nature human behaviour

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41562-018-0321-8

In the format provided by the authors and unedited.

Monophily in social networks introduces similarity among friends-of-friends

Kristen M. Altenburger and Johan Ugander*

Management Science and Engineering Department, Stanford University, Stanford, CA, USA. *e-mail: jugander@stanford.edu

Contents

Supplementary Note 1: Interpreting Homophily and Monophily as Model Parameters

1.1 Distribution of *in*-class Degrees

The notation is explained in the main paper, and we repeat it here for clarity. Note that we use the terminology "nodes" and "individuals" interchangeably. For notational simplicity, we will use $i \in r$ to mean the set of all nodes i with attribute class r, n_r to be the number of nodes with attribute class r, and $(N - n_r) = n_s$ to be the number of nodes with attribute class $s \neq r$ (where we focus primarily on a $k = 2$ class set-up). The *in*-class degree $d_{i,\text{in}}$ denotes the observed number of friendships node i has with individuals that also have the same attribute class r, and the *out*-class degree $d_{i,\text{out}}$ denotes the observed number of friendships node i has with those that do not have attribute class r. We use capital letters $(D_{i,\text{in}}, D_{i,\text{out}})$ when treating the *in*-/*out*-class degrees as random variables. Finally, we represent the probability of an *in*-class link forming as $p_{\text{in},r}$ for nodes in class r and represent the probability of *out*-class forming as p_{out}, where we are assuming $k = 2$ classes in which case p_{out} is necessarily equivalent for both classes.

We analyze a 2-class set-up divided into attribute classes r and s, where we give derivations for all nodes $i \in r$ and the set-up is similar for $i \in s$. For all nodes $i \in r$, node i's total observed degree d_i is partitioned between *in*-class degrees $d_{i,\text{in}}$ and *out*-class degrees $d_{i,\text{out}}$. We observe first that the conditional random variable $(D_{i,\text{in}}|D_i = d_i)$ is approximately binomially distributed, for all i in a particular class, according to the following argument: for large populations (where n_r and $N - n_r = n_s$ are large with $p_{\text{in},r} n_{\text{r}}$ and $p_{\text{out}} n_s$ constant), then $D_{i,\text{in}}$ and $D_{i,\text{out}}$, which are binomial distributed, can be view as approximately Poisson distributed. Under this Poisson approximation, the conditional distribution

$(D_{i,\text{in}} = k | D_{i,\text{in}} + D_{i,\text{in}} = d_i)$ is distributed Bin $\left(d_i, \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right)$. In full formality:

$$\begin{aligned}
\Pr(D_{i,\text{in}} = k | D_i = d_i) &= \Pr(D_{i,\text{in}} = k | D_{i,\text{in}} + D_{i,\text{out}} = d_i) && (1)\\
&= \frac{\Pr(D_{i,\text{in}} = k, D_{i,\text{out}} = d_i - k)}{\Pr(D_{i,\text{in}} + D_{i,\text{out}} = d_i)} && (2)\\
&= \frac{[e^{-n_r p_{\text{in},r}} \cdot \frac{(n_r p_{\text{in},r})^k}{k!} + o(e^{-n_r})][e^{-n_s p_{\text{out}}} \cdot \frac{(n_s p_{\text{out}})^{d_i - k}}{(d_i - k)!} + o(e^{-n_s})]}{[e^{-n_s p_{\text{out}} - n_r p_{\text{in},r}} \cdot \frac{(n_r p_{\text{in},r} + n_s p_{\text{out}})^{d_i}}{d_i!} + o(e^{-n_r - n_s})]} && (3)\\
&= \binom{d_i}{k} \left(\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right)^k \left(\frac{n_s p_{\text{out}}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right)^{d_i - k} + o(1), && (4)
\end{aligned}$$

where $o(1)$ captures an error term that is asymptotically small when n_r and n_s are both large. These steps allow us to identify the conditional distribution $(D_{i,\text{in}} | D_i = d_i)$ as approximately Bin $\left(d_i, \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right)$. When $n_r = n_s$, this distribution reduces to simply Bin $\left(d_i, \frac{p_{\text{in},r}}{p_{\text{in},r} + p_{\text{out}}}\right)$, and when $p_{\text{in},r} = p_{\text{out}}$, this distribution reduces simply to Bin $\left(d_i, \frac{n_r}{n_r + n_s}\right)$.

1.2 Homophily Index as Intercept Term

Here we show that the maximum likelihood estimate of the intercept term in the logistic regression model applied to the *in-* and *out-* degree counts among nodes in a particular class r can be interpreted as the conventional homophily index $\hat{h}_r$. This result is derived specifically for a two-class setting.

Consider $D_{i,\text{in}} | d_i, p_{\text{in},r}, p_{\text{out}} \sim$ Bin $\left(d_i, \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right)$ for nodes $i \in r$, as derived above with $p_{\text{in},r}$ and p_{out} explicitly shown as fixed for clarity and where we define the homophily parameter $h_r = \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}$. Then since the binomial distribution is a member of the exponential dispersion family and can therefore be modeled using a generalized linear model (GLM) with a logit link function, we have that $\text{logit}\left(\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right) = \log\left(\frac{\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}}{\frac{n_s p_{\text{out}}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}}\right) = \log\left(\frac{n_r p_{\text{in},r}}{n_s p_{\text{out}}}\right) = \beta_{0r}$ or equivalently $\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}} = logit^{-1}(\beta_{0r}) = \frac{e^{\beta_{0r}}}{1 + e^{\beta_{0r}}}$.

Given the observed degree counts for nodes with attribute class r represented as $\{(d_{i,\text{in}}, d_i), i \in r\}$, which are approximately independent (but weakly dependent due to combinatorial constraints on the joint distribution of degrees), we derive the maximum likelihood estimate $\hat{\beta}_{0r}$ and show its connection with the homophily index $\hat{h}_r = \sum_{i \in r} d_{i,\text{in}} / \sum_{i \in r} d_i$. First consider the likelihood function:

$$\begin{aligned}
L(\beta_{0r}) &= P(D_{1,\text{in}} = d_{1,\text{in}}, D_{2,\text{in}} = d_{2,\text{in}}, ..., D_{n_r,\text{in}} = d_{n_r,\text{in}}) && (5)\\
&= \prod_{i \in r} \binom{d_i}{d_{i,\text{in}}} \cdot \left(\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right)^{d_{i,\text{in}}} \cdot \left(1 - \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right)^{d_{i,\text{out}}} && (6)\\
&= \prod_{i \in r} \binom{d_i}{d_{i,\text{in}}} \cdot \left(\frac{e^{\beta_{0r}}}{1 + e^{\beta_{0r}}}\right)^{d_{i,\text{in}}} \cdot \left(1 - \frac{e^{\beta_{0r}}}{1 + e^{\beta_{0r}}}\right)^{d_{i,\text{out}}} && (7)\\
&\propto \left(\frac{e^{\beta_{0r}}}{1 + e^{\beta_{0r}}}\right)^{\sum_{i \in r} d_{i,\text{in}}} \cdot \left(1 - \frac{e^{\beta_{0r}}}{1 + e^{\beta_{0r}}}\right)^{\sum_{i \in r} d_{i,\text{out}}}. && (8)
\end{aligned}$$

We transform this likelihood function to a log-likelihood function:

$$\begin{aligned}
l(\beta_{0r}) &= log\left(\frac{e^{\beta_{0r}}}{1 + e^{\beta_{0r}}}\right) \cdot \sum_{i \in r} d_{i,\text{in}} + log\left(1 - \frac{e^{\beta_{0r}}}{1 + e^{\beta_{0r}}}\right) \cdot \sum_{i \in r} d_{i,\text{out}} && (9)\\
&= \beta_{0r} \cdot \sum_{i \in r} d_{i,\text{in}} - log\left(1 + e^{\beta_{0r}}\right) \cdot \sum_{i \in r} d_{i,\text{in}} + log\left(1 - \frac{e^{\beta_{0r}}}{1 + e^{\beta_{0r}}}\right) \cdot \sum_{i \in r} d_{i,\text{out}} && (10)\\
&= \beta_{0r} \cdot \sum_{i \in r} d_{i,\text{in}} - log\left(1 + e^{\beta_{0r}}\right) \cdot \sum_{i \in r} d_{i,\text{in}} - log\left(1 + e^{\beta_{0r}}\right) \cdot \sum_{i \in r} d_{i,\text{out}}, && (11)
\end{aligned}$$

and from here we set $\frac{\mathrm{d}l(\beta_{0r})}{\mathrm{d}\beta_{0r}} = 0$ and solve for $\hat{\beta}_{0r}$:

$$
\begin{aligned}
0 &= \sum_{i\in r} d_{i,\text{in}} - \frac{\sum_{i\in r} d_{i,\text{in}}}{1+e^{\hat{\beta}_{0r}}} \cdot e^{\hat{\beta}_{0r}} - \frac{\sum_{i\in r} d_{i,\text{out}}}{1+e^{\hat{\beta}_{0r}}} \cdot e^{\hat{\beta}_{0r}} && (12)\\
0 &= \sum_{i\in r} d_{i,\text{in}} + \frac{e^{\hat{\beta}_{0r}}}{1+e^{\hat{\beta}_{0r}}} \cdot \left(-\sum_{i\in r} d_{i,\text{in}} - \sum_{i\in r} d_{i,\text{out}} \right) && (13)\\
\frac{e^{\hat{\beta}_{0r}}}{1+e^{\hat{\beta}_{0r}}} &= \frac{\sum_{i\in r} d_{i,\text{in}}}{\sum_{i\in r} d_{i,\text{in}} + \sum_{i\in r} d_{i,\text{out}}} = \frac{\sum_{i\in r} d_{i,\text{in}}}{\sum_{i\in r} d_i}. && (14)
\end{aligned}
$$

Here $\hat{\beta}_{0r}$ is the maximum likelihood estimator, and we use the superscript "MLE" to make this clear. Thus when using binomial logistic regression applied to the *in*-degrees $D_{i,\text{in}}|d_i, p_{\text{in},r}, p_{\text{out}}$, we obtain that $logit^{-1}(\hat{\beta}_{0r}^{MLE}) = \frac{e^{\hat{\beta}_{0r}}}{1+e^{\hat{\beta}_{0r}}} = \frac{\sum_{i\in r} d_{i,\text{in}}}{\sum_{i\in r} d_i} = \hat{h}_r$, the conventional homophily index.

1.3 Properties of Binomial Degree Data

For a realized expected degree sequence (d_i) among nodes i in class r, the conditional distribution of *in*-class degrees is (asymptotically, per Section 2): $D_{i,\text{in}}|d_i, p_{\text{in},r}, p_{\text{out}} \sim \text{Binom}\Big(d_i, \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\Big)$ as previously derived. In this section, we assess the unconditional expectation and variance of the *in*-class degree sequence in settings where $\frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r} + n_s p_{i,\text{out}}}$ is assumed to be constant for all nodes (Model I below) and when $\frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r} + n_s p_{i,\text{out}}}$ is assumed to be random (Model II below). The derivations of Model I and Model II follow those presented in Chapter 10 of [3] and are adapted to this context in terms of *in*- and *out*- class degrees.

1.3.1 Without overdispersion (Model I)

The expectation of $D_{i,\text{in}}$ when there is no overdispersion (when $\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}$ is constant for all nodes) is:

$$
\begin{aligned}
\mathbb{E}\left[D_{i,\text{in}}|d_i\right] &= \mathbb{E}\left[\mathbb{E}\left[(D_{i,\text{in}}|d_i)|\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right]\right] && (15)\\
&= \mathbb{E}\left[\mathbb{E}\left[d_i \cdot \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right]\right] && (16)\\
&= \mathbb{E}\left[d_i \cdot \mathbb{E}\left[\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right]\right] && (17)\\
&= \mathbb{E}\left[d_i \cdot \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right] && (18)\\
&= d_i \cdot \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}. && (19)
\end{aligned}
$$

The variance (again with $\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}$ known) is:

$$
\begin{aligned}
\text{Var}\left[D_{i,\text{in}}|d_i\right] &= \mathbb{E}\left[\text{Var}\left[(D_{i,\text{in}}|d_i)\,|\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right]\right] + \\
&\quad \text{Var}\left[\mathbb{E}\left[(D_{i,\text{in}}|d_i)\,|\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right]\right] && (20)\\
&= \mathbb{E}\left[d_i \cdot \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}} \cdot \left(1 - \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right)\right] + \\
&\quad \text{Var}\left[d_i \cdot \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right]. && (21)
\end{aligned}
$$

Considering each of these two terms, we have:

$$\mathbb{E}\left[d_i \cdot \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}} \cdot \left(1 - \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right)\right] \tag{22}$$

$$= d_i \cdot \left[\mathbb{E}\left[\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right] - \mathbb{E}\left[\left(\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right)^2\right]\right] \tag{23}$$

$$= d_i \cdot \left[\mathbb{E}\left[\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right] - \text{Var}\left[\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right] - \mathbb{E}\left[\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right]^2\right] \tag{24}$$

$$= d_i \cdot \left[\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}} - 0 - \left(\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right)^2\right] \tag{25}$$

$$= d_i \cdot \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}} \cdot \left(1 - \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right), \tag{26}$$

and

$$\text{Var}\left[d_i \cdot \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right] = d_i^2 \cdot \text{Var}\left[\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right] = d_i^2 \cdot 0 = 0. \tag{27}$$

As a result, we obtain that $\text{Var}\left[D_{i,\text{in}}|d_i\right] = d_i \cdot \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}} \cdot \left(1 - \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right)$.

If the expectation and variance are rewritten in terms of h_r, then they are: $\mathbb{E}\left[D_{i,\text{in}}|d_i\right] = d_i \cdot h_r$ and $\text{Var}\left[D_{i,\text{in}}|d_i\right] = d_i \cdot h_r \cdot (1 - h_r)$, respectively.

