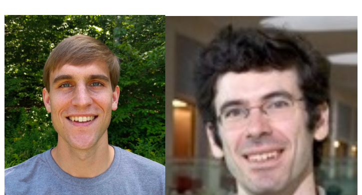
Higher-order homophily is combinatorially impossible

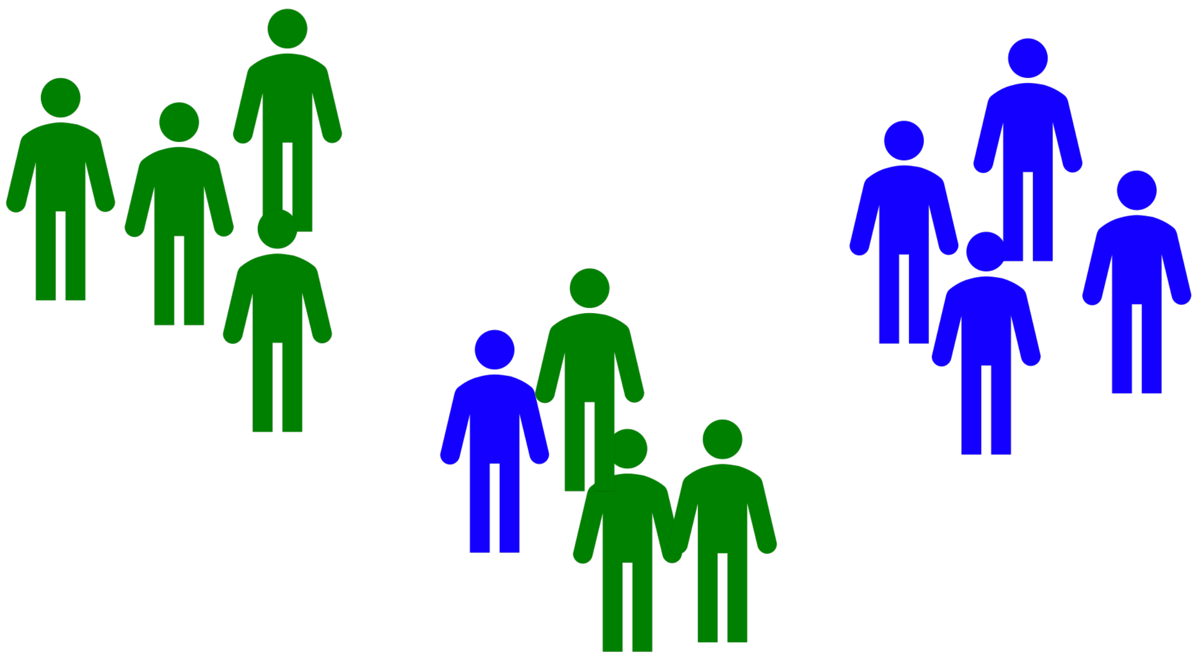
Austin Benson · Cornell University

HONS@Networks 2021



Joint work with
Nate Veldt (Cornell ⟶ Texas A&M) &
Jon Kleinberg (Cornell)

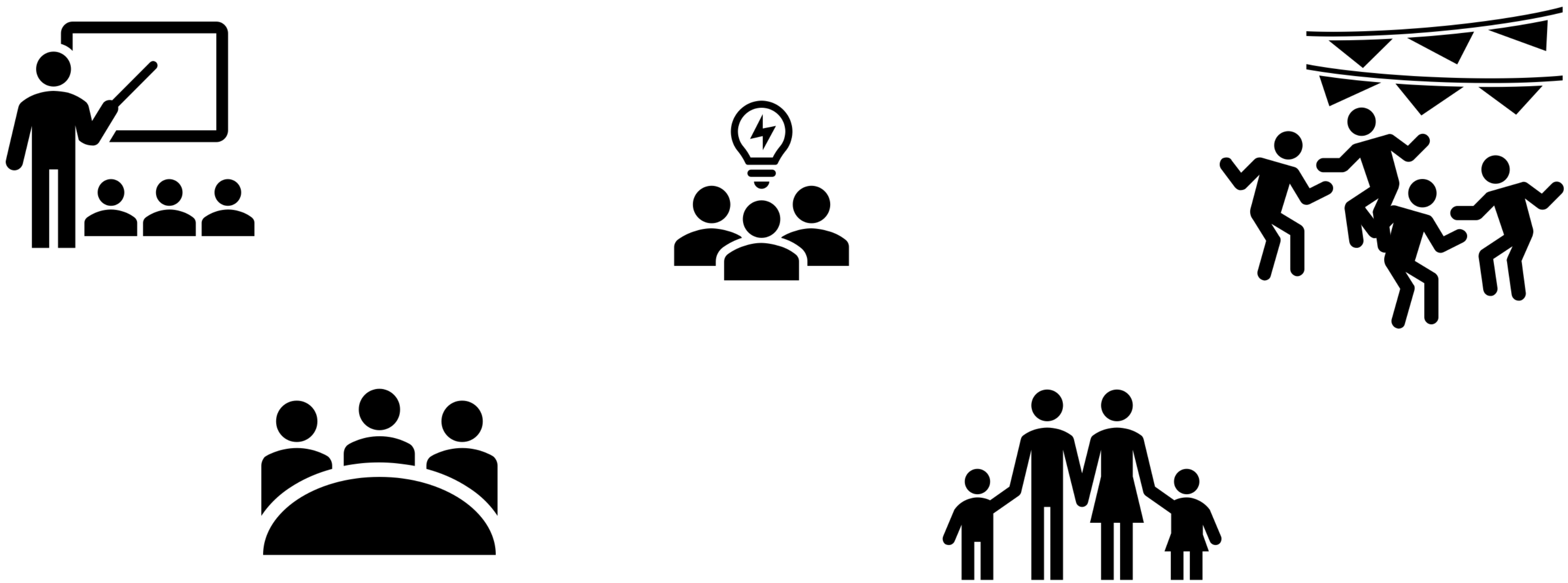
People tend to connect to similar others.



age
race
gender
location
occupation
education level
political affiliation
religious affiliation
attitudes and aspirations

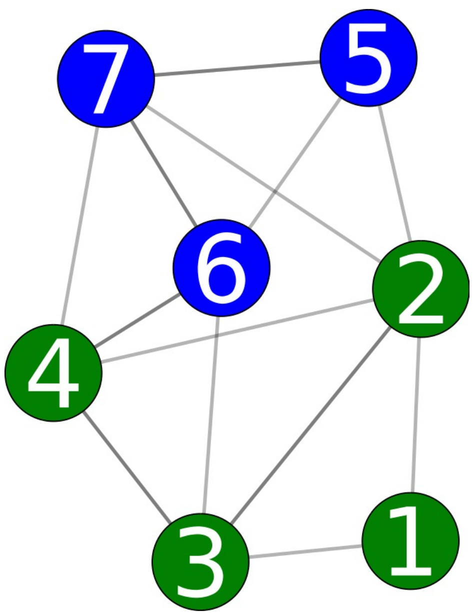
Birds of a feather: Homophily in social networks, McPherson, Smith-Lovin, & Cook, 2001.
Mixing Patterns in Networks, Newman, 2003.

