
Network Inference From Grouped Observations
Using Hub Models

YUNPENG ZHAO¹ AND CHARLES WEKO²

George Mason University¹ and United States Army²

Supplementary Material

S1 Identifiability

In Section 3.2, we introduced the notion of identifiability. For Hub Models, this means for any two sets of parameters $\{A, \rho\}$ and $\{A^*, \rho^*\}$:

$$\mathbb{P}(G = g|A, \rho) = \mathbb{P}(G = g|A^*, \rho^*) \quad \forall g \implies A = A^*, \rho = \rho^*. \quad (\text{S1.1})$$

In this section, we provide a simple counterexample to show that Hub Models are not identifiable if the adjacency matrix is unconstrained and prove Theorem 1.

6 **S1.1 Counterexample**

7 Consider a network of size $n = 4$ with the parameters defined in Table 1.
 8 Note that for the asymmetric case, $n = 4$ is the smallest population size
 9 where the number of possible groups exceeds the number of parameters to
 10 estimate. In this example, nodes v_1 and v_2 both always produce the same
 11 group while nodes v_3 and v_4 produce a different group.

		A_{ij}			
		j			
ρ_i	i	1	2	3	4
0.25	1	1	1	1	0
0.25	2	1	1	1	0
0.25	3	0	1	1	1
0.25	4	0	1	1	1

Table 1: Example of a Set of Parameters Which are Not Identifiable

12 The probability of G has the form:

$$\mathbb{P}(G = g|A, \rho) = \begin{cases} \frac{1}{2} & g = \{1, 1, 1, 0\}, \\ \frac{1}{2} & g = \{0, 1, 1, 1\}, \\ 0 & \text{otherwise.} \end{cases}$$

13 There are an infinite number of parameters yielding the same distribu-
 14 tion, but a simple alternative is as follows. Let $\rho^* = (0.5, 0, 0.5, 0)$, leave the
 15 first and third rows of A unchanged but let all other components of A^* as-
 16 sume arbitrary values. Obviously, we have $\mathbb{P}(G = g|A, \rho) = \mathbb{P}(G = g|A^*, \rho^*)$

17 for all g . This counterexample demonstrates that the model requires an ad-
 18 ditional condition to be identifiable.

19 **S1.2 Theorem Proof**

20 Theorem 1 states that symmetry of the adjacency matrix is a sufficient
 21 condition for identifiability.

22 **Theorem 1.** Let A and A^* be symmetric adjacency matrices with $A_{ii} =$
 23 $A_{ii}^* = 1$ for all i , $A_{ij} < 1$ and $A_{ij}^* < 1$ for all $i \neq j$. If $\mathbb{P}(g|A, \rho) = \mathbb{P}(g|A^*, \rho^*)$
 24 for all g , then $\{A, \rho\} = \{A^*, \rho^*\}$.

25 Let g^x and g^y denote the singleton groups which consist only of nodes
 26 v_x and v_y , respectively. Further, let g^{xy} denote the group representing the
 27 pair of v_x and v_y .

From (3.4) the probability of the singletons is:

$$\mathbb{P}(g^x|A, \rho) = \rho_x(1 - A_{xy}) \prod_{j \neq \{x,y\}} (1 - A_{xj}) \quad (\text{S1.2})$$

$$\mathbb{P}(g^y|A, \rho) = \rho_y(1 - A_{xy}) \prod_{j \neq \{x,y\}} (1 - A_{yj}). \quad (\text{S1.3})$$

28 In (S1.3) we have taken advantage of the symmetry of A to replace A_{yx}
 29 with A_{xy} .

Now, we consider the probability of g^{xy} .

$$\begin{aligned}
 \mathbb{P}(g^{xy}|A, \rho) &= \rho_x A_{xy} \prod_{j \neq \{x,y\}} (1 - A_{xj}) + \rho_y A_{xy} \prod_{j \neq \{x,y\}} (1 - A_{yj}) \\
 &= A_{xy} \left[\rho_x \prod_{j \neq \{x,y\}} (1 - A_{xj}) + \rho_y \prod_{j \neq \{x,y\}} (1 - A_{yj}) \right] \\
 &= A_{xy} \left[\frac{\mathbb{P}(G = g^x|A, \rho)}{(1 - A_{xy})} + \frac{\mathbb{P}(g^y|A, \rho)}{(1 - A_{xy})} \right] \\
 &= \frac{A_{xy}}{(1 - A_{xy})} \left[\mathbb{P}(g^x|A, \rho) + \mathbb{P}(g^y|A, \rho) \right], \tag{S1.4}
 \end{aligned}$$

which implies that:

$$A_{xy} = \frac{\mathbb{P}(g^{xy}|A, \rho)}{\mathbb{P}(g^x|A, \rho) + \mathbb{P}(g^y|A, \rho) + \mathbb{P}(g^{xy}|A, \rho)}. \tag{S1.5}$$

