Supplemental Material

1. Estimating Networks Under PCHM With Various Parameter Settings

1.1 Varying Link Density

In this section, we repeat the analysis performed in Section 5 while varying parameter settings. Firstly, we change the link density, p, of the adjacency matrix to be 0.75 and 1 while holding $n_o = 8$. The generation procedure of G is the same as described in Section 5.2.2.

Table 1: Comparison of Estimation Error from Hub Model and Penalized Component Hub Model ($n_o = 8$, $p = 0.75$)

	MAE(A)		MAE(ρ)	
	HM	PCHM	HM	PCHM
Obs	Avg (StDev)	Avg(StDev)	Avg(StDev)	Avg(StDev)
100	0.1737 (0.0143)	0.1405 (0.0203)	0.0282 (0.0041)	0.0239 (0.0050)
200	0.1594 (0.0145)	0.0860 (0.0221)	0.0198 (0.0039)	0.0138 (0.0048)
500	0.1380 (0.0127)	0.0232 (0.0173)	0.0080 (0.0027)	0.0039 (0.0030)
1000	0.1321 (0.0117)	0.0097 (0.0090)	0.0039 (0.0013)	0.0018 (0.0014)
2000	0.1299 (0.0107)	0.0059 (0.0059)	0.0022 (0.0009)	0.0011 (0.0009)
5000	0.1271 (0.0119)	0.0037 (0.0061)	0.0011 (0.0006)	0.0007 (0.0006)
10000	0.1259 (0.0110)	0.0026 (0.0051)	0.0007 (0.0006)	0.0005 (0.0006)

Table 2: Comparison of Model Selection from Hub Model and Penalized Component Hub Model ($n_o = 8$, $p = 0.75$)

	Estimated n_o		Estimated d		η
	HM	PCHM	HM	PCHM	PCHM
Obs	Avg (StDev)	Avg(StDev)	Avg (StDev)	Avg(StDev)	Avg(StDev)
100	33.9000 (2.1148)	23.4850 (2.9569)	592.7800 (72.1226)	290.8650 (70.6199)	11.4550 (2.5174)
200	36.2700 (2.2611)	16.6600 (3.0365)	677.4350 (83.0795)	150.6950 (53.2094)	11.0650 (2.5109)
500	37.1950 (2.0781)	9.7000 (2.0349)	711.4800 (78.2131)	52.9550 (24.5867)	8.8250 (2.6580)
1000	38.0050 (2.1490)	8.5050 (1.1561)	742.4900 (82.4541)	40.0850 (12.2740)	6.8800 (2.2641)
2000	38.6600 (2.0654)	8.2350 (0.7228)	767.7500 (79.8226)	37.2850 (7.3733)	4.9750 (1.6826)
5000	39.4100 (1.9803)	8.2000 (1.1253)	797.2300 (78.6053)	37.3500 (15.0819)	3.7025 (0.3721)
10000	39.7933 (1.8293)	8.1467 (1.1723)	812.3133 (73.6353)	36.9400 (16.8818)	3.4667 (0.3411)

Table 3: Comparison of Estimation Error from Hub Model and Penalized Component Hub Model ($n_o = 8$, $p = 1$)

	MAE(A)		MAE(ρ)	
	HM	PCHM	HM	PCHM
Obs	Avg (StDev)	Avg(StDev)	Avg(StDev)	Avg(StDev)
100	0.2342 (0.0148)	0.1962 (0.0226)	0.0305 (0.0033)	0.0299 (0.0042)
200	0.2223 (0.0143)	0.1365 (0.0247)	0.0255 (0.0043)	0.0236 (0.0046)
500	0.1971 (0.0118)	0.0662 (0.0270)	0.0136 (0.0032)	0.0130 (0.0060)
1000	0.1862 (0.0113)	0.0438 (0.0271)	0.0066 (0.0019)	0.0036 (0.0033)
2000	0.1802 (0.0103)	0.0216 (0.0219)	0.0037 (0.0015)	0.0018 (0.0019)
5000	0.1764 (0.0102)	0.0088 (0.0164)	0.0018 (0.0009)	0.0009 (0.0010)
10000	0.1746 (0.0113)	0.0080 (0.0236)	0.0012 (0.0010)	0.0007 (0.0010)

Table 4: Comparison of Model Selection from Hub Model and Penalized Component Hub Model ($n_o = 8$, $p = 1$)

	Estimated n_o		Estimated d		η
	HM	PCHM	HM	PCHM	PCHM
Obs	Avg (StDev)	Avg(StDev)	Avg (StDev)	Avg(StDev)	Avg(StDev)
100	37.2900 (2.1977)	24.2950 (3.9581)	715.3200 (83.0604)	314.0650 (95.8730)	12.0000 (2.5153)
200	40.3100 (1.8277)	16.7750 (3.3988)	833.2650 (74.7636)	153.8350 (60.4515)	11.5675 (2.5646)
500	42.3250 (1.8292)	11.8150 (3.1544)	917.5300 (77.9459)	79.6550 (44.3501)	7.9625 (2.5774)
1000	42.9650 (1.7171)	11.5400 (2.9516)	944.9450 (74.2554)	75.6900 (38.9952)	5.2825 (1.0144)
2000	43.9550 (1.4538)	9.6050 (2.2902)	988.0500 (64.2115)	52.5400 (28.2578)	4.8175 (0.9407)
5000	44.8600 (1.5947)	8.5300 (1.8126)	1028.9050 (72.1769)	41.2800 (25.4035)	4.1550 (0.6368)
10000	45.3285 (1.3991)	8.6277 (2.9729)	1049.9708 (64.2236)	44.9197 (55.8912)	3.7701 (0.3930)