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.



5 in 1 BG, 2 BB edges (3 total)

6 in 2 BG, 2 BB edges (4 total)

7 in 2 BG, 2 BB edges (4 total)

h(B) = (2 + 2 + 2) / (3 + 4 + 4) = 6/11

h(G) = (2 + 3 + 3 + 2) / (2 + 5 + 4 + 4) = 2/3

affinity aka *homophily index*

The *baseline* is the probability that a uniformly chosen neighbor is the same class.

b(B) = 2/6 < h(B) ⟶ h(B) / b(B) > 1 ⟶ homophily w/r/t to the blue class

b(G) = 3/6 < h(G) ⟶ h(G) / b(G) > 1 ⟶ homophily w/r/t to the green class

We have lots of social data of group interactions.

Communications

Physical proximity

Collaboration

her-order Homophily is Combinatorially Impossible

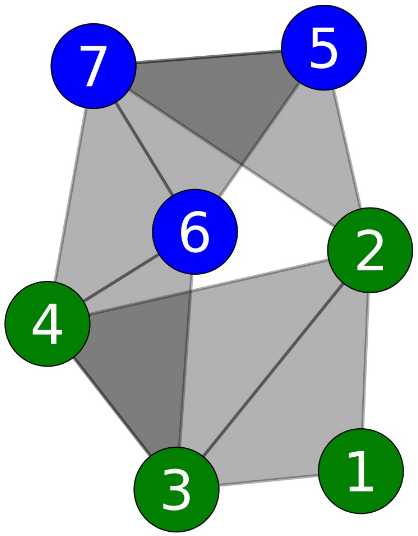
Nate Veldt
Center for Applied Math
Cornell University

Austin R. Benson
Computer Science Dept.
Cornell University

Jon Kleinberg
Computer Science Dept.
Cornell University

Social media

We propose a homophily metric from group interactions.



5 in 0 BGG, 1 BBG, 1 BBB, edges (2 total)

6 in 1 BGG, 1 BBG, 1 BBB edges (3 total)

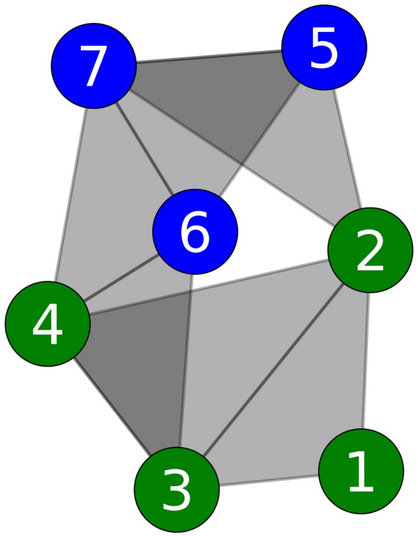
7 In 0 BGG, 2 BBG, 1 BBB edges (3 total)

$h_1(B) = (0 + 1 + 0) / (2 + 3 + 3) = 1/8$
$h_2(B) = (1 + 1 + 2) / (2 + 3 + 3) = 4/8$
$h_3(B) = (1 + 1 + 1) / (2 + 3 + 3) = 3/8$

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

$b_1(B)$ = (4 choose 2) / (6 choose 2) = 2/5 > $h_1(B) \longrightarrow h_1(B) / b_1(B) < 1$
⟶ no *type-1* homophily w/r/t to the blue class
$b_2(B)$ = (2 choose 1) * (4 choose 1) / (6 choose 2) = 8/15 > $h_2(B) \longrightarrow h_2(B) / b_2(B) < 1$
⟶ no *type-2* homophily w/r/t to the blue class
$b_3(B)$ = 1 / (6 choose 2) = 1/15 < $h_3(B) \longrightarrow h_3(B) / b_3(B) > 1$
⟶ yes *type-3* homophily w/r/t to the blue class

We propose a homophily metric from group interactions.



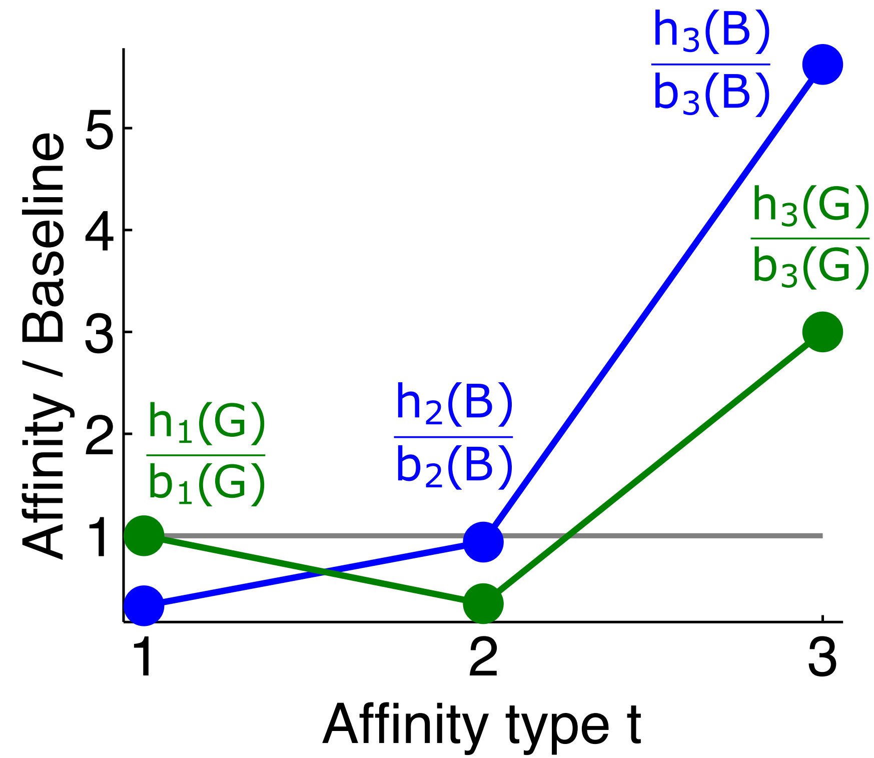
degree	1	2	3	4	Σ	5	6	7	Σ
type-1	0	1	0	1	2	0	1	0	1
type-2	0	0	1	1	2	1	1	2	4
type-3	1	2	2	1	6	1	1	1	3
Σ	1	3	3	3	10	2	3	3	8