30 Therefore, $A_{xy} = A_{xy}^*$ for all x and y .

To complete the proof, consider an arbitrary node v_x which appears as a singleton represented by g^x :

$$\mathbb{P}(g^x|A, \rho) = \rho_x \prod_{j \neq x} (1 - A_{xj}). \tag{S1.6}$$

If $A_{xy} = A_{xy}^*$ for all x and y and $\mathbb{P}(g|A, \rho) = \mathbb{P}(g|A^*, \rho^*)$ for all g , then:

$$\rho_x \prod_{j \neq x} (1 - A_{xj}) = \rho_x^* \prod_{j \neq x} (1 - A_{xj}) \tag{S1.7}$$

31 and it is easy to see that $\rho_x = \rho_x^*$ for all x .

32 \square

33 Equation (S1.5) suggests a method of moments estimator for the adja-
 34 cency matrix based on frequencies of doubletons and singletons. However,
 35 this estimator requires that the probability of doubletons and singletons
 36 be estimated accurately. In practice, this technique would be very ineffi-
 37 cient, because small groups appear infrequently in many real-world datasets.
 38 Thus, we continue to focus on the MLE which presumably uses all available
 39 information.

40 S2 Estimating Equations

41 In Section 3.3, we presented estimating equations for Hub Model parameters
 42 ((3.8) and (3.9)). Here we derive those equations.

43 We begin by taking the derivative of (3.5) with respect to A_{xy} and A_{yx} .

$$\frac{\partial \Lambda(G|A, \rho)}{\partial A_{xy}} = \frac{\partial \mathcal{L}(G|A, \rho)}{\partial A_{xy}} - \lambda_{xy} = 0 \text{ if } x < y, \quad (\text{S2.1})$$

$$\frac{\partial \Lambda(G|A, \rho)}{\partial A_{yx}} = \frac{\partial \mathcal{L}(G|A, \rho)}{\partial A_{yx}} + \lambda_{xy} = 0 \text{ if } x > y. \quad (\text{S2.2})$$

Therefore,

$$\frac{\partial \mathcal{L}(G|A, \rho)}{\partial A_{xy}} = -\frac{\partial \mathcal{L}(G|A, \rho)}{\partial A_{yx}}. \quad (\text{S2.3})$$

We now focus on the derivative of the log likelihood function of (3.5):

$$\sum_t \frac{\rho_x G_x^{(t)} \left[\prod_{j \neq y} A_{xj}^{G_j^{(t)}} (1 - A_{xj})^{1-G_j^{(t)}} \right] \frac{\partial}{\partial A_{xy}} \left(A_{xy}^{G_y^{(t)}} (1 - A_{xy})^{1-G_y^{(t)}} \right)}{\sum_{i=1}^n \rho_i G_i^{(t)} \prod_j A_{ij}^{G_j^{(t)}} (1 - A_{ij})^{1-G_j^{(t)}}}. \quad (\text{S2.4})$$

44 Note that the derivative in the numerator of (S2.4) is equal to 1 if node
 45 v_y is in observation $G^{(t)}$, and -1 if v_y is not in the observation. We represent
 46 this by the function:

$$\gamma(G_y^{(t)}) = \begin{cases} 1 & \text{if } G_y^{(t)} = 1, \\ -1 & \text{if } G_y^{(t)} = 0. \end{cases}$$

Therefore,

$$\frac{\partial}{\partial A_{xy}} \mathcal{L}(G|A, \rho) = \sum_t \frac{\left[\rho_x G_x^{(t)} \prod_{j \neq y} A_{xj}^{G_j^{(t)}} (1 - A_{xj})^{1-G_j^{(t)}} \right] \gamma(G_y^{(t)})}{\sum_{i=1}^n \rho_i G_i^{(t)} \prod_j A_{ij}^{G_j^{(t)}} (1 - A_{ij})^{1-G_j^{(t)}}}. \quad (\text{S2.5})$$

The denominator of (S2.5) is simply the probability of $G^{(t)}$ (see (3.4)).