1.3.2 With overdispersion (Model II)

Following previous notational set-ups, we introduce overdispersion by allowing $\frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r} + n_s p_{i,\text{out}}}$ to vary across nodes such that $\mathbb{E}\left[\frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r} + n_s p_{i,\text{out}}}\right] = \mathbb{E}\left[h_{i,r}\right] = \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}} = h_r$ and (for notational convenience as will be clearer later) that $\text{Var}\left[\frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r} + n_s p_{i,\text{out}}}\right] = \text{Var}\left[h_{i,r}\right] = \phi_r \cdot \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}} \cdot \left(1 - \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right) = \phi_r \cdot h_r \cdot (1 - h_r)$. Note that the only assumption we're making is that ϕ_r is constant across nodes in a given class but we are not making any distributional assumptions on $\frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r} + n_s p_{i,\text{out}}}$.

Then, the unconditional expectation of $D_{i,\text{in}}$ (unconditional on $\frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r} + n_s p_{i,\text{out}}}$) when there is overdispersion (when $\frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r} + n_s p_{i,\text{out}}}$ is random across all nodes) is:

$$\begin{aligned}
\mathbb{E}\left[D_{i,\text{in}}|d_i\right] &= \mathbb{E}\left[\mathbb{E}\left[\left(D_{i,\text{in}}|d_i\right)|\frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r} + n_s p_{i,\text{out}}}\right]\right] &(28)\\
&= \mathbb{E}\left[\mathbb{E}\left[d_i \cdot \frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r} + n_s p_{i,\text{out}}}\right]\right] &(29)\\
&= \mathbb{E}\left[d_i \cdot \mathbb{E}\left[\frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r} + n_s p_{i,\text{out}}}\right]\right] &(30)\\
&= \mathbb{E}\left[d_i \cdot \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\right] &(31)\\
&= d_i \cdot \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}} &(32)
\end{aligned}$$

And the unconditional variance (unconditional on $\frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r}+n_s p_{i,\text{out}}}$) is:

$$\begin{aligned}
\text{Var}\left[D_{i,\text{in}}|d_i\right] &= \mathbb{E}\left[\text{Var}\left[(D_{i,\text{in}}|d_i)\,|\frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r}+n_s p_{i,\text{out}}}\right]\right] + \\
&\quad \text{Var}\left[\mathbb{E}\left[(D_{i,\text{in}}|d_i)\,|\frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r}+n_s p_{i,\text{out}}}\right]\right] \qquad (33)\\
&= \mathbb{E}\left[d_i \cdot \frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r}+n_s p_{i,\text{out}}} \cdot \left(1-\frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r}+n_s p_{i,\text{out}}}\right)\right] + \\
&\quad \text{Var}\left[d_i \cdot \frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r}+n_s p_{i,\text{out}}}\right] \qquad (34)
\end{aligned}$$

Considering each part, we have:

$$\begin{aligned}
&\mathbb{E}\left[d_i \cdot \frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r}+n_s p_{i,\text{out}}} \cdot \left(1-\frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r}+n_s p_{i,\text{out}}}\right)\right] \qquad (35)\\
&= d_i \cdot \left[\mathbb{E}\left[\frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r}+n_s p_{i,\text{out}}}\right] - \mathbb{E}\left[\left(\frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r}+n_s p_{i,\text{out}}}\right)^2\right]\right] \qquad (36)\\
&= d_i \cdot \left[\mathbb{E}\left[\frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r}+n_s p_{i,\text{out}}}\right] - \text{Var}\left[\frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r}+n_s p_{i,\text{out}}}\right] - \mathbb{E}\left[\frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r}+n_s p_{i,\text{out}}}\right]^2\right] \qquad (37)\\
&= d_i \cdot \left[\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r}+n_s p_{\text{out}}} - \phi_r \cdot \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r}+n_s p_{\text{out}}} \cdot \left(1-\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r}+n_s p_{\text{out}}}\right)\right. \qquad (38)\\
&\qquad \left. - \left(\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r}+n_s p_{\text{out}}}\right)^2\right] \qquad (39)\\
&= d_i \cdot \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r}+n_s p_{\text{out}}} \cdot \left(1-\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r}+n_s p_{\text{out}}}\right) \cdot (1-\phi_r)\,, \qquad (40)
\end{aligned}$$

and

$$\begin{aligned}
\text{Var}\left[d_i \cdot \frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r}+n_s p_{i,\text{out}}}\right] &= d_i^2 \cdot \text{Var}\left[\frac{n_r p_{i,\text{in},r}}{n_r p_{i,\text{in},r}+n_s p_{i,\text{out}}}\right] \qquad (41)\\
&= d_i^2 \cdot \phi_r \cdot \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r}+n_s p_{\text{out}}} \cdot \left(1-\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r}+n_s p_{\text{out}}}\right). \qquad (42)
\end{aligned}$$

This derivation means that

$$\begin{aligned}
\text{Var}[D_{i,\text{in}}|d_i] &= d_i \cdot \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r}+n_s p_{\text{out}}} \cdot \left(1-\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r}+n_s p_{\text{out}}}\right) \cdot (1-\phi_r) + \\
&\quad d_i^2 \cdot \phi_r \cdot \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r}+n_s p_{\text{out}}} \cdot \left(1-\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r}+n_s p_{\text{out}}}\right),
\end{aligned}$$

which simplifies to

$$\begin{aligned}
\text{Var}[D_{i,\text{in}}|d_i] &= d_i \cdot \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r}+n_s p_{\text{out}}} \cdot \left(1-\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r}+n_s p_{\text{out}}}\right) \cdot (1-\phi_r+d_i \cdot \phi_r) \qquad (43)\\
&= d_i \cdot \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r}+n_s p_{\text{out}}} \cdot \left(1-\frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r}+n_s p_{\text{out}}}\right) \cdot (1+(d_i-1) \cdot \phi_r)\,. \qquad (44)
\end{aligned}$$

Therefore, when $\phi_r > 0$ then the dispersion in $D_{i,\text{in}}|d_i$ is greater than what would be expected in the setting where $\phi_r = 0$. If the expectation and variance are rewritten in terms of h_r, then they are: $\mathbb{E}\left[D_{i,\text{in}}|d_i\right] = d_i \cdot h_r$ and $\text{Var}[D_{i,\text{in}}|d_i] = d_i \cdot h_r \cdot (1-h_r) \cdot (1+(d_i-1) \cdot \phi_r)$, respectively.

1.4 Algorithm for Estimating Overdispersion Parameter ϕ_r

This section describes the iterative procedure for estimating the overdispersion (ϕ_r) among nodes i in class r by restating the procedure due to Williams [14] using our class-degree notation. The procedure initially assumes the null model without overdispersion (Model I) is true and then iteratively assesses the resulting residual variation via a goodness of fit statistic (X^2) based on the sum of squared residuals, allowing $\hat{\phi}_r > 0$ and then updating the estimates $\hat{\beta}_{0r}$ and $\hat{\phi}_r$ until convergence. The final $\hat{\phi}_r$ at the end of this process is the estimated overdispersion. Note that testing the goodness of fit statistic (X^2) relies on a predetermined significance parameter (α), which in the main paper we test the statistical significance of overdispersion in the networks we study at the $\alpha = 0.001$ significance level. For clarity on the notational differences between our work and that of Williams, we provide the following table that's explained in more detail below:

Supplementary Table 1: Conversion between our notation and Williams' notation [14].

Our notation	**Williams' notation**
d_i	m_i
$d_{i,\text{in}}$	R_i
β_{0r}	λ_i
$h_r = \frac{1}{1+\exp(-\beta_{0r})}$	$\theta = \frac{1}{1+\exp(-\lambda_i)}$
$\text{Var}(h_{i,r}) = \phi_r \cdot h_r \cdot (1-h_r)$	$\text{Var}(P_i) = \phi \cdot \theta_i \cdot (1-\theta_i)$
$v_i = d_i \cdot h_r \cdot (1-h_r)$	$v_i = m_i \cdot \theta_i \cdot (1-\theta_i)$
$w_i^{-1} = 1 + \phi_r \cdot (d_i - 1)$	$w_i^{-1} = 1 + \phi \cdot (m_i - 1)$

Given the observed degree data $\{(d_{i,\text{in}}, d_i), i \in r\}$ and assuming the underlying generative process is $D_{i,\text{in}}|d_i, p_{\text{in},r}, p_{\text{out}} \sim \text{Binom}\Big(d_i, \frac{n_r p_{\text{in},r}}{n_r p_{\text{in},r} + n_s p_{\text{out}}}\Big)$ where d_i is assumed to vary across nodes, the distinguishing feature between the initial Model I (no overdispersion) and the subsequent Model II (overdispersion) is the variance: $\text{Var}\left[D_{i,\text{in}}|d_i\right] = d_i \cdot h_r \cdot (1-h_r) \cdot (1 + (d_i - 1) \cdot \phi_r)$. For notational convenience by allowing $v_i = d_i \cdot h_r \cdot (1-h_r)$ and $w_i^{-1} = (1 + (d_i - 1) \cdot \phi_r)$, then $\text{Var}\left[D_{i,\text{in}}|d_i\right] = v_i \cdot w_i^{-1}$ where Model I strictly enforces $\phi_r = 0$ or equivalently $w_i^{-1} = w_i = 1$. Meanwhile for Model II we allow $\phi_r > 0$ or equivalently $0 < w_i^{-1} < 1$. We note that for Model I, an advantage of interpreting the homophily index within the GLM framework is that it provides a clean approach to computing statistical significance [11] using the p-value for the intercept term.

The steps of Williams' iterative algorithm for jointly estimating $\hat{\beta}_{0r}$ and $\hat{\phi}_r$ are then as follows. Viewed as an iterated algorithm, the iteration is in earnest only over the variables $\hat{\beta}_{0r}$ and $\hat{\phi}_r$ given the input data (d_i) and $(d_{i,\text{in}})$, but a number of auxiliary variables (e.g. w_i, v_i, q) greatly simplify the notation.

1. First assume there's no overdispersion ($\phi_r = 0$) and test for the significance of overdispersion being present ($\phi_r > 0$) by fitting Model I assuming there are no additional explanatory variables. Then $\hat{\beta}_{0r}^{MLE} = logit(\sum_{i \in r} d_{i,\text{in}} / \sum_{i \in r} d_i)$. Evaluate the model fit by computing a goodness-of-fit statistic (X^2) or the sum of squared residuals as $X^2 = \sum_{i \in r} \Big[(d_{i,\text{in}} - d_i \cdot \hat{h}_r)^2 / (d_i \cdot \hat{h}_r \cdot (1 - \hat{h}_r))\Big]$, which under the null should be distributed $\chi^2_{n_r - 1}$. Note as is illustrated in the main paper, we use $w_i = 1$ for this initial goodness-of-fit test, a direct consequence of $\phi_r = 0$ under the initial null model.

2. Compare X^2 with the $\chi^2_{n_r-1}$ distribution which is valid under the null $\phi_r = 0$ being true. Then assuming the null model is true, we compute the statistical significance of X^2 by computing the probability that X^2 would be as extreme if it's assumed to be distributed $\chi^2_{n_r-1}$. If X^2 is significantly large as determined by the α significance level, then reject the null that $\phi_r = 0$ and calculate

an initial estimate $\hat{\phi}_r^0$ as:

$$\hat{h}_r^0 = \frac{1}{1+\exp\left(-\hat{\beta}_{0r}^{MLE}\right)} \tag{45}$$

$$v_i^0 = d_i \cdot \hat{h}_r^0 \cdot (1-\hat{h}_r^0) \tag{46}$$

$$q^0 = \frac{1}{\sum_{i\in r} v_i^0} \tag{47}$$

$$\hat{\phi}_r^0 = \frac{X^2-(n_r-1)}{\sum_{i\in r}[(d_i-1)\cdot(1-v_i^0\cdot q^0)]}. \tag{48}$$

3. Update the weights $\hat{w}_i^t$, re-estimate $\hat{\beta}_{0r}^t$, and update $\hat{h}_r^t$ and $\hat{v}_i^t$:

$$w_i^{t+1} = \frac{1}{1+\hat{\phi}_r^t\cdot(d_i-1)} \tag{49}$$

$$\hat{\beta}_{0r}^{t+1} = \frac{\sum_{i\in r} w_i^{t+1}\cdot[v_i^t\cdot\hat{\beta}_{0r}^t + d_{i,\text{in}} - d_i\cdot\hat{h}_r^t]}{\sum_{i\in r} w_i^{t+1} v_i^t} \tag{50}$$

$$\hat{h}_r^{t+1} = \frac{1}{1+\exp\left(-\hat{\beta}_{0r}^{t+1}\right)} \tag{51}$$

$$v_i^{t+1} = d_i\cdot\hat{h}_r^{t+1}\cdot(1-\hat{h}_r^{t+1}). \tag{52}$$

4. Compute the new sum of squared residuals X^2:

$$X^{2,(t+1)} = \sum_{i\in r}\frac{w_i^{t+1}[d_{i,\text{in}} - d_i\cdot\hat{h}_r^{t+1}]^2}{v_i^{t+1}} \tag{53}$$

sand if this updated value is close to the degrees of freedom n_r-1, then $\hat{\phi}_r = \hat{\phi}_r^t$ is the dispersion estimate and the procedure stops. Otherwise, update

$$q^{t+1} = \frac{1}{\sum_{i\in r} w_i^{t+1} v_i^{t+1}} \tag{54}$$

$$\hat{\phi}_r^{t+1} = \frac{X^{2,(t+1)} - \sum_{i\in r} w_i^{t+1}\cdot(1-w_i^{t+1}\cdot v_i^{t+1}\cdot q^{t+1})}{\sum_{i\in r} w_i^{t+1}\cdot(d_i-1)\cdot(1-w_i^{t+1}\cdot v_i^{t+1}\cdot q^{t+1})} \tag{55}$$

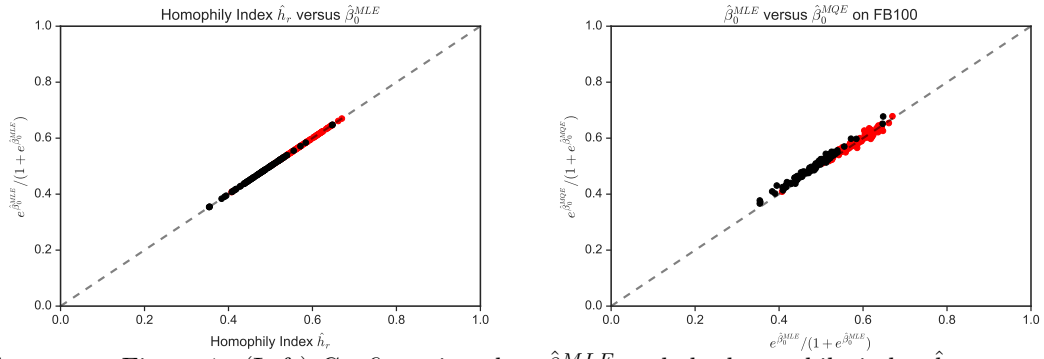
and return to step 3 with $t \leftarrow t+1$.