$\mathbf{h}_1(\mathrm{G}) = 0.2$ $\quad \mathbf{h}_2(\mathrm{G}) = 0.2$ $\quad \mathbf{h}_3(\mathrm{G}) = 0.6$

$\mathbf{b}_1(\mathrm{G}) = 0.2$ $\quad \mathbf{b}_2(\mathrm{G}) = 0.6$ $\quad \mathbf{b}_3(\mathrm{G}) = 0.2$

$\mathbf{h}_1(\mathrm{B}) = 0.12$ $\quad \mathbf{h}_2(\mathrm{B}) = 0.5$ $\quad \mathbf{h}_3(\mathrm{B}) = 0.38$

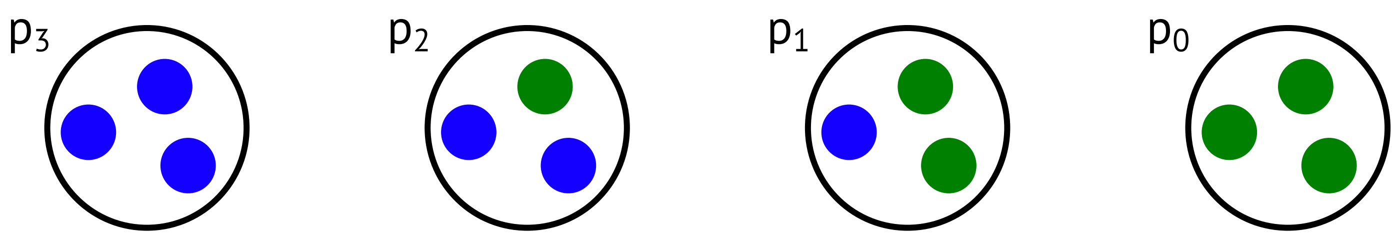
$\mathbf{b}_1(\mathrm{B}) = 0.4$ $\quad \mathbf{b}_2(\mathrm{B}) = 0.53$ $\quad \mathbf{b}_3(\mathrm{B}) = 0.07$



Affinities also have a statistical interpretation.

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

- p_t = prob. exactly *t* of class B 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.

RESEARCH-ARTICLE

Women in computer science research: what is the bibliography data telling us?

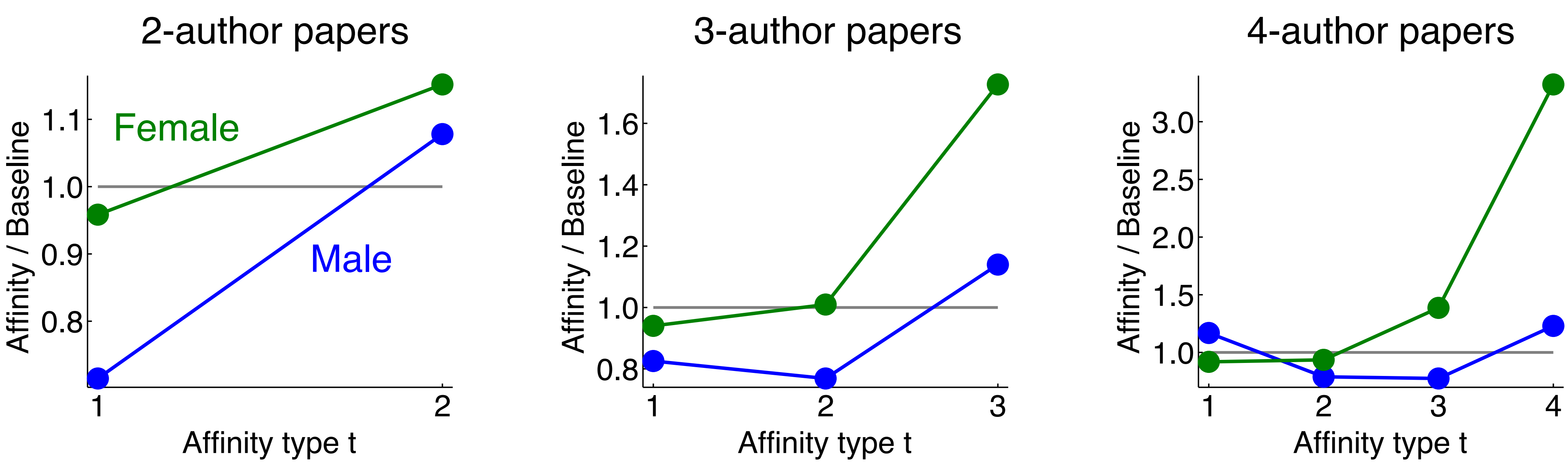
Authors: Swati Agarwal, Nitish Mittal, Rohan Katyal, Ashish Sureka, Denzil Correa