Tables 1-4 show the comparison of estimation errors and model selection for both cases. The results show the same pattern as in Tables 6 and 7 of the main article. That is, the error in A and ρ under PCHM is quickly shrunken and is uniformly smaller than the error under HM. And PCHM can identify a much sparser model and the accuracy of estimated n_o increases with the sample size T.

In addition, the performance in general is getting worse as the link density increases. The reason is that the average group size becomes larger as the network becomes denser, which makes identification of group leaders more difficult and thus affects the parameter estimation and model selection.

1.2 Varying the Number of Leaders

Next we perform a simulation by keeping the link density as 0.5 but changing n_o to be 16 and 32. The results are summarized in Tables 5-8. As in the previous section, PCHM generally yields better parameter estimation on A and can detect the correct n_0 with a moderate number of observations. The estimation of ρ from HM outperforms the estimation from PCHM when the sample size is very small. This is due to the fact that the penalty on ρ from PCHM introduces bias. However, the estimation of A is less affected because PCHM does not directly penalize A.

Note that as n_o gets closer to n, the performance of PCHM begins to resemble that of HM. Recall from Equation (7) of the main article that the reduction of estimated elements does not decrease linearly with the reduction of n_o. Therefore, when $n_o = 32$, HM estimates almost as many elements as PCHM.

Table 5: Comparison of Estimation Error from Hub Model and Penalized Component Hub Model ($n_o = 16$, $p = 0.5$)

	MAE(A)		MAE(ρ)	
	HM	PCHM	HM	PCHM
Obs	Avg (StDev)	Avg(StDev)	Avg(StDev)	Avg(StDev)
100	0.1239 (0.0127)	0.1112 (0.0133)	0.0218 (0.0030)	0.0219 (0.0031)
200	0.0983 (0.0111)	0.0757 (0.0133)	0.0147 (0.0025)	0.0138 (0.0031)
500	0.0671 (0.0080)	0.0194 (0.0071)	0.0050 (0.0011)	0.0038 (0.0014)
1000	0.0627 (0.0066)	0.0098 (0.0010)	0.0028 (0.0004)	0.0021 (0.0004)
2000	0.0604 (0.0066)	0.0070 (0.0011)	0.0018 (0.0003)	0.0014 (0.0003)
5000	0.0578 (0.0065)	0.0044 (0.0006)	0.0010 (0.0002)	0.0009 (0.0002)
10000	0.0560 (0.0064)	0.0031 (0.0002)	0.0007 (0.0001)	0.0006 (0.0001)

Table 6: Comparison of Model Selection from Hub Model and Penalized Component Hub Model ($n_o = 16$, $p = 0.5$)

	Estimated n_o		Estimated d		η
	HM	PCHM	HM	PCHM	PCHM
Obs	Avg (StDev)	Avg(StDev)	Avg (StDev)	Avg(StDev)	Avg(StDev)
100	33.5950 (2.2239)	26.2500 (1.9355)	582.5700 (75.9285)	358.5200 (51.8661)	10.1975 (3.0162)
200	35.8800 (2.1491)	23.9250 (2.1453)	662.9250 (77.8106)	299.4550 (52.3889)	9.8350 (2.8633)
500	34.8150 (2.2553)	16.9050 (1.2901)	624.9800 (79.9768)	151.1700 (23.9367)	7.8375 (2.4811)
1000	35.4200 (2.0505)	16.0250 (0.1565)	646.0900 (73.4720)	135.4250 (2.6608)	4.7375 (1.5898)
2000	36.1200 (1.7030)	16.0350 (0.1842)	670.8300 (62.3208)	135.5950 (3.1321)	3.6175 (0.8980)
5000	36.7750 (1.6577)	16.0300 (0.1710)	694.9550 (61.5129)	135.5100 (2.9073)	2.9775 (0.3020)
10000	36.7650 (1.7042)	16.0000 (0.0000)	694.6600 (63.3309)	135.0000 (0.0000)	2.8000 (0.2923)

Table 7: Comparison of Estimation Error from Hub Model and Penalized Component Hub Model ($n_o = 32$, $p = 0.5$)