In addition, the term in brackets can be made equal to $\mathbb{P}(G^{(t)}, S_x = 1)$ by multiplying $A_{xy}^{G_y^{(t)}} (1 - A_{xy})^{(1-G_y^{(t)})}$. To conserve space, we suppress $\{A, \rho\}$

going forward. This gives:

$$\frac{\partial}{\partial A_{xy}} \mathcal{L}(G) = \sum_t \frac{\gamma(G_y^{(t)}) \mathbb{P}(G^{(t)}, S_x^{(t)} = 1)}{A_{xy}^{G_y^{(t)}} (1 - A_{xy})^{(1-G_y^{(t)})} \mathbb{P}(G^{(t)})}. \quad (\text{S2.6})$$

This equation can be further simplified by observing that $\frac{\mathbb{P}(G^{(t)}, S_x^{(t)} = 1)}{\mathbb{P}(G^{(t)})}$ is equivalent to $\mathbb{P}(S_x^{(t)} = 1 | G^{(t)})$:

$$\frac{\partial}{\partial A_{xy}} \mathcal{L}(G) = \sum_t \frac{\gamma(G_y^{(t)}) \mathbb{P}(S_x^{(t)} = 1 | G^{(t)})}{A_{xy}^{G_y^{(t)}} (1 - A_{xy})^{(1-G_y^{(t)})}}. \quad (\text{S2.7})$$

Plugging (S2.7) into (S2.3), we get:

$$\sum_t \frac{\gamma(G_y^{(t)}) \mathbb{P}(S_x^{(t)} = 1 | G^{(t)})}{A_{xy}^{G_y^{(t)}} (1 - A_{xy})^{(1-G_y^{(t)})}} = - \sum_t \frac{\gamma(G_x^{(t)}) \mathbb{P}(S_y^{(t)} = 1 | G^{(t)})}{A_{yx}^{G_x^{(t)}} (1 - A_{yx})^{(1-G_x^{(t)})}}. \quad (\text{S2.8})$$

By applying symmetry and breaking the summations, this becomes:

$$\begin{aligned} & \sum_{t:G_y^{(t)}=1} \frac{\mathbb{P}(S_x^{(t)} = 1 | G^{(t)})}{A_{xy}} - \sum_{t:G_y^{(t)}=0} \frac{\mathbb{P}(S_x^{(t)} = 1 | G^{(t)})}{1 - A_{xy}} \\ &= - \sum_{t:G_x^{(t)}=1} \frac{\mathbb{P}(S_y^{(t)} = 1 | G^{(t)})}{A_{xy}} + \sum_{t:G_x^{(t)}=0} \frac{\mathbb{P}(S_y^{(t)} = 1 | G^{(t)})}{1 - A_{xy}}. \end{aligned} \quad (\text{S2.9})$$

With some simple algebra, it is easy to see that:

$$\hat{A}_{xy} = \frac{\sum_t G_y^{(t)} \mathbb{P}(S_x = 1 | G^{(t)}) + \sum_t G_x^{(t)} \mathbb{P}(S_y = 1 | G^{(t)})}{\sum_t [\mathbb{P}(S_x = 1 | G^{(t)}) + \mathbb{P}(S_y = 1 | G^{(t)})]}. \quad (\text{S2.10})$$

47 It is worth repeating that (S2.10) is not a closed form solution for \hat{A}_{xy} .
 48 This is because the right hand side of the equation depends on \hat{A}_{xy} .

We next derive the estimating equation for $\hat{\rho}$. By taking the derivative of (3.6) with respect to ρ_x , we get the following:

$$\begin{aligned} \frac{\partial}{\partial \rho_x} \Lambda(G) &= \sum_t \frac{G_x^{(t)} \prod_j A_{xj}^{G_j^{(t)}} (1 - A_{xj})^{1-G_j^{(t)}}}{\mathbb{P}(G^{(t)})} - \lambda_o \\ &= \sum_t \frac{\mathbb{P}(G^{(t)}, S_x^{(t)} = 1)}{\rho_x \mathbb{P}(G^{(t)})} - \lambda_o \\ &= \frac{1}{\rho_x} \sum_t \mathbb{P}(S_x^{(t)} = 1 | G^{(t)}) - \lambda_o. \end{aligned} \quad (\text{S2.11})$$