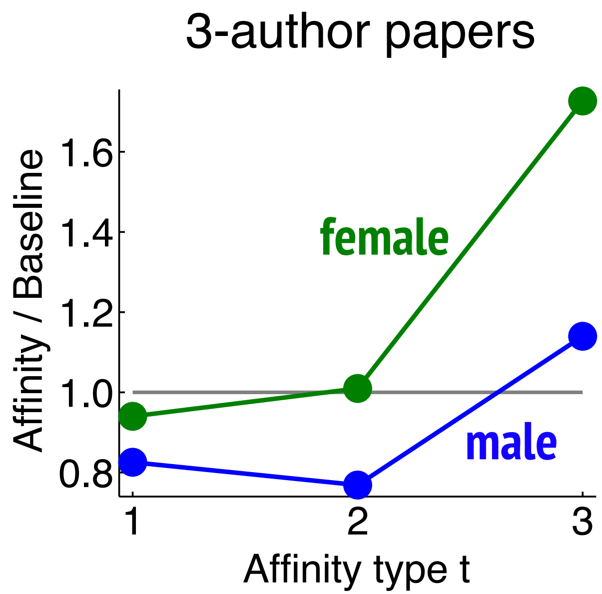
Authors Info & Affiliations

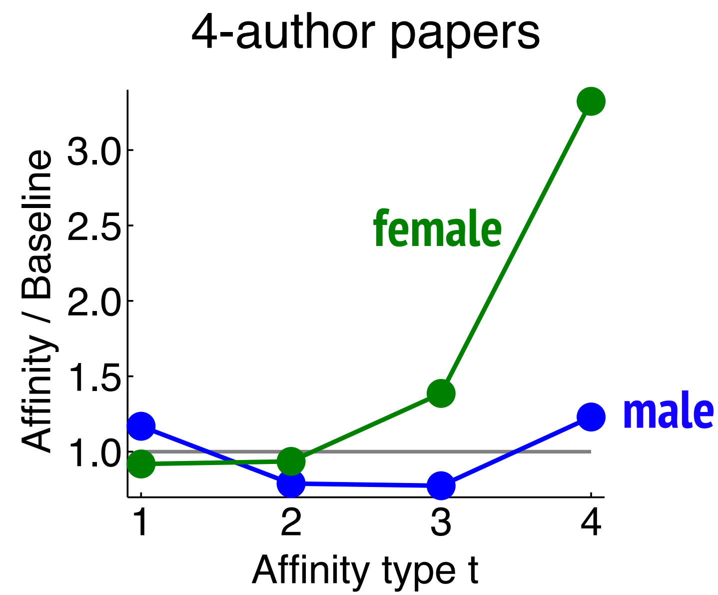
ACM SIGCAS Computers and Society, Volume 46, Issue 1 • 0March 2016 • pp 7–19 • https://doi.org/10.1145/2908216.2908218

Published: 28 March 2016

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





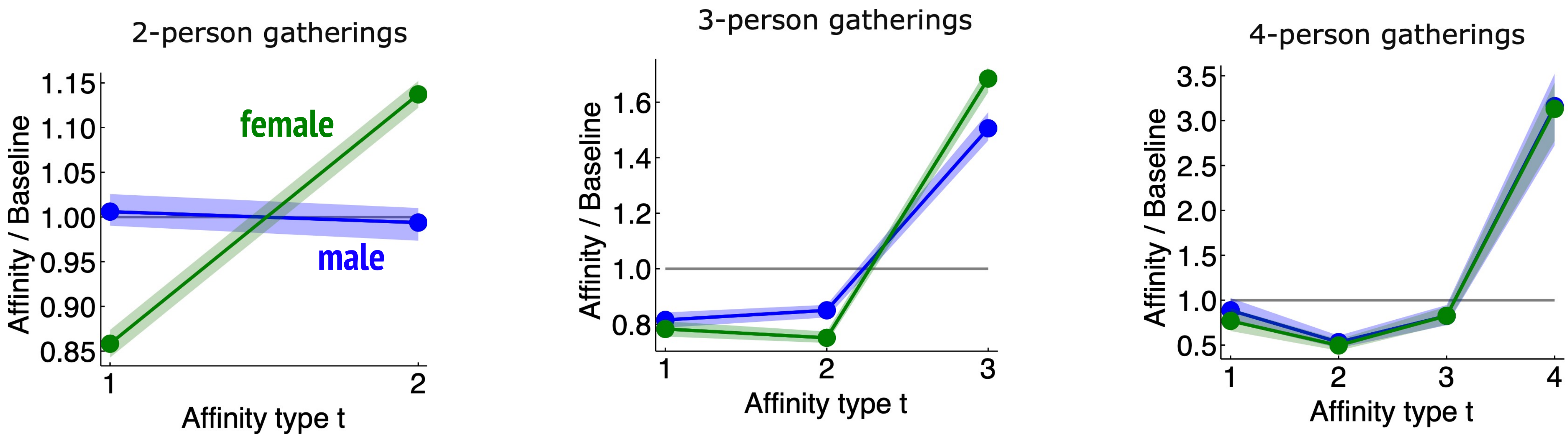


Women are more likely to be in majority-female collaborations than by chance.
Men are only more likely than chance to be in all-male or 1M–3F collaborations.
Women and men cannot both prefer majority same-gender collaborations more than chance!

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!

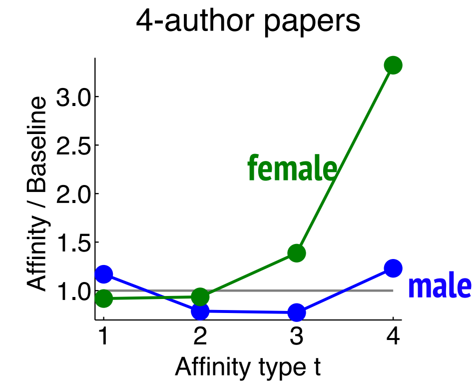
When two classes of people participate in groups of 3, they cannot both have higher than random preferences for all groups where they are in the majority.

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 $t > k / 2$.