We use the R package `dispmod` to compute $\hat{\phi}_r$. The code snippet below assumes that the vectors `deg_same` and `deg_different` contains the degrees $d_{i,\text{in}}$ and $d_{i,\text{out}}$, respectively:

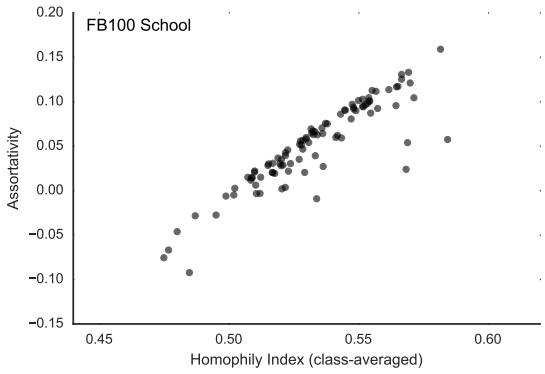
```
compute_monophily_phi <- function(deg_same, deg_different){
    mod <- glm(cbind(deg_same, deg_different) ~ 1, family=binomial(logit))
    mod.disp <- glm.binomial.disp(mod, maxit = 50, verbose = F)
    return(mod.disp$dispersion)
}
```

We note that this iterative procedure results in an estimate of homophily bias that we denote by $\hat{\beta}_{0M}^{MQE}, \hat{\beta}_{0F}^{MQE}$ as distinct from the original homophily index obtained from the GLM model ($\hat{\beta}_{0M}^{MLE}, \hat{\beta}_{0F}^{MLE}$). The estimates are comparable. Supplementary Figure 1 compares the original $\hat{\beta}_{0M}^{MLE}, \hat{\beta}_{0F}^{MLE}$ to the updated $\hat{\beta}_{0M}^{MQE}, \hat{\beta}_{0F}^{MQE}$ under the iterative Williams Method. We observe that the estimates strongly correlate, which is not surprising given that the average degree between males and females is similar and as noted in[14] that "...the difference between the maximum likelihood estimate of β under Model I and the maximum quasi-likelihood estimate of β under Model II is expected to be small when the m_i are of a similar magnitude."

One might also consider measuring homophily in terms of assortativity [7]. We observe that assortativity and the homophily index are highly correlated as shown in Supplementary Figure 2, but a major drawback of assortativity is that it does not give class-specific measures of homophily, unlike the



Supplementary Figure 1: (Left) Confirmation that $\hat{\beta}_0^{MLE}$ and the homophily index $\hat{h}_r$ are equivalent, seen here for female (red) and male (black) classes across the 97 colleges from the FB100 networks. (Right) Comparison of the maximum likelihood estimate $\hat{\beta}_0^{MLE}$ versus the Williams estimate $\hat{\beta}_0^{MQE}$ obtained from maximizing the quasi-likelihood for the same data, confirming that the difference between the estimates is small in practice, as noted by Williams [14].



Supplementary Figure 2: Across the 97 Facebook schools we compare an averaged gender homophily index and gender assortativity.

homophily index due to Coleman. This figure compares the (averaged) gender homophily index versus assortativity across the FB100 schools.

Supplementary Note 2:
Attribute Inference Methods on Social Networks

2.1 Majority Vote Classification

The aim of the 1-hop (immediate friends) and 2-hop (friends-of-friends) Majority Vote classifiers is to aggregate the known labels among an unknown node's friendship network in order to assign classification scores. Given the vector of training labels $a_1, \ldots, a_n$, where $a_i = +1$ for female training label, $a_i = -1$ for Male training label, and $a_i = NA$ is a testing label, we implement the following procedures:

- For the 1-hop Majority Vote, we use the portion of the adjacency matrix corresponding to the unknown testing labels, which we'll refer to as A_{test} and for a specific unknown node u as $A_{u,test} = A_{test}[u, :]$. Then the classification score assigned to unknown node u is based on the relative difference in the proportion of labeled male friends versus labeled female friends:

$$\frac{A_{u,test} \cdot \mathbb{I}[a = -1] - A_{u,test} \cdot \mathbb{I}[a = +1]}{A_{u,test} \cdot \mathbb{I}[a = -1] + A_{u,test} \cdot \mathbb{I}[a = +1]}. \tag{56}$$

  If node u does not have any labeled neighbors meaning that $A_{u,test} \cdot \mathbb{I}[a = -1] + A_{u,test} \cdot \mathbb{I}[a = +1] = 0$, then we assign a score based on the relative proportions in the training sample:

$$\frac{\sum \mathbb{I}[a = -1] - \sum \mathbb{I}[a = +1]}{\sum \mathbb{I}[a = -1] + \sum \mathbb{I}[a = +1]}. \tag{57}$$

- For the 2-hop Majority Vote, we implement a similar procedure as the 1-hop Majority Vote except now weights are based on A^2_{test}, weighted by the number of length-2 paths to labeled friends-of-friends.

- Note that in the case of ties or when an individual has an equal number of female and male friends, then we still assign this relative class proportion since we compare relational inference methods based on their AUC.

2.2 Regularization for LINK

The LINK model introduced by Zheleva and Getoor [15] learns a binary logistic regression classifier where the features are the entire row of the adjacency matrix among users who reveal their attribute and the outcome variable is the user's revealed attribute class. We examine these model fitting issues in this section as they pertain to binary gender inference on the datasets we study, where regularization should be given careful consideration given the large number of predictors and small number of training observations (e.g. $p \gg n_{train}$) where n_{train} denotes the size of the training sample. The method of ℓ_2-regularized logistic regression minimizes the following cost function where β represents the parameter vector corresponding to each of the N nodes in the graph with β_0 intercept, observed gender values $y_i \in \{-1, 1\}$, gain parameter C, and X_i corresponding to the ith row of the adjacency matrix for user i:

$$\min_{\beta_0, \beta} \frac{1}{2}\beta^T\beta + C \cdot \sum_{i \in train} \log\big(\exp\big(-y_i(X_i^T\beta + \beta_0)\big) + 1\big). \tag{58}$$

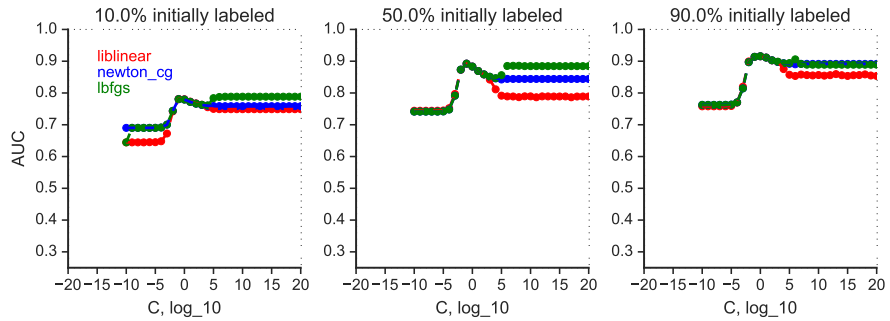
Here C captures the inverse of the regularization strength, where for concreteness small (large) values of C correspond with large (small) amounts of regularization. After exploring the sensitivity of the C parameter on classification performance, we find minimal improvement from incorporating ℓ_2-regularization. To minimize this cost function we use the implementation in Python's `scikit-learn` library.

We evaluated several different optimization methods for minimizing the regularized loss across a wide range of gains C; in these evaluations we focus on the Amherst College network from the FB100 dataset, the same network featured in the individual network analyses in the main paper. We evaluate `scikit-learn`'s (package version 0.18.1) `lbfgs`, `newton-cg`, and `liblinear` solvers, all with their default parameter settings, as illustrated in Supplementary Figures 3 and 4 across varying regularization gains. Note that only the `liblinear` solver is evaluated with ℓ_2-regularization and ℓ_1-regularization in Python, while `lbfgs` and `newton-cg` are only evaluated with ℓ_2-regularization. These different solvers sometimes choose rather different models, as seen in the sometimes large differences in AUC. But ultimately across all three solvers we observe robust evidence that ℓ_2-regularization does not help the AUC,

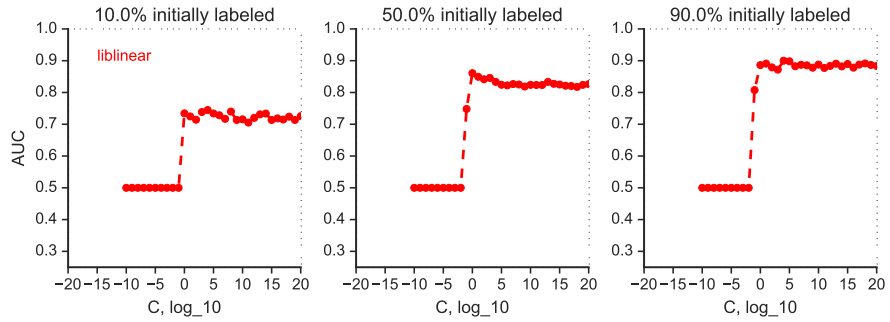
and therefore choose to learn the LINK models throughout this paper (outside this section) using a very large regularization gain C with the `lbfgs` solver, effectively disabling regularization.

On the subject of ℓ_1-regularization within LINK, we briefly note that such a regularization would in a sense be trying to find a small subset of individuals to use as features for the entire graph, an insight motived by the "subset selection" interpretation of ℓ_1-regularization [13]. Since each of the n nodes is only connected to a small fraction of the graph, there's a formal sense to which we require some $O(n)$ fraction of the nodes to have non-zero weights, not $o(n)$, contradicting the subset selection motivation. As a result, the lack of improvement from ℓ_1-regularization is expected, and indeed this is what we see in Supplementary Figure 4.

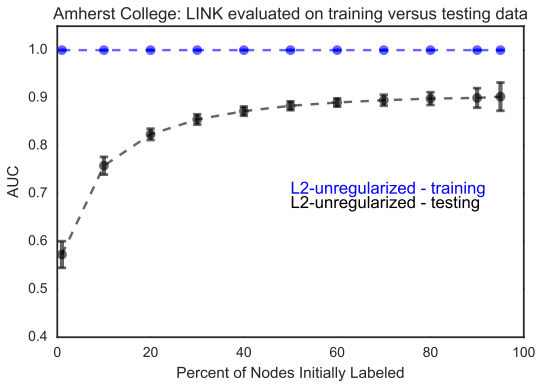
We observe, as has been previously noted [5], that this unregularized model with a very large number parameters is still empirically good at distinguishing between classes. We also observe a tendency toward separability in Amherst, as seen in Supplementary Figure 5. Then, as noted in [10], behavior with a large gain C is similar to choosing an ℓ_2 max-margin classifier. For prediction on the political blog network and the Noordin Top membership network we employ C=1.



Supplementary Figure 3: Evaluate sensitivity to regularization parameter, C, on Amherst College, for 1 fold, for ℓ_2-regularization.



Supplementary Figure 4: Evaluate sensitivity to regularization parameter, C, on Amherst College, for 1 fold, for ℓ_1-regularization.



Supplementary Figure 5: Evaluating separability on Amherst College.

2.3 Interpretation of LINK as a 2-hop Method

We claim that the LINK family of linear classifiers (based on linear weights applied to the columns of the adjacency matrix as feature vectors), which include LINK-Logistic Regression, LINK-SVM, and LINK-Naive Bayes, obtain their predictive power from 2-hop paths to individuals friends-of-friends. This section establishes this 2-hop connection explicitly by deriving the weights used by LINK in a Naive Bayes classifier. We demonstrate the Naive Bayes classifier reduces to a linear aggregation over a nodes friends of a nonlinear aggregation over those nodes' friends (the friends of friends of the classification subject). Note that we employ a Laplace smoothing factor ("+1") to handle the case when a node does not have any male or female friends in the training sample.