Solving this equation for zero, we obtain:

$$\rho_x = \frac{1}{\lambda_o} \sum_t \mathbb{P}(S_x^{(t)} = 1 | G^{(t)}). \quad (\text{S2.12})$$

Using the constraint on ρ (3.3), we get:

$$\hat{\rho}_x = \frac{\sum_{t=1}^T \mathbb{P}(S_x^{(t)} = 1 | G^{(t)})}{T}. \quad (\text{S2.13})$$

49 **S3 Data Analysis**

50 In Section 6, we performed analysis on a dataset from *Dream of the Red*
 51 *Chamber* to show how Hub Models can provide sharper contrast between

52 the relationships with a population. Here we present analysis of two addi-
53 tional datasets.

54 The first dataset records co-sponsorship of legislation in the Senate of
55 the 110th United States Congress. The rules of the Senate require that each
56 piece of legislation have a unique sponsor; however, other members may co-
57 sponsor the bill (Fowler, 2006a). These rules mean that the data conform
58 to the assumption of the Hub Model.

59 The second dataset has been extracted from the USDA plant database.
60 Unlike the other two datasets, this one does not deal with “social” data, but
61 with “spatial” data. Each observation represents a single species of plant
62 along with each North American state or territory in which the plant is ob-
63 served to grow. For this analysis, states and territories represent nodes. As
64 with the *Dream of the Red Chamber* dataset, we find that the Hub Models
65 return meaningful information about underlying structure even when the
66 assumption of a single hub node is not valid.

67 **S3.1 Senate of the 110th United States Congress**

68 The first dataset records co-sponsorship of legislation in the Senate of the
69 110th United States Congress. The rules of the Senate require that each
70 piece of legislation have a single, unique sponsor; however, other members

71 may co-sponsor the bill (Fowler, 2006a). These rules mean that the data
72 conform to the assumption of the Hub Model.

73 The United States Senate is a chamber in the bicameral legislature of
74 the United States, and together with the U.S. House of Representatives
75 makes up the U.S. Congress. A key function of both chambers of Congress
76 is to originate legislation. Each piece of legislation can have only one orig-
77 inating sponsor; however, since the mid-1930s, Senators have had an op-
78 portunity to express support for a piece of legislation by signing it as a
79 co-sponsor (Fowler, 2006a).

80 The 110th United States Congress occurred between January 3, 2007
81 and January 3, 2009. The Democratic Party controlled a majority in both
82 chambers for the first time since 1995 with a voting share of 50.5 % of the
83 Senate membership.

84 The United States Senate consists of 100 members with each state rep-
85 resented by 2 Senators at any time. During this session of Congress, there
86 were a total of 102 individuals who served in the Senate. One original mem-
87 ber died and a second resigned to become a lobbyist. Both members were
88 replaced by appointed state representatives.

89 Data for legislative co-sponsorship are available in the Library of Congress
90 Thomas legislative database. This database includes more than 280,000

91 pieces of legislation proposed in the U.S. House and Senate with over 2.1
92 million co-sponsorship signatures. Most bills do not pass, and cosponsors
93 need not invest time and resources crafting legislation; so co-sponsorship
94 is a relatively costless way to signal one’s position on issues important to
95 constituents and fellow legislators. For the purposes of this study, we in-
96 clude all forms of legislation including all available resolutions, public and
97 private bills, and amendments (Fowler, 2006b). During the 110th Congress,
98 the Senate initiated 10,327 pieces of legislation.

99 It is a trivial task to apply the KHM to this dataset when we treat
100 the sponsor as known; therefore, we focus on the case where the sponsor is
101 unknown. That is, we intentionally confound sponsors with co-sponsors so
102 that the only data that we have is G . We would like to investigate whether
103 the HM can provide a meaningful estimate of the latent social structure
104 even when the information of hub nodes is missing. The average difference
105 between edges estimated by KHM and HM is 0.03, which suggests that the
106 HM estimate is very accurate even when we confound the hub nodes.

107 In Figure 1, we plot the co-occurrence matrix, half-weight index, and the
108 adjacency matrix of the Hub Model using the force directed graph drawing
109 technique of Fruchterman-Reingold. Each Senator is represented as a node
110 where the color of the node represents the Senator’s official political party.