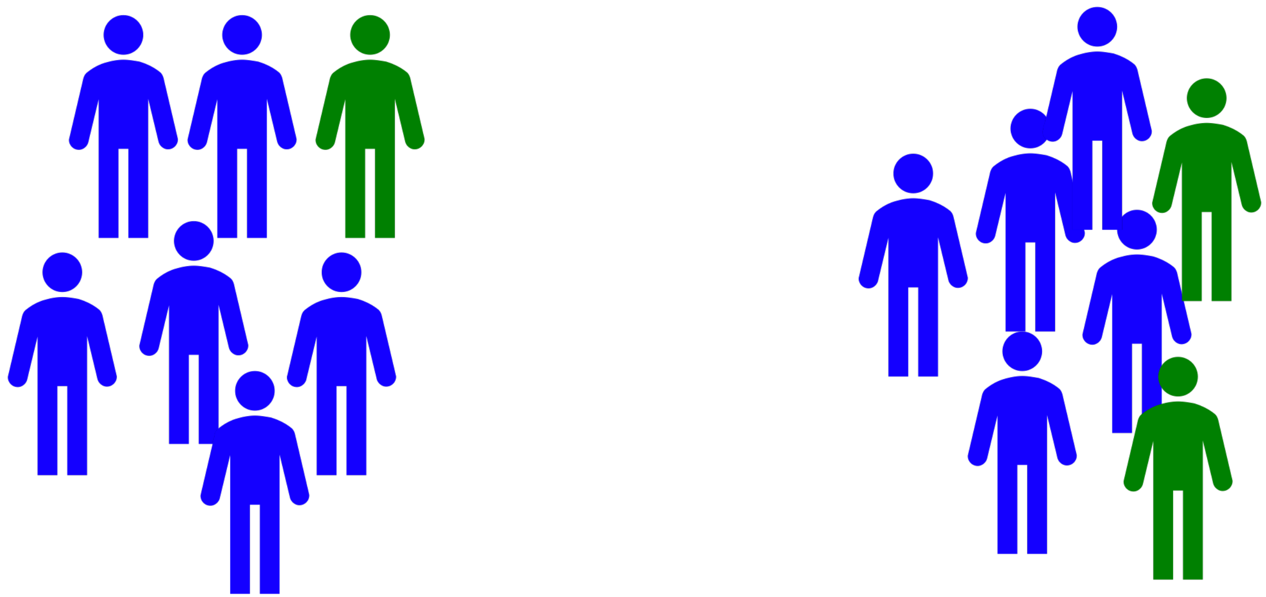
[these are the same if $k = 2$]

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 on 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]

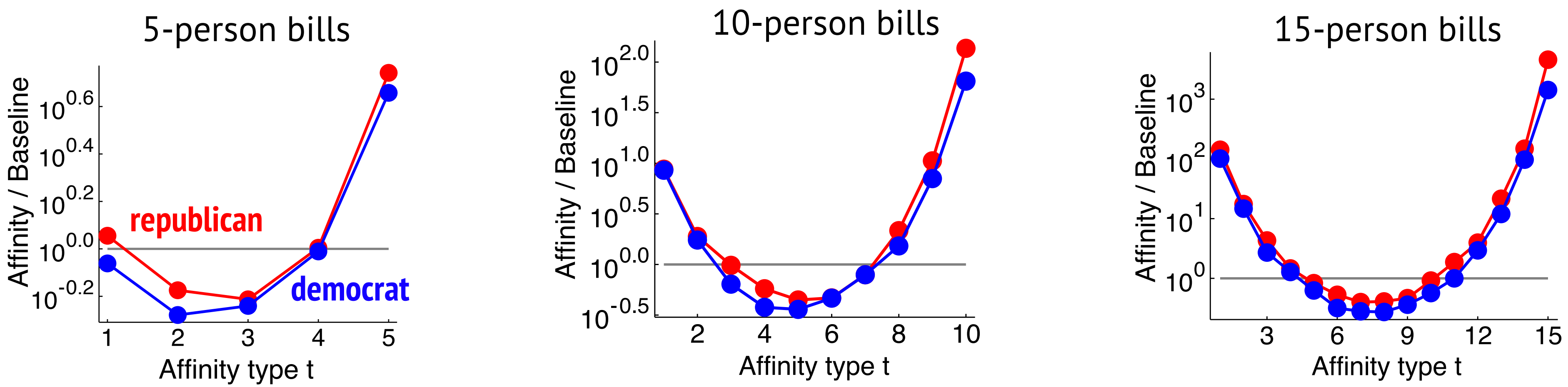
Intuition. **Majority groups for one class are minority groups for the other class.**



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 $t = 1, 2, \ldots, k$.

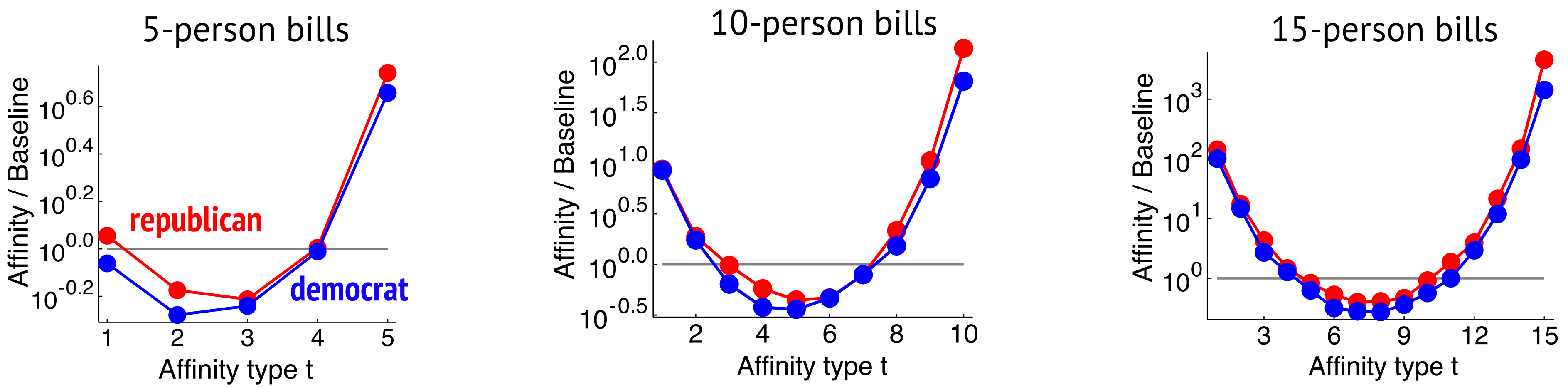
Proof. $h_1(X) + \ldots + h_t(X) = 1 = b_1(X) + \ldots + b_t(X)$.