Suppose that for our training data we observe labels $y_{train} \in \{M, F\}$ with the corresponding observed features $x_i \in \{0, 1\}$ and random variable $X \in \{0, 1\}^N$ where N is the total number of nodes in the graph. This set-up represents the observed gender labels, y_{train}, and the observed friendships that all nodes in the network have with these training data. From this information, we construct the Naive Bayes classification rule by making the standard conditional independence assumption and studying the likelihood ratio:

$$
\begin{aligned}
LR(x) &= \frac{P(y_{train} = F) \cdot P(X | y_{train} = F)}{P(y_{train} = M) \cdot P(X | y_{train} = M)} && (59) \\
&= \frac{P(y_{train} = F) \cdot \prod_{i=1}^{N} P(X_i = x_i | y_{train} = F)}{P(y_{train} = M) \cdot \prod_{i=1}^{N} P(X_i = x_i | y_{train} = M)} && (60) \\
&= \frac{P(y_{train} = F)}{P(y_{train} = M)} \\
&\quad \cdot \prod_{i:x_i=0} \frac{P(X_i = x_i | y_{train} = F)}{P(X_i = x_i | y_{train} = M)} \cdot \prod_{i:x_i=1} \frac{P(X_i = x_i | y_{train} = F)}{P(X_i = x_i | y_{train} = M)}. && (61)
\end{aligned}
$$

Note that in this expression we've separated the non-neighbors and neighbors of a test node by the restriction ($i : x_i = 0$ and $i : x_i = 1$), to be considered separately.

We now have the following empirical estimates, where $n_{F,train}$ ($n_{M,train}$) denotes the number of females (males) in the training sample, $d_{i,F,train}$ ($d_{i,M,train}$) denotes node i's degree with F (M) nodes in the training sample, and we finally assume $d_{i,F,train} + d_{i,M,train} = d_{i,train}$. For simplicity in notation, we'll remove the *train* subscript for the rest of this section. Including +1 Laplace smoothing we have

the following standard maximum likelihood estimates for the "parameters" of the Naive Bayes model:

$$\hat{P}(y_{train} = F) = \frac{n_F}{n_F + n_M} \tag{62}$$

$$\hat{P}(y_{train} = M) = \frac{n_M}{n_F + n_M} \tag{63}$$

$$\hat{P}(X_i = 1 | y_{train} = F) = \frac{d_{i,F} + 1}{n_F + 2} \tag{64}$$

$$\hat{P}(X_i = 1 | y_{train} = M) = \frac{d_{i,M} + 1}{n_M + 2} \tag{65}$$

$$\hat{P}(X_i = 0 | y_{train} = F) = \frac{n_F - d_{i,F} + 1}{n_F + 2} \tag{66}$$

$$\hat{P}(X_i = 0 | y_{train} = M) = \frac{n_M - d_{i,M} + 1}{n_M + 2}. \tag{67}$$

Substituting these empirical estimates into the earlier likelihood ratio, we obtain the following likelihood ratio for classifying a test node as belonging to class F:

$$LR(x) = \frac{n_F}{n_M} \prod_{i:x_i=0} \frac{\left(\frac{n_F - d_{i,F} + 1}{n_F + 2}\right)}{\left(\frac{n_M - d_{i,M} + 1}{n_M + 2}\right)} \prod_{i:x_i=1} \frac{\left(\frac{d_{i,F} + 1}{n_F + 2}\right)}{\left(\frac{d_{i,M} + 1}{n_M + 2}\right)} \tag{68}$$

$$= \frac{n_F}{n_M} \prod_{i:x_i=0} \left(\frac{n_F - d_{i,F} + 1}{n_M - d_{i,M} + 1}\right) \left(\frac{n_M + 2}{n_F + 2}\right) \prod_{i:x_i=1} \left(\frac{d_{i,F} + 1}{d_{i,M} + 1}\right) \left(\frac{n_M + 2}{n_F + 2}\right) \tag{69}$$

$$= \frac{n_F}{n_M} \left(\frac{n_M + 2}{n_F + 2}\right)^N \prod_{i:x_i=0} \left(\frac{n_F - d_{i,F} + 1}{n_M - d_{i,M} + 1}\right) \prod_{i:x_i=1} \left(\frac{d_{i,F} + 1}{d_{i,M} + 1}\right) \tag{70}$$

$$= \frac{n_F}{n_M} \left(\frac{n_M + 2}{n_F + 2}\right)^N \prod_{i=1}^{N} \left(\frac{d_{i,F} + 1}{d_{i,M} + 1}\right)^{x_i} \left(\frac{n_F - d_{i,F} + 1}{n_M - d_{i,M} + 1}\right)^{1 - x_i}. \tag{71}$$

Then considering the log of the likelihood-ratio, where
$C = \log(n_F/n_M) + N \log((n_M + 2)/(n_F + 2))$ is a constant:

$$\log(LR(x)) = C + \log\left(\prod_{i=1}^{N} \left(\frac{d_{i,F} + 1}{d_{i,M} + 1}\right)^{x_i} \cdot \left(\frac{n_F - d_{i,F} + 1}{n_M - d_{i,M} + 1}\right)^{1 - x_i}\right) \tag{72}$$

$$= C + \sum_{i=1}^{N} x_i \cdot \log\left(\frac{d_{i,F} + 1}{d_{i,M} + 1}\right) + \sum_{i=1}^{N} (1 - x_i) \cdot \log\left(\frac{n_F - d_{i,F} + 1}{n_M - d_{i,M} + 1}\right) \tag{73}$$

$$= C + \sum_{i=1}^{N} \log\left(\frac{n_F - d_{i,F} + 1}{n_M - d_{i,M} + 1}\right) + \sum_{i=1}^{N} x_i \cdot \left(\log\left[\frac{d_{i,F} + 1}{d_{i,M} + 1}\right] - \log\left[\frac{n_F - d_{i,F} + 1}{n_M - d_{i,M} + 1}\right]\right) \tag{74}$$

$$= C + \sum_{i=1}^{N} \log\left[\frac{n_F - d_{i,F} + 1}{n_M - d_{i,M} + 1}\right] + \sum_{i=1}^{N} x_i \cdot \log\left[\left(\frac{d_{i,F} + 1}{d_{i,M} + 1}\right) \cdot \left(\frac{n_M - d_{i,M} + 1}{n_F - d_{i,F} + 1}\right)\right]. \tag{75}$$

If we assume that the network is sparse we have that $n_F, n_M >> d_{i,F}, d_{i,M}$ and can therefore simplify:

$$\log(LR(x)) \approx \underbrace{C + \sum_{i=1}^{N} \log\left[\frac{n_F}{n_M}\right]}_{C'} + \sum_{i=1}^{N} x_i \cdot \log\left[\left(\frac{d_{i,F}+1}{d_{i,M}+1}\right) \cdot \left(\frac{n_M}{n_F}\right)\right] \quad (76)$$

$$= C' + \sum_{i=1}^{N} x_i \cdot \log\left[\left(\frac{d_{i,F}+1}{d_{i,M}+1}\right) \cdot \left(\frac{n_M}{n_F}\right)\right] \quad (77)$$

Thus, we see that LINK-Naive Bayes in particular (and the LINK method in general) is a 2-hop method since the classification procedure relies on the degrees $d_{i,F}$ and $d_{i,M}$ of an unlabeled test node's neighbor, effectively incorporating information about the labels of nodes two hops away. In the case when $n_M = n_F$, then we can directly observe this 2-hop relation as the log-likelihood ratio $\log(LR(x))$ reduces to $C' + \sum_{i=1}^{N} x_i \cdot \log\left[\left(\frac{d_{i,F}+1}{d_{i,M}+1}\right)\right]$. Ignoring the +1 (which comes from the use of Laplacian smoothing), the likelihood a test node is F is scored based on the relative tendency of the test node's neighbors to form friendships with F nodes relative to M nodes. Note that the scoring is *not* based on the neighbor's labels but rather on the neighbor-of-neighbor's labels in the training data.

As a final comment, the use of the adjacency matrix to classify nodes pre-dates the LINK paper, to at least McSherry [6], where the dominant eigenvector of the adjacency matrix was used as an unsupervised method to classify nodes in a stochastic block model (SBM). One can think of the McSherry algorithm as a dimensionality reduction step (to dimension 1) applied to LINK, with theoretical guarantees for recovering the block structure of SBMs. The McSherry work further emphasizes the focus of adjacency-matrix-based methods on homophily, given that the target problem there is related to stochastic block models (SBMs), which model homophily and not monophily.

2.4 Evaluation Metrics for Assessing Predictive Performance

The aim of this section is to provide further details on how to measure classification performance, illustrate and interpret alternative metrics to assess classification performance, and to describe the inherent trade-offs when selecting a performance metric and how the choice of performance metric is specific to the prediction problem. For example, predicting gender versus predicting suspicious persons of interest represent settings where there are likely different misclassification considerations and costs. We report all classification performance results in the main paper using a weighted Area Under the Curve (AUC) [1]. We first summarize typical metrics for evaluating classifier performance such as accuracy, AUC, precision, and recall and then provide these alternative evaluations.

The output of a classifier is the likelihood an observation is predicted to be in a target class of interest. Choosing an evaluation metric for a classifier then depends on whether one is interested in evaluating the classification scores themselves or whether a cutoff threshold is already pre-selected for labeling based on predicted scores. For any given threshold cutoff, a confusion matrix is a useful tool for understanding the various ways classifier performance can be measured either for a fixed classifier threshold or across various classifier thresholds. As shown in Supplementary Table 2.4, for binary classification there are two ground-truth classes, r and s, and the output from a classifier at a given cutoff threshold is a prediction that an observation is in one of the classes. A perfect classifier would mean all the actual classes are correctly predicted, and that there are no False Negatives (FN) or False Positives (FP).

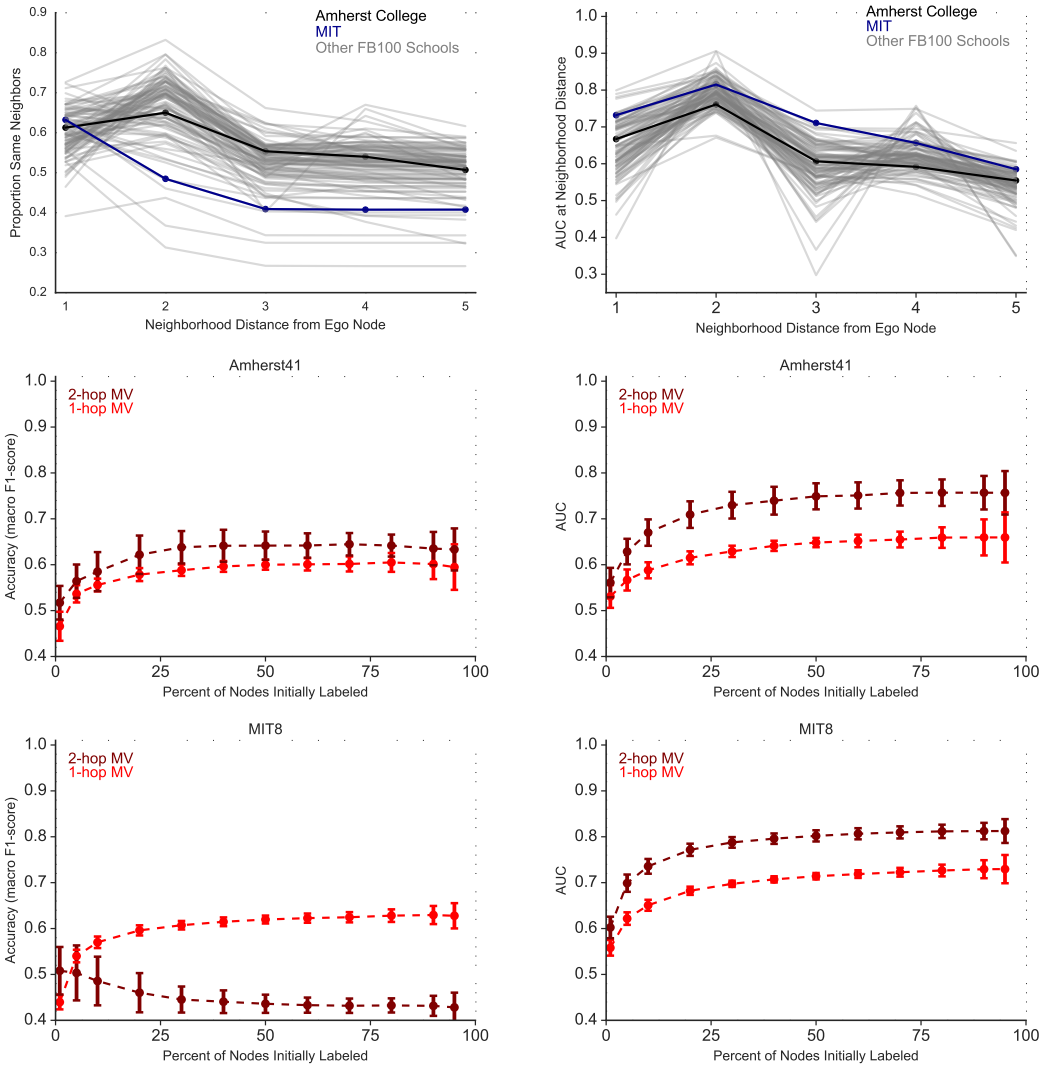
	True Class r	**True Class** s
Predicted Class r	True Positive (TP)	False Positive (FP)
Predicted Class s	False Negative (FN)	True Negative (TN)

Supplementary Table 2: Confusion matrix that illustrates the various ways to categorize predicted labels from a binary classifier at one particular threshold cutoff for an observation to be predicted to be in a specific class. Here, we're assuming the model is to predict an observation belonging to class r.

Using this confusion matrix, we next define the following potential evaluation metrics that can be used for assessing and comparing classifier performance and describe limitations of each. For concreteness, we assume a binary classification setting where possible labels $y_i \in s, r$ for individual i and our classifier predicts $P(y_i = r)$.