	MAE(A)		MAE(ρ)	
	HM	PCHM	HM	PCHM
Obs	Avg (StDev)	Avg(StDev)	Avg(StDev)	Avg(StDev)
100	0.1445 (0.0115)	0.1430 (0.0119)	0.0183 (0.0019)	0.0212 (0.0020)
200	0.1078 (0.0097)	0.1099 (0.0117)	0.0142 (0.0016)	0.0180 (0.0025)
500	0.0444 (0.0049)	0.0355 (0.0045)	0.0048 (0.0008)	0.0053 (0.0011)
1000	0.0336 (0.0034)	0.0206 (0.0012)	0.0031 (0.0004)	0.0030 (0.0004)
2000	0.0285 (0.0032)	0.0143 (0.0008)	0.0021 (0.0003)	0.0021 (0.0003)
5000	0.0241 (0.0029)	0.0090 (0.0004)	0.0013 (0.0002)	0.0013 (0.0002)
10000	0.0211 (0.0028)	0.0064 (0.0004)	0.0009 (0.0001)	0.0009 (0.0001)

Table 8: Comparison of Model Selection from Hub Model and Penalized Component Hub Model ($n_o = 32$, $p = 0.5$)

	Estimated n_o		Estimated d		η
	HM	PCHM	HM	PCHM	PCHM
Obs	Avg (StDev)	Avg(StDev)	Avg (StDev)	Avg(StDev)	Avg(StDev)
100	35.8000 (1.8968)	30.3350 (1.8411)	659.5100 (69.0910)	475.9600 (56.4881)	9.4850 (3.3971)
200	39.8000 (1.9336)	31.7600 (2.3601)	812.7800 (78.2581)	522.0000 (76.7643)	8.9225 (3.3102)
500	38.9250 (1.7621)	32.6350 (1.1699)	777.5850 (69.4107)	548.5200 (39.0467)	5.4225 (1.8233)
1000	38.9400 (1.4756)	32.0400 (0.1965)	777.7150 (58.1947)	528.3200 (6.4829)	3.1225 (0.9491)
2000	39.6300 (1.2811)	32.0100 (0.0997)	804.9000 (51.5175)	527.3300 (3.2917)	2.7300 (0.7296)
5000	40.2700 (1.4690)	32.0000 (0.0000)	831.0450 (59.9956)	527.0000 (0.0000)	2.4050 (0.4896)
10000	40.1050 (1.3009)	32.0100 (0.0997)	824.1000 (52.9504)	527.3300 (3.2917)	2.1775 (0.2831)

2. Jester Dataset

Jester is a joke recommender system developed at UC Berkeley to study social information filtering. Between April 1999 and May 2003, researchers collected a dataset from 24,983 users who rated 36 or more of a set of 100 jokes on a continuous scale (-10 to +10).

For our analysis, joke ratings have been converted to binary results. Let r_{tj} be the rating that the t^{th} user gave the j^{th} joke.

$$G_j^{(t)} = \begin{cases} 1 & \text{if } r_{tj} > 0 \\ 0 & \text{otherwise} \end{cases}$$

This translation converts user ratings into collections of jokes which the users preferred.

The task of inferring humor preferences from user ratings is a challenging problem because the human sense of humor is based on a mix of social, emotional, and intellectual characteristics which are further influenced by gender, age, and upbringing (**?**). Here we follow the example of **?** by focusing on joke preference rather than joke content.

Additionally, this dataset is interesting because the average group size is large. As discussed earlier, datasets with numerous singletons resist sparsity.

In this section, we focus on the effect that increasing the number of observations has on the degree of simplification that can be achieved. For example, if new information is continually arriving, an important question to address is up to what value of T that PCHM is beneficial and beyond which a researcher might transition to using HM.

In Table 9, we show the optimal η and the associated n_o based on the minimum BIC. Along with this we show the minimum number of non-zero elements in ρ for all η.

There are several important points to make about Table 9. First, we focus on the column of "Optimal η". Notice that when the number of observation is low (up to 1000), optimal η selected by BIC gives a small n_o. It suggests that the reduction of parameters can improve the performance of the model to a large extent when the sample size is relatively small.

As the number of observations increases, there comes a point ($T = 2000$) where the optimal η becomes 1. This demonstrates a useful attribute of PCHM. In simulations, we showed that when the underlying network structure is sparse PCHM detects the true structure with relatively few observations and continues to do so as the number of observations increases. Here, where the assumption of a sparse structure may not be valid, PCHM provides feedback about the point at which PCHM ceases to benefit the analysis.

Now we shift our attention to the column of "Minimum n_o". In addition to n_o corresponding to the optimal η selected by BIC, we also give the minimum n_o which could be obtained by PCHM. This is the minimum number of leaders over the complete range of η. It shows that when $n = 2000$ and $n = 5000$, the number of parameters can still be heavily shrunken by PCHM with a larger η even though this is not optimal based on the BIC. It suggests that if a researcher were interested in using an alternative criterion to BIC, PCHM is still capable of significantly simplifying the model.

Table 9: Effect of increasing observations on optimal solutions

	Optimal η		Minimum n_o
Obs	η	n_o	n_o
100	13.8	35	35
200	19.2	19	18
500	17.6	17	17
1000	15.9	18	17
2000	1.0	77	23
5000	1.0	83	27