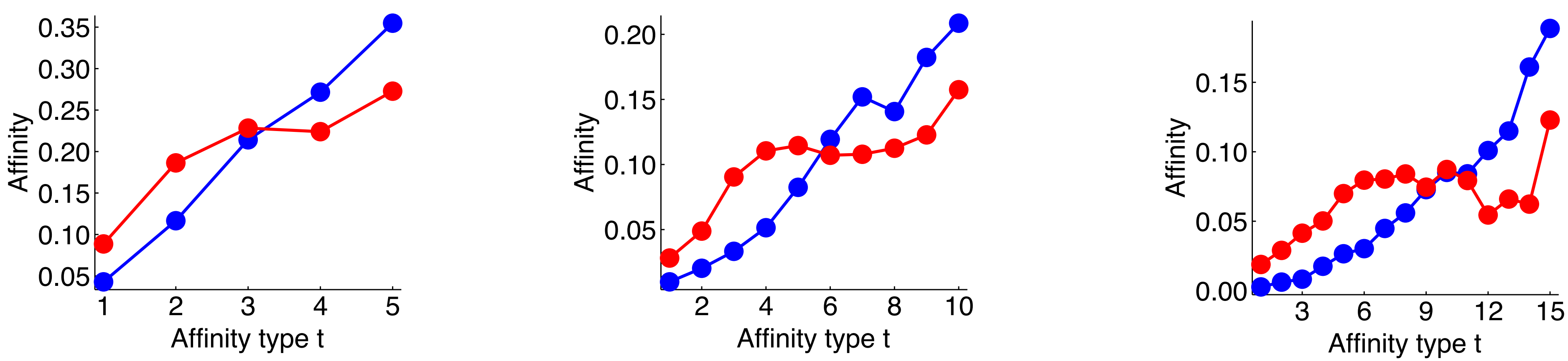
1,718 congresspersons, 810 / 908 republican / democrat, co-sponsoring 883,105 bills

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	6	6	6

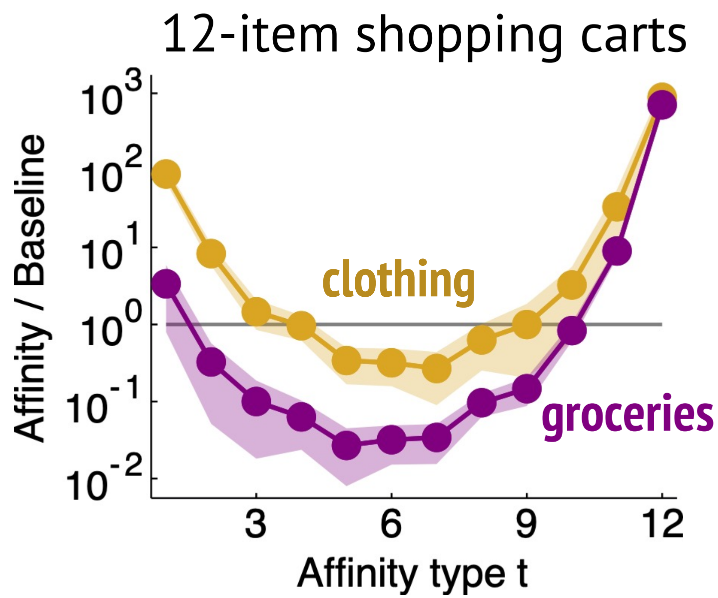
Group Homophily Index (GHI) = number of top affinity scores above baseline



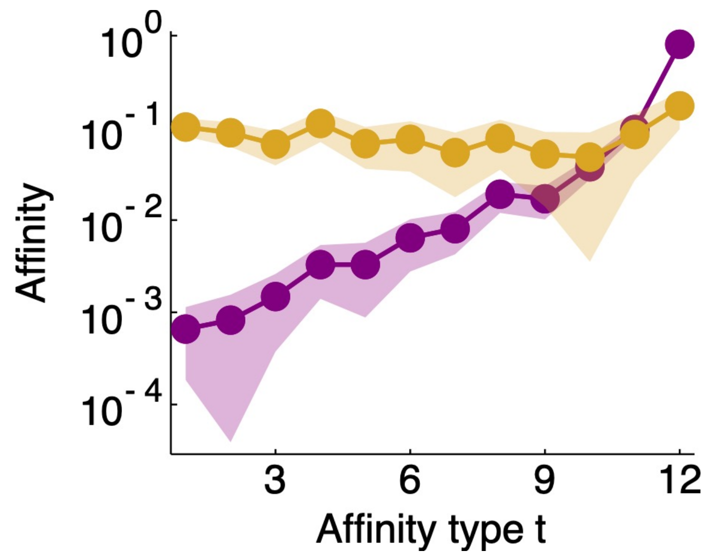
1,718 congresspersons, 810 / 908 republican / democrat, co-sponsoring 883,105 bills



