	0	0.03	1	1.00	6	25611	22144	28790.48	40221	0.40
6	137	0	135.2	222	0.66	19	0	0	0.02	2	1.00	7	4790	2072	3846.76	7118	0.36
7	50	0	115.49	208	0.52							8	893	74	456.97	1700	0.20
8	31	0	88.94	182	0.54							9	115	0	49.49	642	0.20
9	11	0	58.57	132	0.54							10	7	0	3.75	69	0.24
10	3	0	38.08	94	0.56							11	0	0	0.28	6	1.00
11	1	0	20.92	57	0.54							12	0	0	0.02	2	1.00
12	2	0	10.73	44	0.70							Inf	40000	21194	33728.52	44276	0.24