- **Accuracy:** Accuracy is a commonly chosen metric if one is interested in evaluating a classifier performance at a pre-determined threshold. At a classifier threshold C, the prediction classification scores defined by $P(y_i = r)$ would be converted to labels as follows: $P(y_i = r) \geq C$ indicates $\hat{y} = r$ and otherwise $\hat{y} = s$. Then, the accuracy metric assesses the overall proportion of correctly labeled observations computed by $\left(\frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}\right)$. We comment on two main limitations of using accuracy for our purposes. First, in settings with imbalanced classes, this metric can be very misleading since a naive classifier that assigns all observations to the majority class can falsely result in a classifier with high predictive accuracy [2]. Second, this metric relies on sufficient domain knowledge of the prediction setting to make an informed choice of a decision threshold cutoff. Given these limitations of interpreting accuracy when classes are imbalanced and differently balanced in different datasets (i.e. in the datasets we explore) and of choosing an informed classifier threshold, we instead primarily focus on AUC in the main text, which provides a more comprehensive overview of classifier performance.

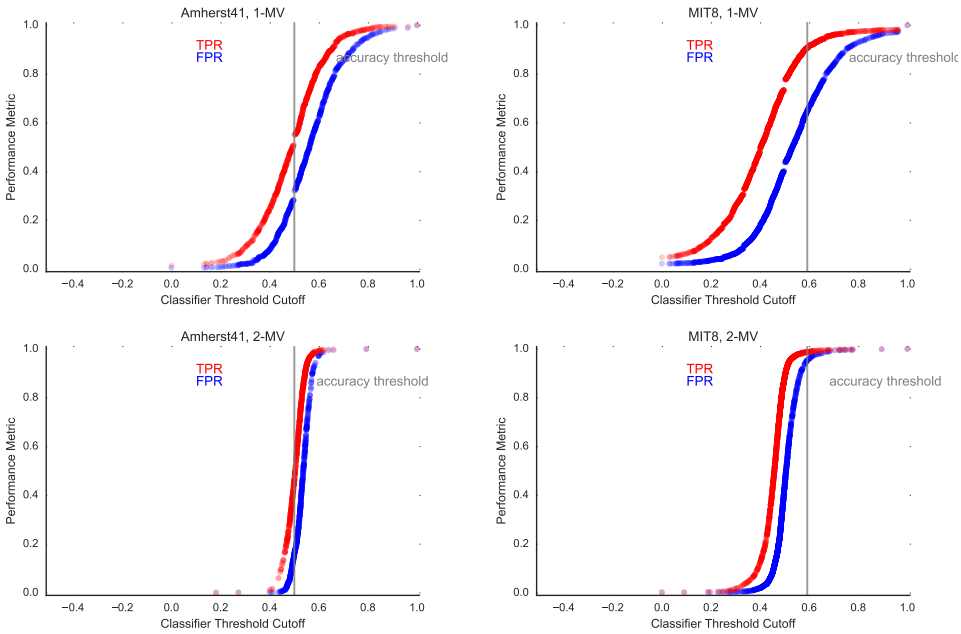
  In order to illustrate accuracy results on Amherst College and other networks, we use the relative class proportion as the classifier cut-off and show accuracy performance in Supplementary Figure 6. In the left panels we illustrate accuracy-based evaluations, where accuracy is equivalent to the proportion of same-class neighbors. More concretely, we compute for each node the proportion of nodes that are in the same-class, and then across all nodes in the network, count the proportion of nodes where the majority of their neighbors are in the same class, using weights based on the number of length-k paths to labeled k-hop friends. We define majority relative to the class proportions in the whole network. We then highlight two schools, Amherst College and MIT, which exhibit differing behavior in top top panel. Amherst College has a larger percentage of nodes with same-class neighbors at 2-hop whereas MIT has a smaller percentage at 2-hop compared to 1-hop. We emphasize that the analog set of figures for a percent same at k-hop figure is evaluating classifier performance based on accuracy, shown in the middle-left and bottom-left of Supplementary Figure 6 where we confirm 2-hop MV is higher than 1-hop MV for Amherst, and the reverse for MIT.



Supplementary Figure 6: (Left Column) Accuracy interpretation of classifier performance on FB100 schools for a fixed decision threshold. We highlight results for Amherst College and MIT where the 2-hop and 1-hop performance changes differ. (Right Column) AUC interpretation of classifier performance on FB100 schools, across varying decision thresholds.

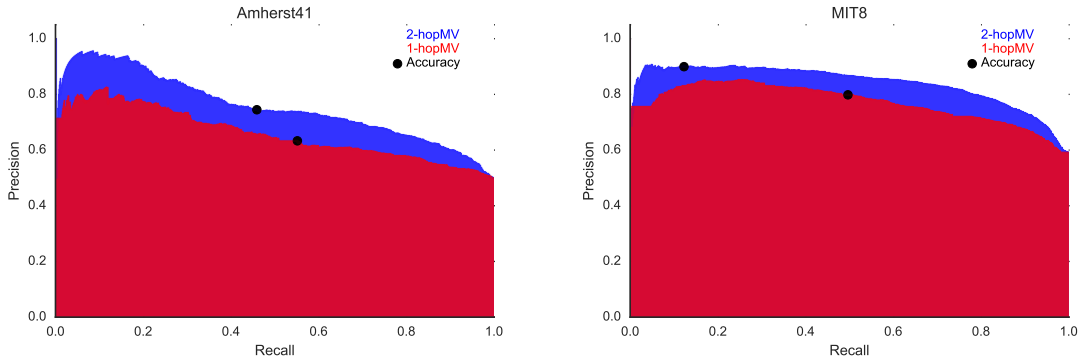
- **Area Under the Curve (AUC):** AUC is a commonly chosen metric to evaluate overall classifier performance across a range of classifier thresholds. It is based on the ROC curve which varies classifier thresholds to obtain a global sense of classifier performance across a range of cutoffs[12]. In the right panel of Supplementary Figure 6, we illustrate across the FB100 schools AUC performance for k-hop majority vote classification (k=1,2,3,...). We consistently observe higher performance at 2-hop relative to 1-hop, which we further illustrate in the middle-right and bottom-right for Amherst and MIT, respectively.

  In Supplementary Figure 7 we provide a fuller characterization of the AUC for Amherst and MIT from the FB100 dataset, evaluating 1-hop and 2-hop majority vote on these networks. The x-axis represents the various threshold cutoffs while the y-axis plots the False-Positive Rate (FPR) and the True-Positive Rate (TPR), which are what the area under the ROC curve concisely summarizes into one statistic. We also illustrate how an accuracy metric provides a limited view of classifier performance by using a simple decision threshold cutoff based on relative gender proportions in the school population. We observe for MIT using 2-hop MV that this threshold is not optimal as TPR and FPR are comparable and a slightly lower decision threshold would result in a relatively unchanged TPR but a lower FPR.



Supplementary Figure 7: Visualizations of what the AUC captures from the ROC curve and corresponding accuracy threshold cutoff for Amherst College (left) and MIT (right) for 1-hop MV (top row) and 2-hop MV (bottom row). The gray vertical line reports the TPR and FPR values at the specific threshold cut-off used if we reported accuracy based on relative class proportions.

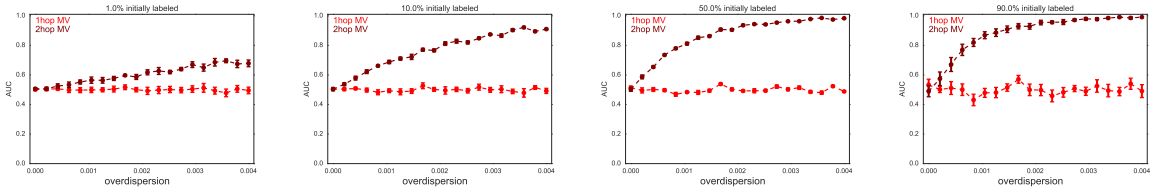
- **Precision and Recall:** These metrics are aimed at overcoming the limitations of accuracy by providing an assessment of the proportion of predicted observations that are truly of the predicted class (precision) and the proportion of the class of interest that are actually predicted as being in the predicted class of interest (recall); the formulas for precision and recall are $\left(\frac{\text{TP}}{\text{TP+FP}}\right)$ and $\left(\frac{\text{TP}}{\text{TP+FN}}\right)$, respectively. A precision-recall curve generalizes these metrics by varying the classification thresholds and comparing precision vs. recall. We illustrate the precision-recall curves for Amherst College and MIT in Supplementary Figure 8. Observe that the curves for the 2MV methods almost completely dominate the curves for the 1MV methods on both schools.



Supplementary Figure 8: Visualization of Precision-Recall Curves for Amherst College (left) and MIT (right). The black point indicates where an accuracy metric would be if labels were assigned based on relative class proportions.

2.5 Sensitivity of Predictive Performance to Overdispersion

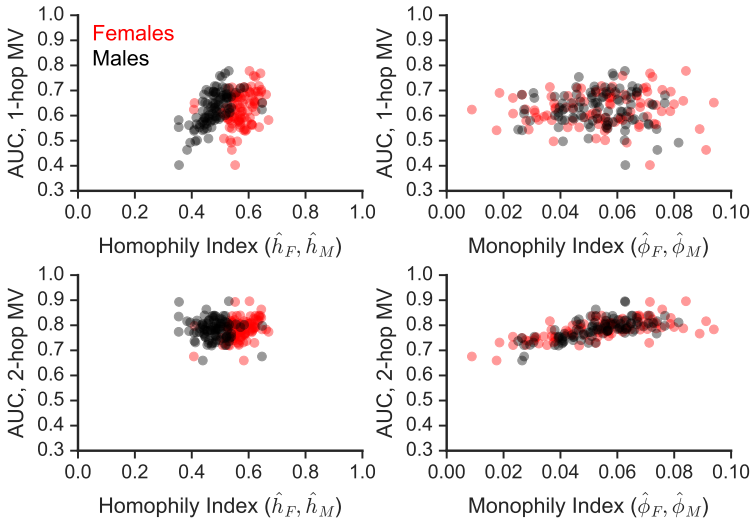
As a way to examine how overdispersion is driving the improved predictive performance of 2-hop MV relative to 1-hop MV in the non-homophilous network setting is to compare the relative performance of these classifiers on the oSBM as we increase the overdispersion. As is illustrated in Figure 3A in the main text, we generate an oSBM network (described in detail in Section 4) that for fixed $x\%$ of nodes initially labeled (1%, 10%, 50%, 90%), we create different instances of the oSBM without homophily but increasing the amount of monophily (overdispersion). The goal is then to compare the relative predictive performance of 2-hop versus 1-hop MV. As overdispersion increases we see how 2-hop MV improves in predictive performance relative to 1-hop MV, verifying that the increasing overdispersion in the oSBM can account for the improved 2-hop MV performance in the non-homophilous setting.



Supplementary Figure 9: Sensitivity of 1-hop vs. 2-hop MV to increasing overdispersion, where the x-axes denote $\phi^*_{in} = \phi^*_{out}$ in the oSBM. Different panels show different percentages of initially labeled nodes. We observe that as the overdispersion increases (while all other parameter values remain unchanged) the 2-hop MV performance increases relative to 1-hop MV.

Moving to empirical networks, classification for 2-hop Majority Vote considerably outperforms classification based on 1-hop Majority Vote across the population of FB100 schools, as illustrated in Supplementary Figure 10. We attribute this performance difference to the monophily in the network.

As a way to evaluate the predictive performance specifically of LINK in terms of homophily versus monophily, we consider the $\hat{\beta}_j$ coefficients from the fitted LINK model. The idea is that the coefficients

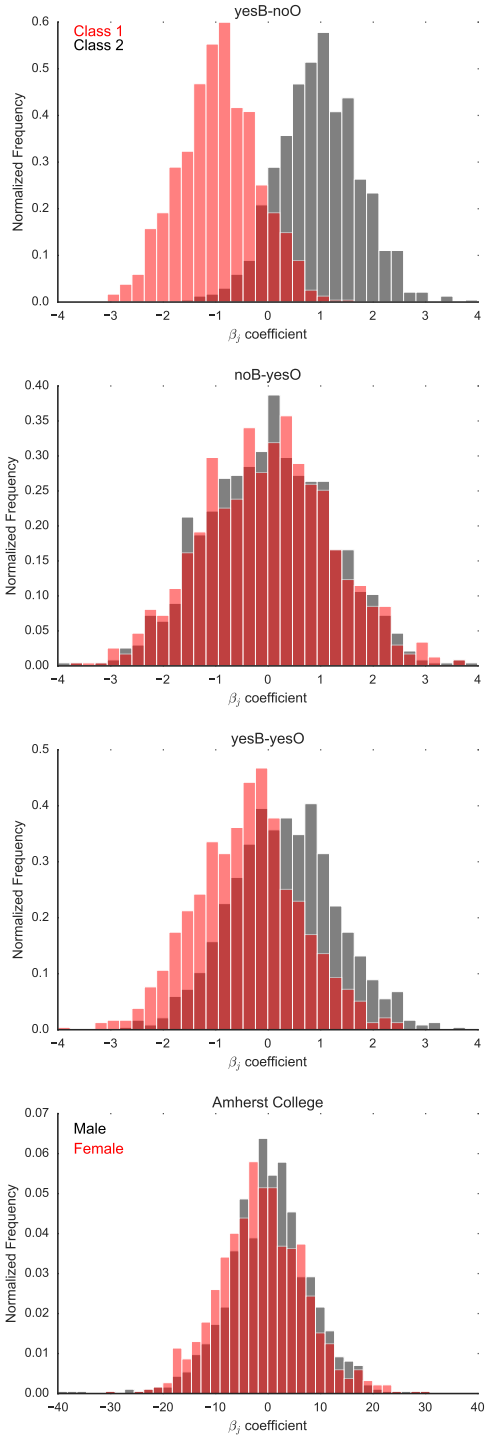


Supplementary Figure 10: Across FB100 networks we compare the correlation between 1-hop and 2-hop Majority Vote (with 50% initially labeled nodes) versus gender homophily and gender monophily. We observe that homophily has high explanatory power for the 1-hop Majority Vote AUC across schools while monophily has very little. Meanwhile, homophily has weak explanatory power of the 2-hop Majority Vote AUC across schools while monophily has strong explanatory power for that method.

can help inform us *who* is useful for prediction by comparing the sign of the estimated coefficient $\hat{\beta}_j$ (as non-negative or negative) to that to that node's own class label.