[(810 choose 5) * (908 choose 5)] /
[(810 choose 8) * (908 choose 2)] = 7.99

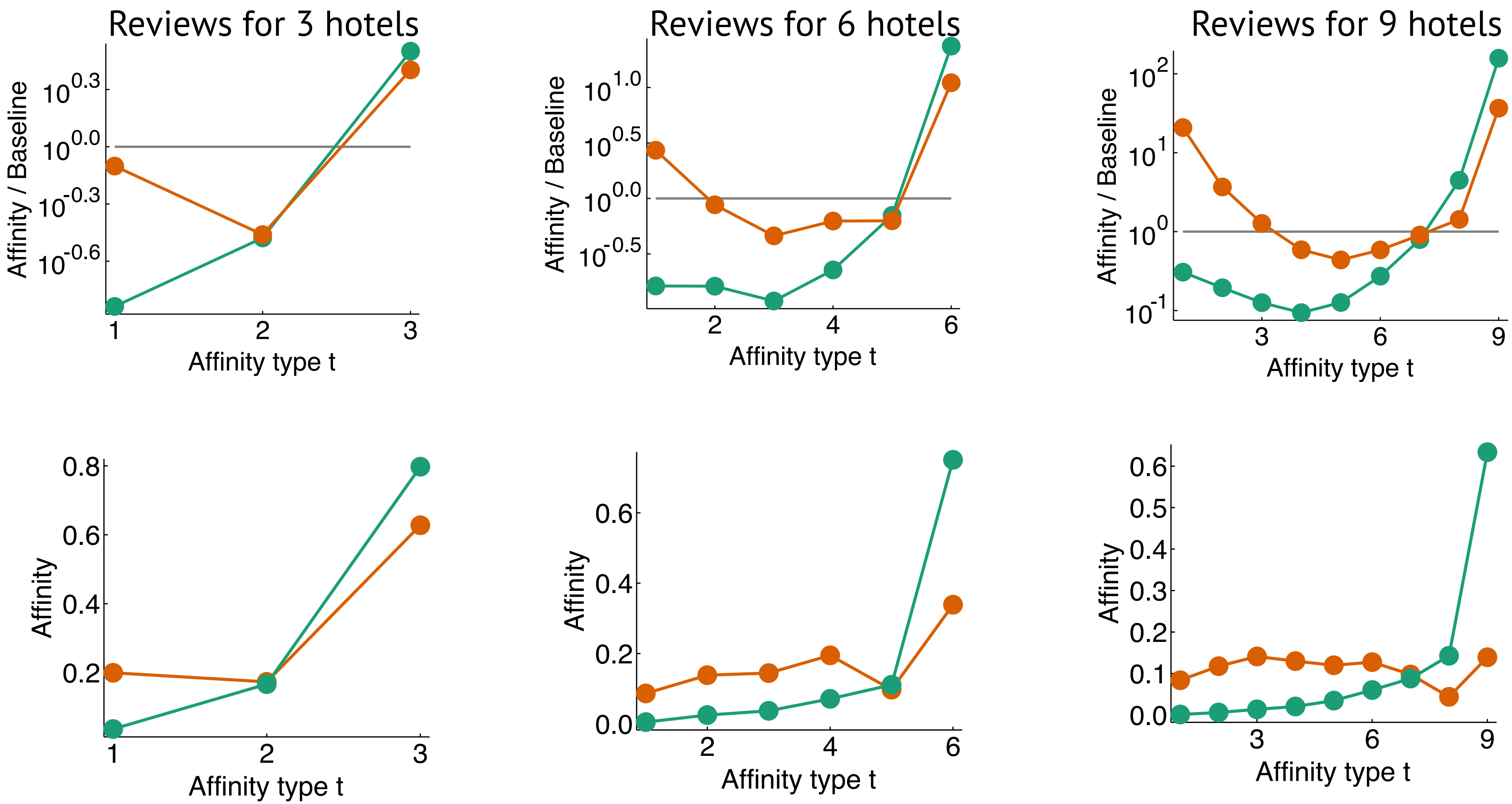


More shopping trips highly focused on clothes or groceries than expected by chance.



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.

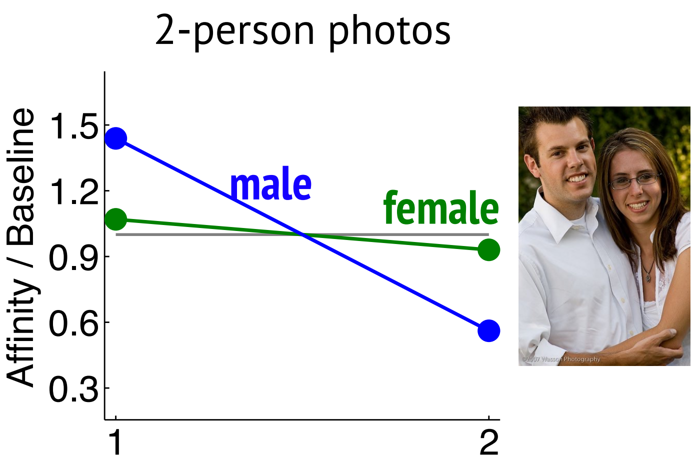
48,480 products purchased at Walmart

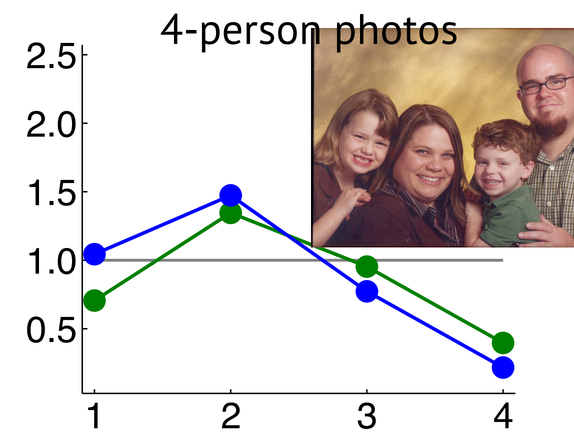


8,956 hotels reviewed by 128,494 users on `tripadvisor.com`

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	1	**2**	3	3	3	3

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





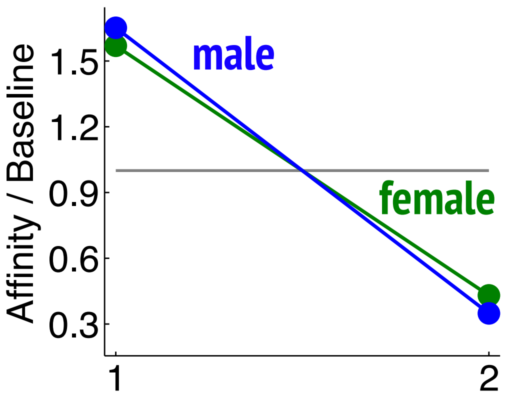
“family portrait” query on Flickr
⟶ 1,051 images

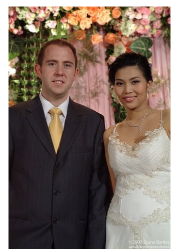
Pairwise reduction graph homophily

Male	0.43
Female	0.41

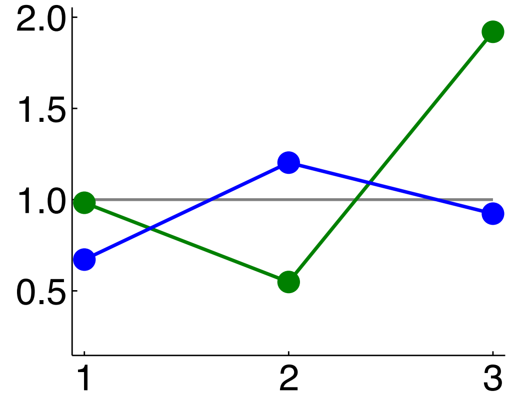
Understanding Groups of Images of People, Gallagher & Chen, 2009.