111 Red nodes represent Republicans while blue nodes represent Democrats.

112 First, notice that each estimated adjacency matrix produces a different
 113 layout because of the different weights estimated for each relationship. For
 114 the co-occurrence matrix and the half weight index, the estimated adjacency
 115 matrices do not separate Republican and Democrat Senators while the Hub
 116 Model produces an adjacency matrix which clusters Senators by political
 117 party.

118 Second, in Figure 1, we plot only the 5% strongest relationships in each
 119 estimate. Again, the Hub Models provide insight into the individuals in
 120 the population which have key relationships. In Figure 1c, individuals with
 121 many strong connections are closer to the center of the layout while in
 122 Figures 1a and 1b a number of these individuals are actually present at the
 123 perimeter of the layout. We have not presented all relationships because
 124 this results in a nearly complete graph.

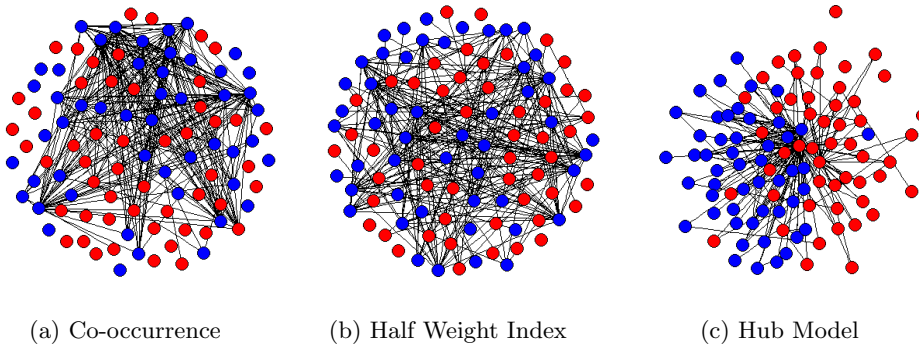


Figure 1: Comparison of Estimation Techniques for the 110th Senate

This implies that the HM provides more meaningful information about the community structure of this dataset than classical measures. This is further supported by the normalized cut value corresponding to the Senator’s official party membership (Shi and Malik, 2000) .

$$\frac{\sum_{i \in C_1, j \in C_2} A_{ij}}{\sum_{i, j \in C_1} A_{ij}} + \frac{\sum_{i \in C_1, j \in C_2} A_{ij}}{\sum_{i, j \in C_2} A_{ij}}. \quad (\text{S3.14})$$

Equation (S3.14) gives the normalized cut value for two communities C_1 and C_2 . Lower normalized cut values indicate stronger community differentiation. For the Senate data, Table 2 presents the values for each inferred network. The normalized cut value for HM is lower than the co-occurrence matrix and half weight index, which strengthens the visual intuition that the estimates from HM provide better distinction between communities.

	Normalized Cut Value
Co-Occurrence Matrix	0.837
Half Weight Index	0.823
Hub Model	0.757

Table 2: Normalized Cut Ratios of 110th Senate for Different Inference Techniques

130

131 **S3.2 North American Flora**

132 The final dataset has been extracted from the USDA plant database. In
133 the previous examples, we have worked with datasets which are essentially

134 from the social sciences. However, we believe that Hub Models are useful in
135 other situations where observations are the result of nodes coalescing around
136 a single node or observations are the result of some resource dispersing
137 outward from a single node to multiple nodes.

138 As a demonstration of how this kind of data can be used to estimate the
139 relationship between different regions, we use a dataset from the University
140 of California Irvine Machine Learning Repository which had been extracted
141 and encoded from the USDA plants database (Hamalainen and Nykanen,
142 2008). 34,781 plant species or geneses are included in the dataset. For each
143 plant, the dataset indicates in which of 68 areas the plant is found. These
144 areas include all United States states, Canadian provinces and territories,
145 along with the Virgin Islands, Puerto Rico, Greenland, and St. Pierre and
146 Miquelon (islands off the northeast coast of Canada). For simplicity, we
147 will refer to all of these areas as “states”.

148 We would expect that contiguous states would tend to have many flora
149 in common while states which are far apart would be less likely to share com-
150 mon flora. For example, Connecticut and Massachusetts are small states
151 which share a common boarder; therefore, we would expect them to appear
152 together in the regions of many plants. Conversely, California and Green-
153 land are very far apart and at different latitudes; therefore, we would expect

154 a weak relationship.