To review, the interpretation of a β_j coefficient from a LINK-Logistic regression model that for example predicts the likelihood of being male ($y_i = 1$) versus female ($y_i = 0$) is that the coefficient β_j indicates being friends with individual j, keeping everything else fixed, impacts one's own log-odds of being Male by β_j. If $\beta_j > 0$, then this means that one has an increased chance of being male while the reverse is true if $\beta_j < 0$. Therefore, if $\beta > 0$ *among* nodes that are primarily male, then this means that males are revealing male gender of their friends or homophily is driving the predictions. In a non-homophilous but monophilous setting, we'd expect a mixture of female and male nodes to be predictive of being male.

We illustrate the learned coefficients across 4 networks (three oSBM networks and Amherst College) in Supplementary Figure 11. We note that in the homophily-only setting (yesB-noO), we observe a clear separation that Class 2 (Class 1) nodes on average are predictive of Class 2 (Class 1). In the monophily-only setting (noB-yesO), we observe a mixture of Class 1 and Class 2 nodes both being predictive indicated by complete overlap in the distributions. For the homophily and monophily setting, we observe some overlap in the distributions but with some Class 2 nodes being indicative of Class 2. Finally, another way to observe that overdispersion is resulting in high predictive performance is that there's almost complete overlap in the coefficient distribution for Amherst College.

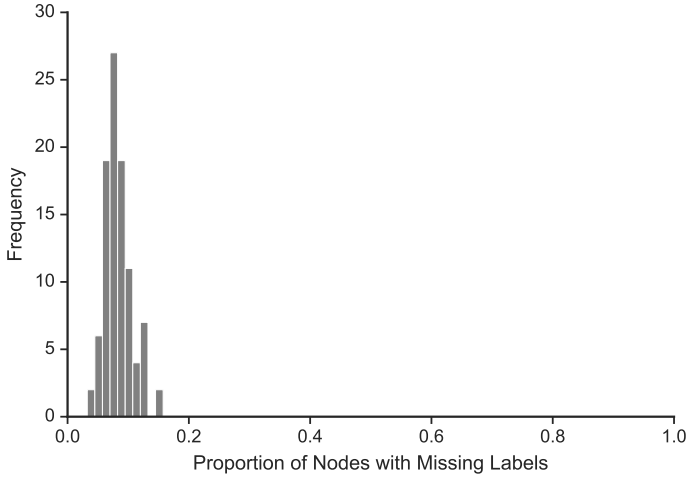


Supplementary Figure 11: An analysis of *which* nodes are useful for prediction based on the LINK model, comparing the coefficient value for nodes within each class label. (yesB-noO) On an oSBM with only block structure and no overdispersion, we'd expect nodes of a specific class to be predictive of their own class. That is, given that a node is similar to their neighbors in a homophilous network, the useful nodes for prediction should reveal their own class, which is illustrated in the two set of figures that males tend to reveal males and females tend to reveal females. (noB-yesO) Next, we compare an oSBM without block structure but with overdispersion where we see complete overlap in terms of which nodes are predictive of class membership. (yesB-yesO) We observe both some separation in the histograms along with significant overlap in which classes are predictive. (Amherst College) We observe that this network is monophily-driven for gender prediction as both male and female nodes are predictive of gender.

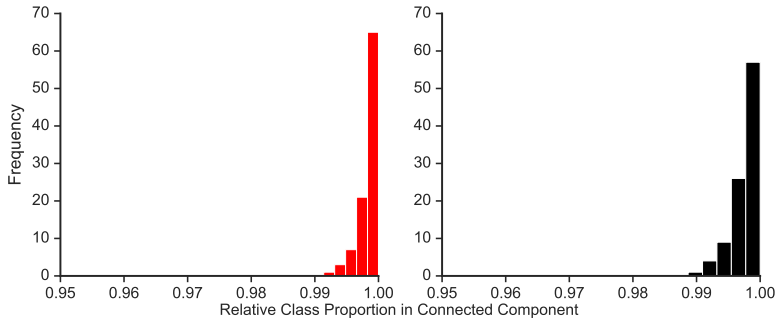
Supplementary Note 3:
Analysis of Homophily and Monophily on Social Networks

3.1 Facebook Data Pre-processing

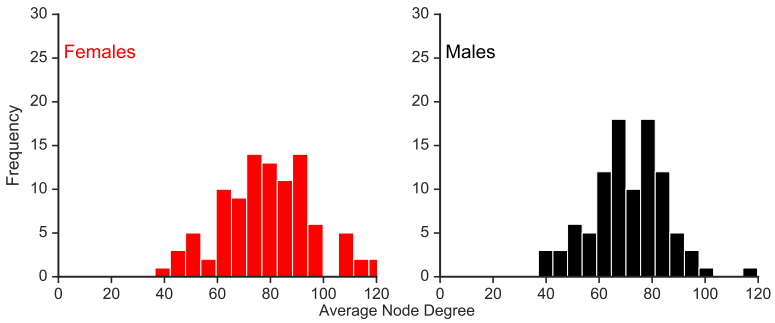
Supplementary Figure 12 shows that across the Facebook schools there's a very small percentage of individuals with truly unknown gender labels, always less than 16% with an average of 8.3%. We acknowledge that these unknown individuals can be useful for revealing the gender of others using e.g. LINK or 2-hop Majority Vote, but given that we are not able to include them in a training/testing cross-validation set-up, we completely remove them in this work since our goal is to compare the relative performance of inference methods. Supplementary Figure 13 illustrates the relative proportion of nodes in the largest connected component, and shows that these nodes comprise the majority of individuals in each graph, so subsetting is a minimal change compared to the original dataset. Supplementary Figure 14 illustrates that across the population of schools, males and females have comparable average degrees. The mean average degree is 71.35 for males and 79.22 for females.



Supplementary Figure 12: The proportion of nodes with missing gender labels in the original network dataset across the 97 schools from the FB100 schools.

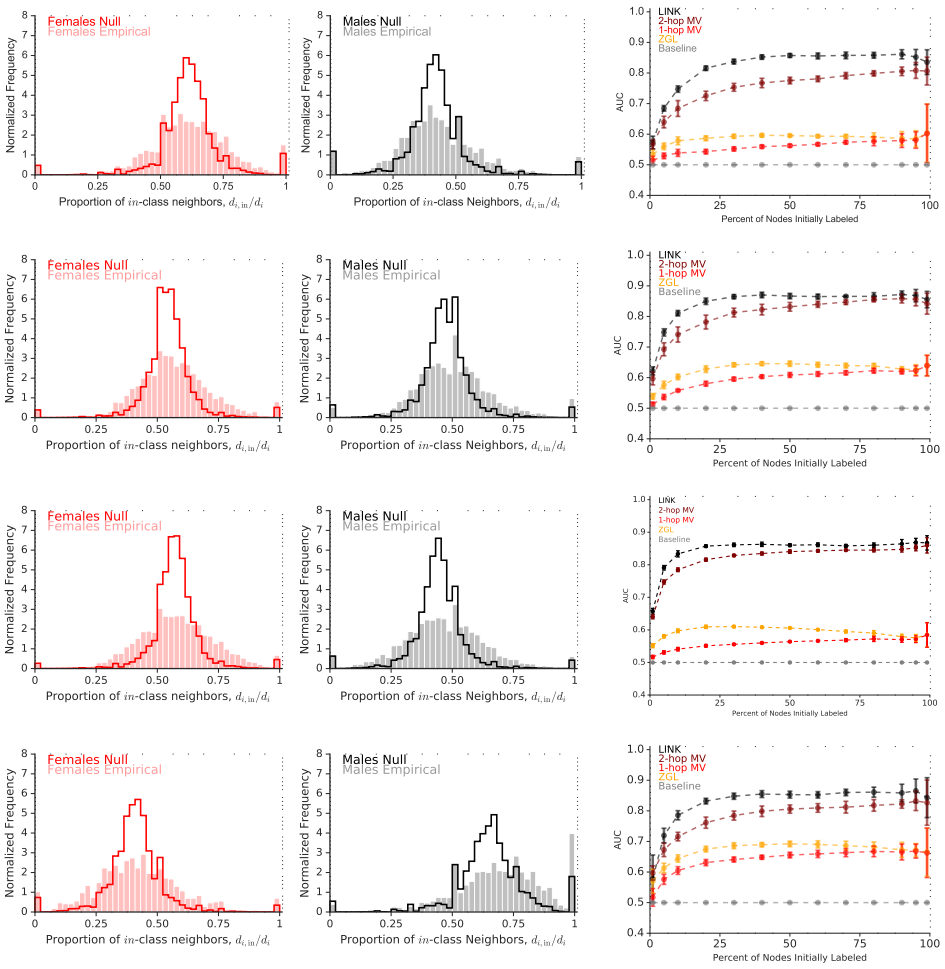


Supplementary Figure 13: The relative proportion of the original nodes preserved after subsetting to the largest connected component across the 97 schools from the FB100 schools.



Supplementary Figure 14: The average node degree among nodes in the largest connected component across the 97 schools from the FB100 schools.

3.2 Additional FB100 Figures (gender)



Supplementary Figure 15: (A) American University, (B) Boston College, (C) Indiana University, and (D) University of Michigan.

3.3 Add Health Analysis (gender)

This section provides an analysis of gender inference on the Add Health dataset [9]. We evaluate homophily and monophily on the undirected and directed degree sequences, where the measures generalize cleanly to the directed setting. We follow the same data pre-processing steps as we did for the FB100, restricting the analysis to only nodes that disclose their gender and restricting to nodes in the largest (weakly) connected component.

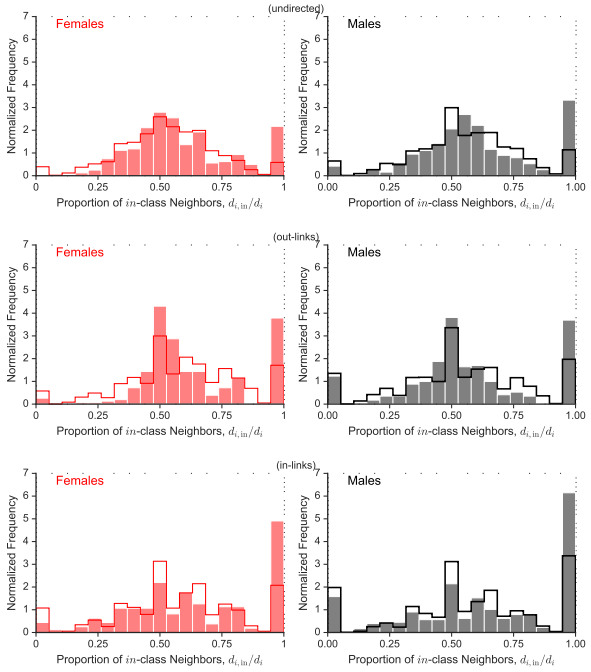
A particularly remarkable property of the Add Health networks is that they result from a directed friendship nomination survey that limited the number of male and female friends that each person could nominate, up to five of each. While a survey under such constraints strongly limits the presence of homophily, it does not limit monophily in the in-directed network. For instance, if all females nominate person u, then "having u as a friend" would be a key feature for inferring the gender of individuals who have kept that information private. We observe that it is therefore still possible to achieve high classification accuracy in the directed Add Health networks collected with constrained surveys using methods that can harness monophily, a clear demonstration that constrained surveys cannot guarantee privacy. We leave as an open question how to limit the predictive performance of directed network surveys for inferring private attributes.

The out-directed graph is described by an adjacency matrix $A_{ij} = 1$ if student i *nominated* student j. The data collection was restricted such that $\sum_j A_{ij} \leq 10$, $\sum_{j \in M} A_{ij} \leq 5$, and $\sum_{j \in F} A_{ij} \leq 5$. The in-

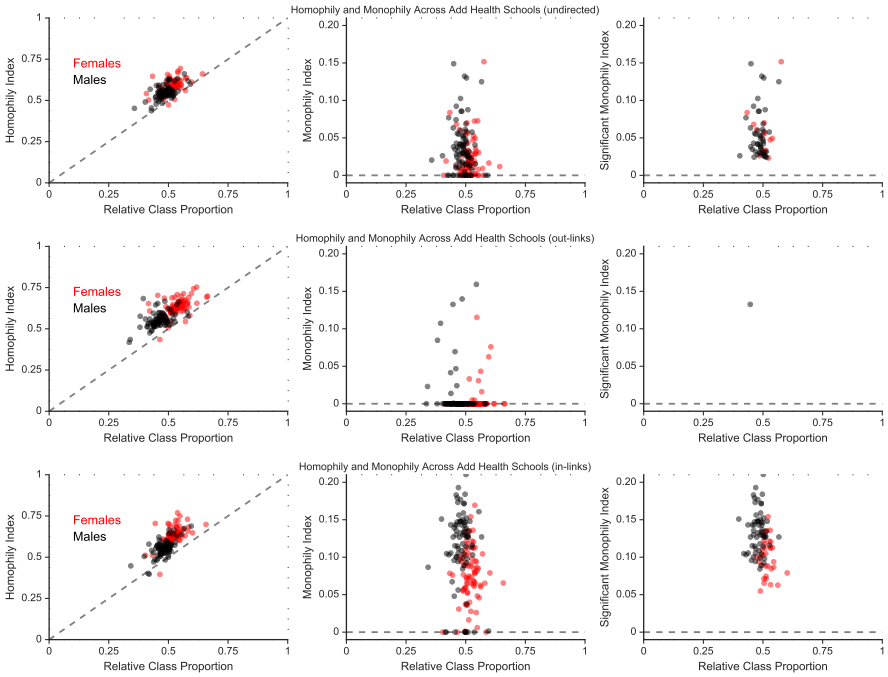
directed graph is defined by an adjacency matrix $A_{ij} = 1$ if student i *was nominated by* student j and is not restricted as $0 \leq \sum_j A_{ij} \leq N$. Due to the restriction on the out-directed degree sequences, we expect the gender preferences to be *underdispersed*, which is evident in Supplementary Figure 17. Lastly, we drop one school (School #27) due to the high proportion of male students (99.8%) at that school.