2-person photos

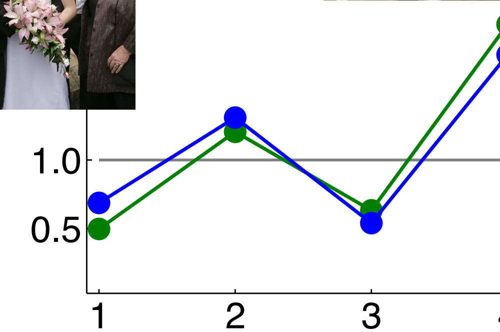




3-person photos



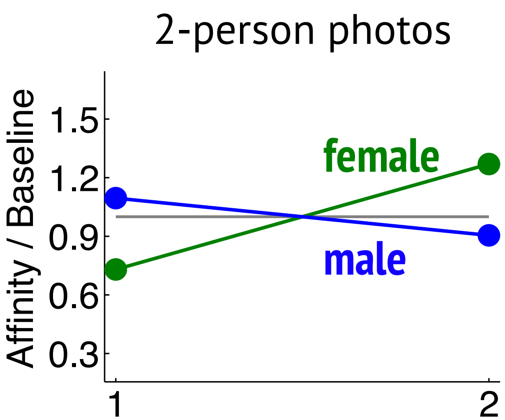
4-person

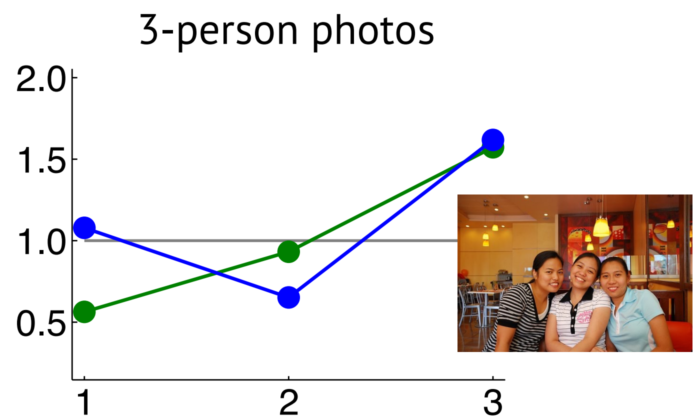


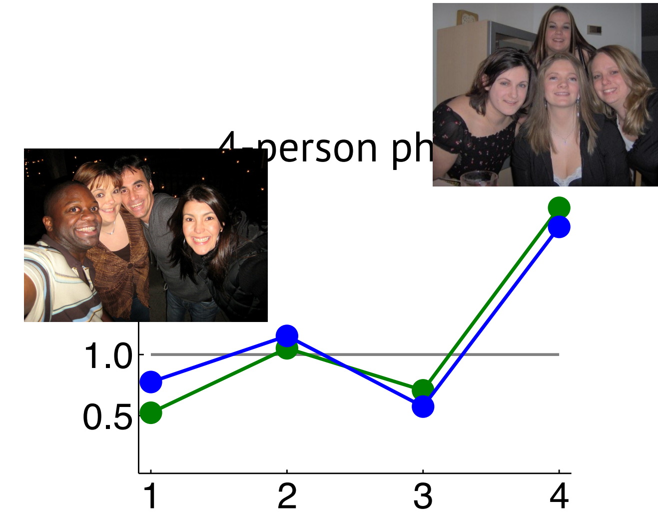
“wedding + bride + groom + portrait” query on Flickr
⟶ 662 images

Pairwise reduction
graph homophily

Male	0.57
Female	0.54







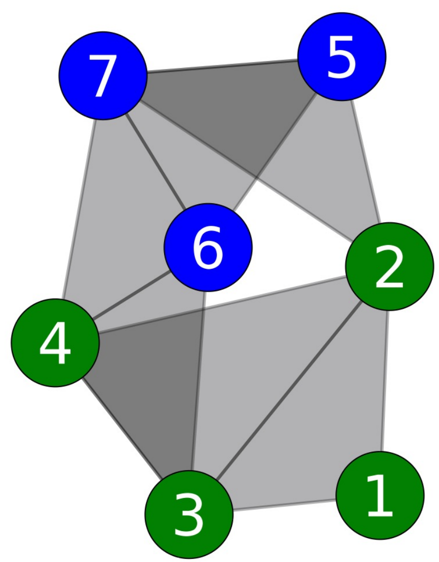
“group shot” or
“group photo” or
“group portrait”
query on Flickr
⟶ 963 images

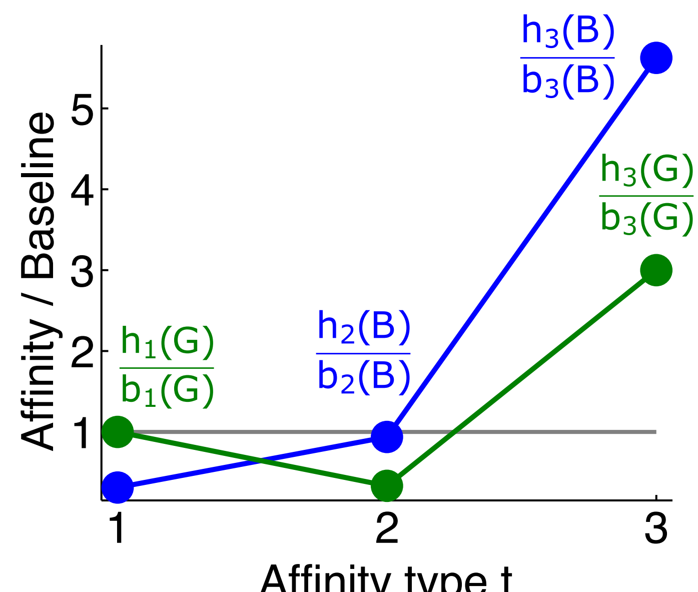
Pairwise reduction
graph homophily

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





THANKS! Austin Benson

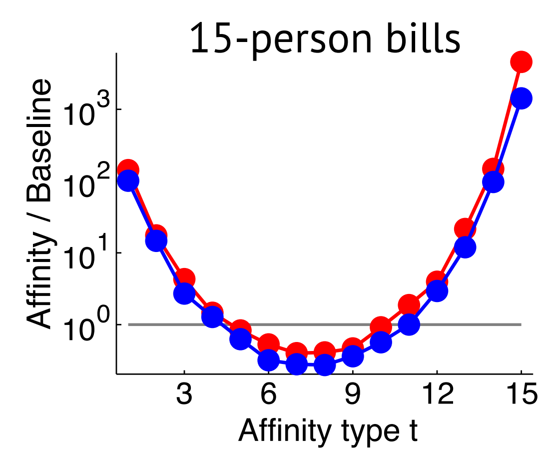
`http://cs.cornell.edu/~arb`

@austinbenson

arb@cs.cornell.edu

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.