Higher-order homophily is combinatorially impossible

Austin Benson · Cornell University
HONS@Networks 2021

Joint work with
Nate Veldt (Cornell → Texas A&M) &
Jon Kleinberg (Cornell)
People tend to connect to similar others.

*Birds of a feather: Homophily in social networks*, McPherson, Smith-Lovin, & Cook, 2001. *Mixing Patterns in Networks*, Newman, 2003.
Homophily is used to understand groups.

The duality of persons and groups, Breiger, 1974.
Sex and race homogeneity in naturally occurring groups, Mayhew et al., 1995.
Testing a dynamic model of social composition, McPherson & Rotolo, 1996.
Community-Affiliation Graph Model for Overlapping Network Community Detection, Yang & Leskovec, 2012.
Even though homophily is used to understand groups, we measure it from pairwise interactions.

\[
\begin{align*}
\text{in 1 BG, 2 BB edges (3 total)} \\
\text{in 2 BG, 2 BB edges (4 total)} \\
\text{in 2 BG, 2 BB edges (4 total)} \\
\end{align*}
\]

\[
h(B) = \frac{2 + 2 + 2}{3 + 4 + 4} = \frac{6}{11}
\]

\[
h(G) = \frac{2 + 3 + 3 + 2}{2 + 5 + 4 + 4} = \frac{2}{3}
\]

\text{affinity aka homophily index}

The \textit{baseline} is the probability that a uniformly chosen neighbor is the same class.

\[
b(B) = \frac{2}{6} < h(B) \rightarrow h(B) / b(B) > 1 \rightarrow \text{homophily w/r/t to the blue class}
\]

\[
b(G) = \frac{3}{6} < h(G) \rightarrow h(G) / b(G) > 1 \rightarrow \text{homophily w/r/t to the green class}
\]
We have lots of social data of group interactions.

Communications

Physical proximity

Collaboration

Social media

her-order Homophily is Combinatorially Impossible

Jade Veldt
Center for Applied Math
Cornell University

Austin R. Benson
Computer Science Dept.
Cornell University

Jon Kleinberg
Computer Science Dept.
Cornell University
We propose a homophily metric from group interactions.

The *t*-baseline is the probability that there are *t* of a given class if other 2 are random.

\[ b_1(B) = \frac{4 \text{ choose } 2}{6 \text{ choose } 2} = \frac{2}{5} > h_1(B) \rightarrow h_1(B) / b_1(B) < 1 \]
\[ \rightarrow \text{ no type-1 homophily w/r/t to the blue class} \]

\[ b_2(B) = \frac{2 \text{ choose } 1}{6 \text{ choose } 2} = \frac{8}{15} > h_2(B) \rightarrow h_2(B) / b_2(B) < 1 \]
\[ \rightarrow \text{ no type-2 homophily w/r/t to the blue class} \]

\[ b_3(B) = \frac{1}{6 \text{ choose } 2} = \frac{1}{15} < h_3(B) \rightarrow h_3(B) / b_3(B) > 1 \]
\[ \rightarrow \text{ yes type-3 homophily w/r/t to the blue class} \]
We propose a homophily metric from group interactions.

| degree | 1 | 2 | 3 | 4 | Σ  | 5 | 6 | 7 | Σ  |
|--------|---|---|---|---|----|---|---|---|----|
| type-1 | 0 | 1 | 0 | 1 | 2  | 0 | 1 | 0 | 1  |
| type-2 | 0 | 0 | 1 | 1 | 2  | 1 | 1 | 2 | 4  |
| type-3 | 1 | 2 | 2 | 1 | 6  | 1 | 1 | 1 | 3  |
| Σ      | 1 | 3 | 3 | 3 | 10 | 2 | 3 | 3 | 8  |

\[
\begin{align*}
    h_1(G) &= 0.2, & h_2(G) &= 0.2, & h_3(G) &= 0.6, \\
    b_1(G) &= 0.2, & b_2(G) &= 0.6, & b_3(G) &= 0.2, \\
    h_1(B) &= 0.12, & h_2(B) &= 0.5, & h_3(B) &= 0.38, \\
    b_1(B) &= 0.4, & b_2(B) &= 0.53, & b_3(B) &= 0.07.
\end{align*}
\]
Affinities also have a statistical interpretation.