155 Of course, the map of North America is well known and our objective
 156 here is not to compare the Hub Model to spacial modeling. Instead, we are
 157 using the regions of North America as a proxy for a system of distribution.

158 To demonstrate the ability of the Hub Model to capture the connections
 159 between states, we split them into 13 different regions. The United States
 160 is identified by the 9 divisions of the US Census Bureau. The Canadian
 161 provinces are identified by three regions. The final region includes islands
 162 in the Atlantic ocean which are not included in any other region.

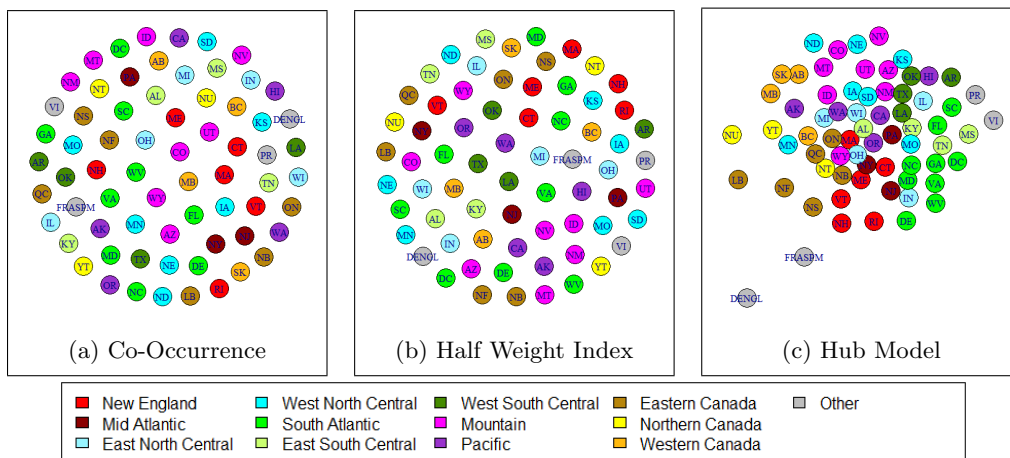


Figure 2: Adjacency Matrix Estimates for North American Flora Data

163 In Figure 2, we apply a force directed graph drawing technique to layout
 164 the states according to the estimated adjacency matrices of each approach.
 165 The HM graph is striking in how closely states in the same region are

166 grouped. Southern states are generally on the right side of Figure 2c while
167 northern states are generally to the left. Eastern states are generally at the
168 bottom of the figure while western states are generally at the top. Atlantic
169 islands are on the outer edge of the plot.

170 In Figures 2a and 2b, there is almost no distinction between the or-
171 ganization of the states. This suggests that even in situations where the
172 data does not clearly conform to the Hub Model assumption that valuable
173 information about the relationship between nodes can be identified.

174 **S4 Discussion**

175 **S4.1 Theoretical Curve**

176 We gave a simple counterexample in Section S1.1 to demonstrate that with-
 177 out any constraints, the model is not identifiable. Here we will randomly
 178 select parameters to explore the issue in general. We randomly generate
 179 an asymmetric adjacency matrix with $n = 4$ (see Table 3). The number of
 180 observed groups is set high ($T = 100,000$) to ensure good performance of
 181 the algorithm. We ran Algorithm 1 100 times, and obtained 100 different
 182 estimators with the same or similar likelihoods. Figure 3 gives a scatter-
 183 plot of the one hundred pairs of $\{\hat{A}_{1,2}, \hat{A}_{2,1}\}$ indicated by blue circles. This
 184 plot clearly demonstrates a non-linear relationship between these two values
 185 where increases in one are associated with decreases in the other.