Focusing on a single representative school, School #23 (for additional schools see Supplementary Figure 19), we evaluate the variance of the empirical distributions for individuals to nominate same-gender friends relative to a null model as shown in Supplementary Figure 16, which compares to Figure 1A (for the Amherst College Facebook network) in the main paper. Then evaluating the relative performance of LINK-logistic regression [15] on the undirected versus directed degree sequences, we observe that overdispersion again drives the improved performance of LINK in the directed setting. Applying a majority vote classifier in the directed setting is less clear for k-hop at $k > 1$ as its not clear how to propagate labels forward are backward. We leave directed majority vote classifiers as beyond the scope of the present paper. Note, we are able to rely on LINK to compare the relative predictability of attributes along with obtaining a sense of which nodes are predictive of gender based on the β_j LINK coefficients.

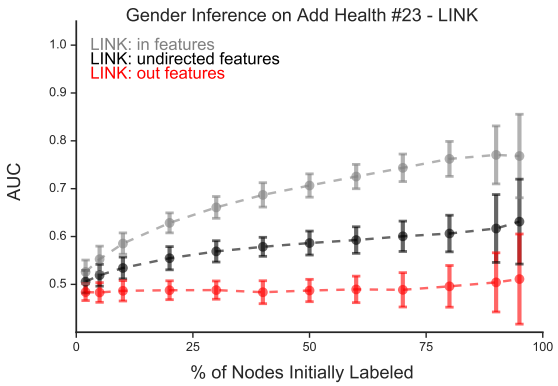
We observe the classification performance of LINK as being driven by *in*-nominations. That is, a machine learning model based on the feature that user i is nominated by all females will be more useful than a model based on the feature on who user i nominates. Therefore, in Supplementary Figure 18 we observe that the "in features" created based on a model fit on the out-directed adjacency matrix is the most useful in learning the relationship of correlating received nominations from particular genders. We attribute LINK's limited performance on the undirected graph as shown in Supplementary Figure 18 to the underdispersion inherent in the Add Health data collection process.



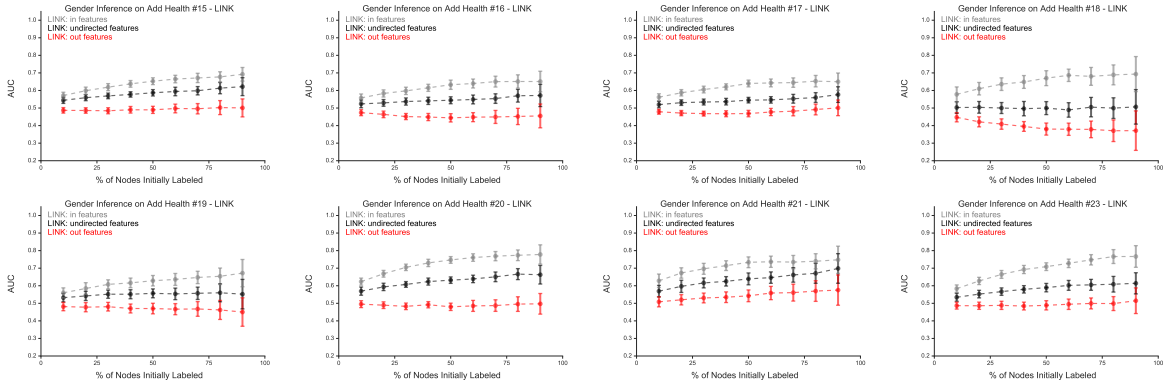
Supplementary Figure 16: For Add Health School #23, we compare the variance of the empirical *in*-class preference distribution (filled bars) relative to the simulated null distribution (solid lines) for the undirected network, out-link network, and in-link network. We observe that the out-link network is underdispersed, which is in part due to the restriction on nominating male and female friends. Meanwhile, we observe the in-link network to be overdispersed which seems to be due to a larger than expected number of individuals nominating all male or all female friends.



Supplementary Figure 17: (Left) Homophily index $\hat{h}_r$, (Middle) monophily index $\hat{\phi}_r$, and (Right) statistically significant monophily index values at the 0.001 level across Add Health schools, in three different arrangements: (Top) undirected, (Center) out-directed, and (Bottom) in-directed.

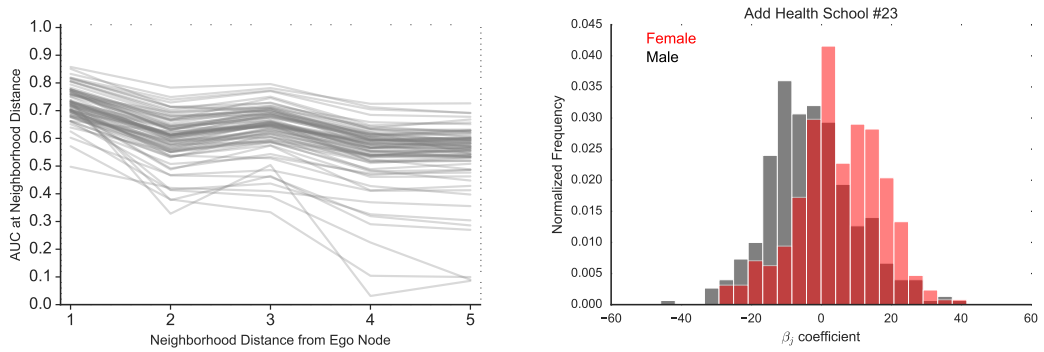


Supplementary Figure 18: Gender classification on school #23 in Add Health with $n_F = 309$, 302, 291 and $n_M = 369$, 337, 324 for the different undirected, in-directed, and out-directed graph versions, respectively.



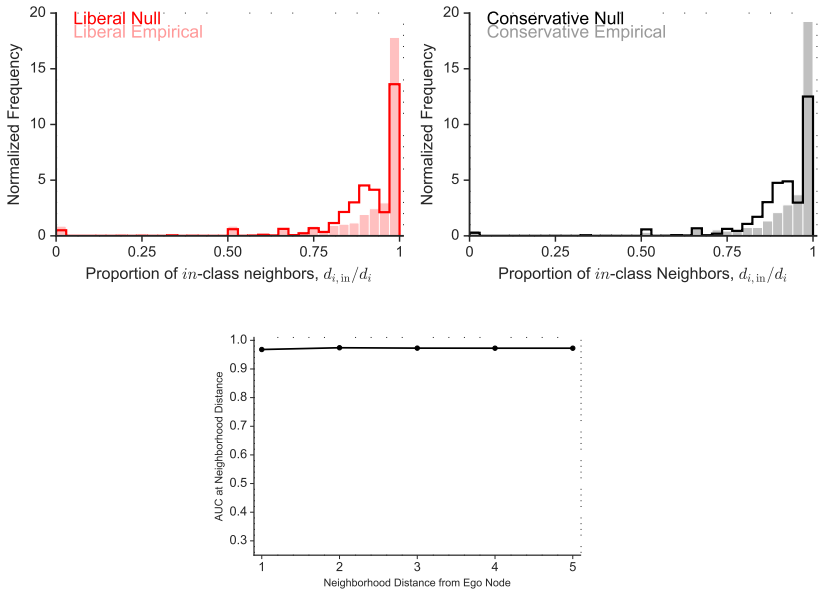
Supplementary Figure 19: Gender classification on a sample of other Add Health schools, illustrating the relative performance of (undirected, in-directed, and out-directed) LINK where we use a stratified random sampling set-up for train/test splits.

We compare a majority vote classifier across the Add Health schools by first converting the directed networks to undirected networks. We then compare AUC performance on the fully labeled network at k-hops, shown in Supplementary Figure 20. Here AUC performance is greater at 1-hop than it is 2-hop. Given the constraints in the out-link nominations, the performance at 2-hop is likely limited on the undirected network which combines both the in- and out-link signals. We compare the β_j coefficients from LINK for the Add Health School #23 as we previously evaluated on the oSBM in Supplementary Figure 11. Here we note that the predictive performance of LINK appears to be driven by both homophily due the slight separation in the histograms and monophily due to the overlap as shown in Supplementary Figure 20.



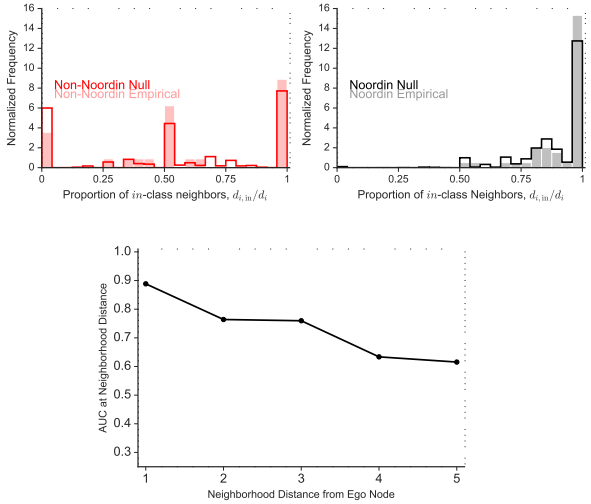
Supplementary Figure 20: Interpretation of gender classification for Add Health School. For Add Health School #23, we compare the majority vote classifier performance at k-hop neighborhood distance (left) on the undirected graph. As already noted, the out-degree constraints likely limits the performance of a 2-hop based classifier on the undirected version of this graph. Based on evaluating the coefficients from a learned LINK logistic regression model (right) we observe that predictions are both homophily and monophily driven as can be seen by this figure's similarity to the yesB-noO graph in Supplementary Figure 11.

3.4 Overdispersion among Political Blogs (political affiliation)



Supplementary Figure 21: (Top) Illustration in political affiliation preferences between Liberal and Conservative classes. We compute the empirical distribution for each class (filled bars) and compare to a null distribution (solid lines). (Bottom) Varying the number of k at which k-hop MV is performed the AUC for k-hop MV on fully labelled network when predicting political affiliation.

3.5 Overdispersion among Persons of Interest (Noordin Top)



Supplementary Figure 22: (Top) Illustration in Noordin Top group membership between non-member and member classes. We compute the empirical distribution for each class (filled bars) and compare to a null distribution (solid lines). (Bottom) Varying the number of k at which k-hop MV is performed the AUC for k-hop MV on fully labelled network when predicting membership.

3.6 Homophily and Monophily Summary Statistics and Visualization

This section provides a table summarizing the networks considered in the main paper (Amherst College, Political Blogs, and Noording Top). In Supplemenatry Table 3, we report the number of nodes in each attribute class along with the homophily and monophily indices.

Network dataset	**Attribute**	$\lvert V_r \rvert$	$\hat{h}_r$	$\hat{\phi}_r$
Amherst College	Female (F) Male (M)	$\lvert V_F \rvert$=1,015 $\lvert V_M \rvert$=1,017	$\hat{h}_F^{***}$=0.55 $\hat{h}_M^{***}$=0.51	$\hat{\phi}_F^{***}$=0.04 $\hat{\phi}_M^{***}$=0.04
Political Blogs	Liberal (L) Conservative (C)	$\lvert V_L \rvert$=586 $\lvert V_C \rvert$=636	$\hat{h}_L^{***}$=0.90 $\hat{h}_C^{***}$=0.91	$\hat{\phi}_L^{***}$=0.23 $\hat{\phi}_C^{***}$=0.19
Noordin Top	Member (M) Non-Member (N)	$\lvert V_M \rvert$=47 $\lvert V_N \rvert$=27	$\hat{h}_M^{***}$=0.89 $\hat{h}_N$= 0.55	$\hat{\phi}_M$=0.02 $\hat{\phi}_N$=0.00

Supplementary Table 3: We report summary statistics for network size, homophily indices $\hat{h}_r$, and monophily indices $\hat{\phi}_r$ evaluated for three demonstration networks: gender on a FB100 network (Amherst College), political affiliation of political blogs, and terrorist group affiliation in a communication network. For the homophily and monophily indices, we indicate statistical significance ($^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$).