Hypergraph stochastic block model for size-\(k\) groups and classes \(B\) & \(G\)

- \(p_t = \text{prob. exactly } t \text{ of class } B \text{ in a hyperedge}\)

- Type-\(t\) node degrees are asymptotically independent
- For an observed set of degrees, \(h_t(B)\) is the MLE for \(p_t\)

Monophily in social networks introduces similarity among friends-of-friends
Altenburger & Ugander, 2018.
74,134 papers in 81 CS conferences with 2, 3, or 4 authors each, covering 105,256 total authors, 21.5% of which are female.
Women are more likely to be in majority-female collaborations than by chance. Men are only more likely than chance to be in all-male or 1M–3F collaborations. **Women and men cannot both prefer majority same-gender collaborations more than chance!**

Women exhibit monotonically increasing preferences for more female authors. Men don’t have this pattern. **Women and men cannot both have monotonically increasing majority-gender preferences!**
When two classes of people participate in groups of 3, they cannot both have higher than random preferences for all groups where they are in the majority.

This is not a social finding... it is a combinatorial impossibility of hypergraphs!
242 students at a primary school with gatherings of students if they all made contact within 20 seconds as measured by wearable sensors
Our theory captures these ideas precisely.

In group interactions of size $k$, we say that class $X$ exhibits

- **majority homophily** if $h_t(X) > b_t(X)$ for $t > k/2$;
- **monotonic homophily** if $h_t(X) / b_t(X) > h_{t-1}(X) / b_{t-1}(X)$ for $t > k/2$.

[these are the same if $k = 2$]

**Theorem [Veldt-Benson-Kleinberg 21]**

- For $k$ odd, both classes *cannot* simultaneously exhibit majority homophily or monotonic homophily.
- For $k$ even, both classes *cannot* exhibit majority homophily if $h_{k/2}(X) / b_{k/2}(X) > h_{k/2-1}(X) / b_{k/2-1}(X)$ for at least one class $X$.
- For $k$ even, both classes *can* exhibit majority homophily but need $h_{k/2}(X) > b_{k/2}(X)$ for at least one class $X$.

[these results also covers another homophily measure and many types of baselines]
Intuition. Majority groups for one class are minority groups for the other class.
A weak homophily impossibility result is easy to prove.

No class can have all affinities above baselines, i.e., there cannot be a class where $h_t(X) > b_t(X)$ for $t = 1, 2, \ldots, k$.

Proof. $h_1(X) + \ldots + h_t(X) = 1 = b_1(X) + \ldots + b_t(X)$. 
1,718 congresspersons, 810 / 908 republican / democrat, co-sponsoring 883,105 bills

| group size $k$ |
|---------------|
| 5  | 6  | 7  | 8  | 9  | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Rep. GHI     | 2   | 2   | 2   | 3   | 3   | 3   | 4   | 4   | 4   | 5   | 5   | 5   | 6   | 7   | 7   | 6   |
| Dem. GHI     | 1   | 2   | 2   | 2   | 3   | 3   | 3   | 4   | 4   | 4   | 5   | 5   | 5   | 5   | 6   | 6   | 6   |

*Group Homophily Index (GHI) = number of top affinity scores above baseline*
1,718 congresspersons, 810 / 908 republican / democrat, co-sponsoring 883,105 bills

\[
\frac{\binom{810}{5} \times \binom{908}{5}}{\binom{810}{8} \times \binom{908}{2}} = 7.99
\]
More shopping trips highly focused on clothes or groceries than expected by chance.

More common to go on a clothing-focused trip and get a few groceries than a grocery-focused trip and get a couple of clothing items.

48,480 products purchased at Walmart
8,956 hotels reviewed by 128,494 users on tripadvisor.com

| group size $k$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---------------|---|---|---|---|---|---|---|---|----|----|----|----|
| N. America GHI | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 3  | 3  | 3  | 4  |
| Europe GHI    | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 3 | 3  | 3  | 3  | 3  |
“family portrait” query on Flickr → 1,051 images

Pr(2 boys) = 1/4
Pr(2 girl) = 1/4
Pr(1 boy, 1 girl) = 1/2

Pairwise reduction
graph homophily
Male 0.43
Female 0.41

Understanding Groups of Images of People, Gallagher & Chen, 2009.
“wedding + bride + groom + portrait” query on Flickr → 662 images

Pairwise reduction graph homophily
Male 0.57
Female 0.54
“group shot” or “group photo” or “group portrait” query on Flickr → 963 images

Pairwise reduction graph homophily
Male  0.60
Female 0.58
There is lots of structure when analyzing higher-order interactions where nodes are in one of two classes.

1. Homophily is (in some sense) impossible for higher-order networks.
2. This is a combinatorial fact, so social insights need care.
3. (near-)homogeneous groups are often homophilous: physical contacts, political teams, co-reviews, certain photos
4. Reducing to pairwise destroys insights
Higher-order homophily is combinatorially impossible. Nate Veldt, Austin R. Benson, and Jon Kleinberg. arXiv:2103.11818, 2021.

Code & Data. github.com/nveldt/HypergraphHomophily

Supported by ARO MURI, ARO Award W911NF19-1-0057, NSF Award DMS-1830274, and JP Morgan Chase & Co.