		A_{ij}			
		j			
ρ_i	i	1	2	3	4
0.5499	1	1.0000	0.7854	0.9063	0.7957
0.3269	2	0.7032	1.0000	0.8324	0.5885
0.1016	3	0.9464	0.8817	1.0000	0.9334
0.0216	4	0.7452	0.8594	0.9478	1.0000

Table 3: True Adjacency Matrix in Identifiability Example

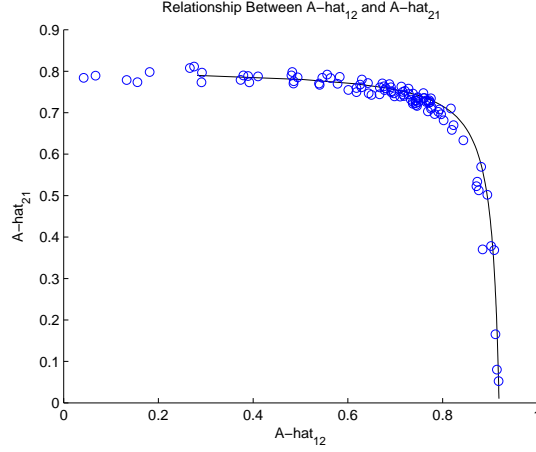


Figure 3: Nonlinear Relationship of Symmetric Elements for \hat{A}

186 It is possible to derive the theoretical relationship between symmetric
 187 elements of the adjacency matrix. As in Section S1.2, let g^x and g^y repre-
 188 sent the observed groups which contain only node v_x and v_y respectively.
 189 Further, let g^{xy} represent the group that is observed to contain only the
 190 pair v_x and v_y .

Under the Hub Models $\mathbb{P}(G = g^x)$ and $\mathbb{P}(G = g^y)$ are given by (S1.2)
 and (S1.3). In the asymmetric case, the probability of the pair is:

$$\mathbb{P}(G = g^{xy}) = \rho_x A_{xy} \prod_{j \neq \{x,y\}} (1 - A_{xj}) + \rho_y A_{yx} \prod_{j \neq \{x,y\}} (1 - A_{yj}). \quad (\text{S4.15})$$

By simply reordering the terms of Equations S1.2 and S1.3, they can

be plugged back into Equation S4.15 to give:

$$\mathbb{P}(G = g^{xy}) = \frac{A_{xy}}{(1 - A_{xy})} \mathbb{P}(G = g^x) + \frac{A_{yx}}{(1 - A_{yx})} \mathbb{P}(G = g^y) \quad (\text{S4.16})$$

By some simple algebra, we find the following relationship between the elements of the adjacency matrix:

$$A_{xy} = \frac{\mathbb{P}(G = g^{xy}) - A_{yx} [\mathbb{P}(G = g^y) + \mathbb{P}(G = g^{xy})]}{[\mathbb{P}(G = g^x) + \mathbb{P}(G = g^{xy})] - A_{yx} [\mathbb{P}(G = g^x) + \mathbb{P}(G = g^y) + \mathbb{P}(G = g^{xy})]}. \quad (\text{S4.17})$$

191 Using (S4.17), we can calculate the relationship between $A_{1,2}$ and $A_{2,1}$
 192 from the example above. This is represented in Figure 3 by the solid line.
 193 Clearly the observed solutions are falling along this theoretical curve.

194 S4.2 Self-sparsity

195 In Section 6, we introduced a property of the Hub Model estimators which
 196 we refer to as *self-sparsity*. When T is small relative to n , the model tends to
 197 produce a sparse adjacency matrix. Rabbat et al. (2008) observed similar
 198 behavior in their research. Sparsity in the adjacency matrix is achieved
 199 without any penalty in the log-likelihood, hence the name.

To begin, observe that the true probability of co-occurrence is related

to the parameters by the following equation:

$$\mathbb{P}(v_i \text{ and } v_j \text{ co-occur}) = \sum_{k=1}^n \rho_k A_{ki} A_{kj}. \quad (\text{S4.1})$$

Suppose that there is a pair of nodes, $\{v_i, v_j\}$, for which the probability of co-occurrence is exactly zero. Equation (S4.1) implies that:

$$\rho_k A_{ki} A_{kj} = 0 \quad \forall k.$$

200 Hence, for every k , at least one of the following is true: $\rho_k = 0$, $A_{ki} = 0$,
 201 or $A_{kj} = 0$. At a minimum, this requires that there be n elements of the
 202 parameters which are exactly equal to zero for every pair of nodes which
 203 fails to co-occur.

204 Clearly, $O_{ij} = 0$ implies $A_{ij} = 0$. However, self-sparsity shows that
 205 the absence of co-occurrence contains even more information than just the
 206 relationship between two nodes. Absence of co-occurrence means that no
 207 member of the population chooses to simultaneously interact with both
 208 nodes.

209 However, the question remains as to why self-sparsity occurs in the
 210 estimation of the parameters when T is small relative to n . We observe
 211 that sparsity in \hat{A} or $\hat{\rho}$ is a consequence of the EM-algorithm of HM.