Supplementary Note 4: Sampling graphs from the Overdispersed Stochastic Block Model (oSBM)

This section introduces an algorithm for sampling graphs from an overdispersed stochastic block model on k blocks assuming a latent Beta distribution on friendship preferences and adopting the Beta parameterization in [8]. The algorithm takes as input parameters to control the block structure, p_{in} and p_{out}, as well as parameters to control the dispersion, ϕ^*_{in} and ϕ^*_{out}. Note we assume the same parameter value for $p_{in,r}$ across all attribute classes r, so denote this by p_{in} instead of $p_{in,r}$ for a given class r. In settings where one wants to preserve a given overall average node degree $\bar{d} = \frac{1}{N} \cdot \sum_{i=1}^{N} d_i$, we give the following parameterization for p_{in} and p_{out} that relies on a block structure parameter $\lambda \geq 1$ and assumes $k > 1$ blocks:

$$\frac{\lambda}{N} = p_{\text{in}} / \ \bar{d} \quad \Longrightarrow \quad p_{\text{in}} = \lambda \cdot \frac{\bar{d}}{N} \tag{78}$$

$$\bar{d} = \frac{1}{N} \cdot \sum_{i=1}^{k} n_i \cdot (p_{\text{in}} \cdot n_i + p_{\text{out}} \cdot \sum_{j=1, j \neq i}^{k} n_j)) \quad \Longrightarrow \quad p_{\text{out}} = \frac{\bar{d} \cdot N - p_{\text{in}} \cdot \sum_{i=1}^{k} {n_i}^2}{\sum_{i=1}^{k} n_i \cdot (\sum_{j=1, j \neq i}^{k} n_j)}. \tag{79}$$

Algorithm 1 Sample from Overdispersed Block Model

procedure SAMPLE FROM OVERDISPERSED BLOCK MODEL
Input: Given $k \geq 2$ mutually exclusive attribute class labels $\{a_1,...,a_k\}$ for unique class blocks of size $n_1,...,n_k$ nodes, respectively, let $N = \sum_{j=1}^{k} n_j$. Also given: block structure $p_{\text{in}}, p_{\text{out}}$, dispersion parameters $\phi^*_{\text{in}}, \phi^*_{\text{out}}$.
    Model affinity probabilities for *in*- and *out*-class degree distributions using latent Beta distribution [8] with parameters $0 < p_{\text{in}} < 1$ and $0 < p_{\text{out}} < 1$:
    Create *in*-class parameters:
        set $\alpha_{\text{in}} = p_{\text{in}} \cdot (\frac{1}{\phi^*_{\text{in}}}) \cdot (1 - \phi^*_{\text{in}})$
        set $\beta_{\text{in}} = (1 - p_{\text{in}}) \cdot \left(\frac{1}{\phi^*_{\text{in}}}\right) \cdot (1 - \phi^*_{\text{in}})$
    Create *out*-class parameters:
        set $\alpha_{\text{out}} = p_{\text{out}} \cdot \left(\frac{1}{\phi^*_{\text{out}}}\right) \cdot (1 - \phi^*_{\text{out}})$
        set $\beta_{\text{out}} = (1 - p_{\text{out}}) \cdot \left(\frac{1}{\phi^*_{\text{out}}}\right) \cdot (1 - \phi^*_{\text{out}})$
    for each attribute class $\{1, ..., k\}$ **do**
        for each node i in specific class r **do**
            $p_{i,\text{in}} \sim \text{Beta}(\alpha_{\text{in}}, \beta_{\text{in}})$
            $d_{i,\text{in}} = p_{i,\text{in}} \cdot n_r$
            $p_{i,\text{out}} \sim \text{Beta}(\alpha_{\text{out}}, \beta_{\text{out}})$
            $d_{i,\text{out}} = p_{i,\text{out}} \cdot \sum_{j=1, j \neq r}^{k} n_j = p_{i,\text{out}} \cdot (N - n_r)$
    for each affiliation pair a_i, a_j s.t. $a_i \leq a_j$ **do**
        if $a_i = a_j = r$ **then**
            Create Chung-Lu graph with expected degree sequence $(d_{i,\text{in}})$ $\forall$ $i \in r$:
            for each node pair $i, j \in r$, $i \leq j$: **do**
                $A_{ij} | p_{i,\text{in}}, p_{j,\text{in}} \sim \text{Bern}\left(\frac{d_{i,\text{in}} d_{j,\text{in}}}{n_r^2 \cdot p_{\text{in}}}\right)$
                Set $A_{ji} := A_{ij}$
        if $a_i = r \neq a_j = s$ **then**
            Create bipartite Chung-Lu graph (adopting [4]) with expected degree sequence $(d_{i,\text{out}})$ and $(d_{j,\text{out}})$ $\forall i \in r$, $\forall j \in s$:
            for each node pair $i \in r$, $j \in s$: **do**
                $A_{ij} | p_{i,\text{out}}, p_{j,\text{out}} \sim \text{Bern}\left(\frac{d_{i,\text{out}} d_{j,\text{out}}}{(N - n_r) \cdot (N - n_s) \cdot p_{\text{out}}}\right)$,
                Set $A_{ji} := A_{ij}$
    Output: Symmetric adjacency matrix A.

4.1 Properties based on oSBM parameterization.

We assume an underlying model where nodes have an individual affinity for *in*- and *out*-class friends where the mean *in*-class affinity is p_{in} and the mean *out*-class affinity is p_{out}. For all nodes i in class r, we draw *in*- and *out*-class affinity probabilities from a Beta distribution with the following parameterization to preserve the mean *in*- and *out*- affinities and the given dispersion parameters ϕ_{in} and ϕ_{out}:

$p_{i,\text{in}} \sim \text{Beta}(\alpha_{\text{in}}, \beta_{\text{in}})$

$$\begin{aligned} \text{where } \alpha_{\text{in}} &= p_{\text{in}} \cdot \left(\frac{1}{\phi_{\text{in}}}\right) \cdot (1-\phi_{\text{in}}) && (80) \\ \beta_{\text{in}} &= (1-p_{\text{in}}) \cdot \left(\frac{1}{\phi_{\text{in}}}\right) \cdot (1-\phi_{\text{in}}) = \left(\frac{1}{\phi_{\text{in}}}\right) \cdot (1-\phi_{\text{in}}) - \alpha_{\text{in}}, && (81) \end{aligned}$$

$p_{i,\text{out}} \sim \text{Beta}(\alpha_{\text{out}}, \beta_{\text{out}})$

$$\begin{aligned} \text{where } \alpha_{\text{out}} &= p_{\text{out}} \cdot \left(\frac{1}{\phi_{\text{out}}}\right) \cdot (1-\phi_{\text{out}}) && (82) \\ \beta_{\text{out}} &= (1-p_{\text{out}}) \cdot \left(\frac{1}{\phi_{\text{out}}}\right) \cdot (1-\phi_{\text{out}}) = \left(\frac{1}{\phi_{\text{out}}}\right) \cdot (1-\phi_{\text{out}}) - \alpha_{\text{out}}. && (83) \end{aligned}$$

4.2 Confirm mean attribute affinity is preserved.

For *in*-class attribute affinity probabilities, we confirm that the above parameterization of the latent Beta distribution is such that $\mathbb{E}[p_{i,\text{in}}] = p_{\text{in}}$:

$$\mathbb{E}[p_{i,\text{in}}] = \frac{\alpha_{\text{in}}}{\alpha_{\text{in}} + \beta_{\text{in}}} = \frac{p_{\text{in}} \cdot \left(\frac{1}{\phi_{\text{in}}}\right) \cdot (1-\phi_{\text{in}})}{\left(\frac{1}{\phi_{\text{in}}}\right) \cdot (1-\phi_{\text{in}})} = p_{\text{in}}. \tag{84}$$

A similar check shows that the *out*-class affinity probability is such that $\mathbb{E}[p_{i,\text{out}}] = p_{\text{out}}$.

4.3 Confirm variance of attribute affinity is preserved.

For *in*-class affinity probabilities, we confirm that the above parameterization is such that ϕ_{in} introduces extra binomial variation such that $\text{Var}[p_{i,\text{in}}] = \phi_{\text{in}} \cdot p_{\text{in}} \cdot (1-p_{\text{in}})$:

$$\begin{aligned} \text{Var}[p_{i,\text{in}}] &= \frac{\alpha_{\text{in}}\beta_{\text{in}}}{(\alpha_{\text{in}}+\beta_{\text{in}})^2(\alpha_{\text{in}}+\beta_{\text{in}}+1)} && (85) \\ &= \frac{\alpha_{\text{in}}}{\alpha_{\text{in}}+\beta_{\text{in}}} \frac{\beta_{\text{in}}}{\alpha_{\text{in}}+\beta_{\text{in}}} \frac{1}{\alpha_{\text{in}}+\beta_{\text{in}}+1} && (86) \\ &= p_{\text{in}} \cdot (1-p_{\text{in}}) \cdot \frac{1}{\frac{1}{\phi_{\text{in}}} \cdot (1-\phi_{\text{in}}) + 1} && (87) \\ &= p_{\text{in}} \cdot (1-p_{\text{in}}) \cdot \frac{1}{\frac{1}{\phi_{\text{in}}} - 1 + 1} && (88) \\ &= p_{\text{in}} \cdot (1-p_{\text{in}}) \cdot \phi_{\text{in}}. && (89) \end{aligned}$$

A similar check for the out-community affinity probability shows that $\text{Var}[p_{i,\text{out}}] = \phi_{\text{out}} \cdot p_{\text{out}} \cdot (1-p_{\text{out}})$.

4.4 Confirm that expected degrees are approximately preserved.

We confirm the oSBM parameterization for edge probabilities, outlined in Algorithm 1 such that $P(A_{ij} = 1|d_{i,\text{in}}, d_{j,\text{in}})$ for *in*-class edges or $P(A_{ij} = 1|d_{i,\text{out}}, d_{j,\text{out}})$ for *out*-class edges, approximately preserves $(d_{i,\text{in}})$ and $(d_{i,\text{out}})$ as the class-specific expected degrees, and will therefore also approximately preserve (d_i) as the overall expected degrees.

Let node i be in class r, and we want to show that $\mathbb{E}\left[\sum_{j\in r} A_{ij}|p_{i,\text{in}}, p_{j,\text{in}}\right] = d_{i,\text{in}}$. We have:

$$\begin{aligned}
\mathbb{E}\left[\sum_{j\in r} A_{ij}|p_{i,\text{in}}, p_{j,\text{in}}\right] &= \sum_{j\in r} \mathbb{E}[A_{ij}|p_{i,\text{in}}, p_{j,\text{in}}] && (90)\\
&= \sum_{j\in r} \frac{d_{i,\text{in}} \cdot d_{j,\text{in}}}{n_r^2 \cdot p_{\text{in}}} && (91)\\
&= \frac{d_{i,\text{in}}}{n_r \cdot p_{\text{in}}} \sum_{j\in r} \frac{p_{j,\text{in}} \cdot n_r}{n_r} && (92)\\
&= \frac{d_{i,\text{in}}}{n_r \cdot p_{\text{in}}} \sum_{j\in r} p_{j,\text{in}} && (93)\\
&\approx \frac{d_{i,\text{in}}}{n_r \cdot p_{\text{in}}} (n_r \cdot p_{\text{in}}) && (94)\\
&= d_{i,\text{in}}. && (95)
\end{aligned}$$

Let node i be in class r and $\mathbb{E}\left[\sum_{j\in s\neq r} A_{ij}|p_{i,\text{out}}, p_{j,\text{out}}\right] = d_{i,\text{out}}$, then we have the following:

$$\begin{aligned}
\mathbb{E}\left[\sum_{j\in s\neq r} A_{ij}|p_{i,\text{out}}, p_{j,\text{out}}\right] &= \sum_{j\in s\neq r} \mathbb{E}[A_{ij}|p_{i,\text{out}}, p_{j,\text{out}}] && (96)\\
&= \sum_{j\in s\neq r} \frac{d_{i,\text{out}} \cdot d_{j,\text{out}}}{(N-n_r)\cdot(N-n_s)\cdot p_{\text{out}}} && (97)\\
&= \sum_{j\in s\neq r} \frac{p_{i,\text{out}} \cdot (N-n_r) \cdot p_{j,\text{out}} \cdot (N-n_s)}{(N-n_r)\cdot(N-n_s)\cdot p_{\text{out}}} && (98)\\
&= \frac{p_{i,\text{out}}}{p_{\text{out}}} \cdot \sum_{j\in s\neq r} p_{j,\text{out}} && (99)\\
&\approx \frac{p_{i,\text{out}}}{p_{\text{out}}} \cdot (N-n_r) \cdot p_{\text{out}} && (100)\\
&= p_{i,\text{out}} \cdot (N-n_r) && (101)\\
&= d_{i,\text{out}}. && (102)
\end{aligned}$$

Supplementary References

[1] Andrew P Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159, 1997.

[2] Nitesh V Chawla, Nathalie Japkowicz, and Aleksander Kotcz. Special issue on learning from imbalanced data sets. *ACM SIGKDD Explorations Newsletter*, 6(1):1–6, 2004.

[3] Paul H Garthwaite, Ian T Jolliffe, and Byron Jones. *Statistical Inference*. Oxford University Press on Demand, 2002.

[4] Daniel B Larremore, Aaron Clauset, and Abigail Z Jacobs. Efficiently inferring community structure in bipartite networks. *Physical Review E*, 90(1):012805, 2014.

[5] Saskia Le Cessie and Johannes C Van Houwelingen. Ridge estimators in logistic regression. *Applied Statistics*, pages 191–201, 1992.

[6] Frank McSherry. Spectral partitioning of random graphs. In *Proceedings 2001 IEEE International Conference on Cluster Computing*, pages 529–537, 2001.

[7] Mark EJ Newman. Assortative mixing in networks. *Physical Review Letters*, 89(20):208701, 2002.

[8] Ross L Prentice. Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors. *Journal of the American Statistical Association*, 81(394):321–327, 1986.

[9] Michael D Resnick, Peter S Bearman, Robert Wm Blum, Karl E Bauman, Kathleen M Harris, Jo Jones, Joyce Tabor, Trish Beuhring, Renee E Sieving, Marcia Shew, et al. Protecting adolescents from harm: findings from the national longitudinal study on adolescent health. *JAMA*, 278(10):823–832, 1997.

[10] Saharon Rosset, Ji Zhu, and Trevor Hastie. Boosting as a regularized path to a maximum margin classifier. *Journal of Machine Learning Research*, 5:941–973, 2004.

[11] Vito Signorile and Robert M O'Shea. A test of significance for the homophily index. *American Journal of Sociology*, 70(4):467–470, 1965.

[12] John A Swets. Measuring the accuracy of diagnostic systems. *Science*, 240(4857):1285–1293, 1988.

[13] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[14] David A Williams. Extra-binomial variation in logistic linear models. *Applied Statistics*, pages 144–148, 1982.

[15] Elena Zheleva and Lise Getoor. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In *Proceedings of the 18th International Conference on World Wide Web*, pages 531–540, 2009.