212 First, it is easy to see that zero is an absorbing state for \hat{A} and $\hat{\rho}$
213 in the EM-algorithm. If, at the m^{th} iteration, $\hat{\rho}_i^{(m)} = 0$, then by (4.10)
214 $\mathbb{P}(S_i = 1|G^{(t)}) = 0$ for all i , and by (S2.13) $\hat{\rho}_i^{(m+1)} = 0$. Therefore, $\hat{\rho}_i = 0$ is
215 an absorbing state. For a similar reason, $\hat{A}_{ij} = 0$ is also an absorbing state.

216 Recall that in the EM-algorithm we set $\hat{A}_{ij} = 0$ when it is below a
217 certain threshold after estimation is complete. But why the estimate ap-
218 proaches zero is not fully understood. This aspect of the model will be
219 explored in future work.

220 S5 Text Mining Protocol

221 In Section 6, we alluded to a text mining protocol for creating grouped
222 data from the Chinese novel *Dream of the Red Chamber*. Here we present
223 an overview of this technique.

224 We first downloaded a Chinese text version of the novel from the public
225 domain and pasted the text into a Microsoft Word document. Then we
226 developed a list of character names in Chinese and assigned each character
227 a distinguishing letter from the Latin alphabet. For each character name, we
228 performed a find-and-replace search for the Chinese characters and replaced
229 them with a highlighted version of their distinguishing Latin letter. We also
230 performed a find-and-replace for paragraph breaks and represented them

231 with an identifying character. This resulted in text like that shown in
 232 Figure 4. Then we removed all the text in the Word document which was
 233 not highlighted to give us Figure 5.

□□·B·方进入房时，只见两个人搀着一位鬓发如银的老母迎上来，·B·便知是他外祖母。方欲拜见时，早被他外祖母一把搂入怀中，心肝儿肉叫着大哭起来。当下地下侍立之人，无不掩面涕泣，·B·也哭个不住。一时众人慢慢解劝住了，·B·方拜见了外祖母。——此即冷子兴所云之史氏太君，·O··P·之母也。当下·N·一指与·B·：“这是你大舅母，这是你二舅母，这是你先珠大哥的媳妇大嫂子。”·B·一一拜见过。·N·又说：“请姑娘们来。今日远客才来，可以不必上学去了。”众人答应了一声，便去了两个。……¶

Figure 4: Example of text from *Dream of the Red Chamber* after the Chinese names of characters have been replaced with Latin alphabet code letters.

¶
 BBBBOPNBBN¶
 ¶

Figure 5: Result of deleting non-highlighted text from document.

234 We pasted the remaining text into the first column of an Excel docu-
 235 ment. This gave us a text string which indicates the individuals mentioned
 236 in each paragraph. We deleted any rows with no data (these rows represent
 237 paragraphs in which none of the characters were mentioned). Then in the
 238 Excel document, we created a column for each individual where the value
 239 of a cell is 1 if the individual is mentioned in the associated text string and
 240 0 otherwise.

241 Please contact the corresponding author if you are interested in raw
242 data or estimated parameters.

243 **Bibliography**

244 Carreira-Perpinan, M. A. and Renals, S. (2000). Practical identifiability of
245 finite mixtures of multivariate bernoulli distributions. *Neural Computa-*
246 *tion*, 12:141–152.

247 Fowler, J. H. (2006a). Connecting the congress: A study of cosponsorship
248 networks. *Political Analysis*, 14(4):456–487.

249 Fowler, J. H. (2006b). Legislative cosponsorship networks in the u.s. house
250 and senate. *Social Networks*, 28(4):454–465.

251 Hamalainen, W. and Nykanen, M. (2008). Efficient discovery of statistically
252 significant association rules. *Proceedings of the 8th IEEE International*
253 *Conference on Data Mining*, pages 203–212.

254 Rabbat, M., Figueiredo, M., and Nowak, R. (2008). Network inference
255 from co-occurrences. *IEEE Transactions on Information Technology*,
256 54(9):4053–4068.

257 Shi, J. and Malik, J. (2000). Normalized cuts and image segmentation.
258 *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905.

259 Teicher, H. (1961). Identifiability of mixtures. *The Annals of Mathematical*
260 *Statistics*, 32(1):244–248.