Literature Review 1

21.04.2021

Naman Jain (IITGN)
Mentor: Dr Minsu Park (NYU AD)

Friendship paradox redux: Your Friends Are More Interesting Than You

Extended FP to **Virality** (better content) and **Activity** (longer online sessions)

Your friends' superior social connectivity puts them at a greater risk of bio-attack

More friends → More Information → Cognitive limit to appreciate

Twitter Dataset (20-30%, Jun to Dec, 2009)

Discuss about the information flow through virality in overloaded and underloaded users

Can we do a similar study on assessing flows in minority and majority ? Who gets more influenced based on activity and cascade sharing ?

How active a node in the graph is correlates to the information or influence on it

- We make contacts with people who are easiest to discover, who are more active

What influences overloaded users ?

- Any individual tweet's visibility is greatly diluted for overloaded users

Future Work: Using Activity paradox to model social networks over time

AAAI 2013

Friendship paradox biases perceptions in directed networks

Four variants of the paradox exist have not been comprehensively analyzed. 2 exists in all, 2 exists only if in/out degree correlate.

Correlation between nodes' trait and their out-degree amplifies this perception bias

Checking bias in Twitter through #'s for topics and traits

Polling Algo for global prevalence/actual value

Concept of adding attention to tweets is useful (attentive users help in amplifying the bias)

Analysis can be extended to weighted networks

Polling + Homophily

Future Work: How link recommendation can alleviate perception bias, allowing information to reach users based on attention and attribute

Nature Comm 2020

WHAT MAKES YOU THINK YOU'RE SO POPULAR?

Comparison between Tesser and Feld i.e. subjective perception (opinion) and objective perception (network structure)

Feld says people have less friends than their friends

Tesser says people tend to think they are more popular than their friends

Direction of Work: Seeks to understand how individuals develop mental representations of their positions in social network structure

Can we justify actions/transactions/links using mental representation + current structure of the network ?

ASA 2001

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

Homophily (that is, 'the company you keep') - $\mathbf{h_r}$ - perception bias

Monophily (that is, 'the company you're kept in') - $\mathbf{phi_r}$ such that $\mathbf{phi_r * h_r(1 - h_r)}$ define perception variability

They define the terms and run prediction algorithms on networks with homophily and monophily

Two hop methods works for networks without homophily

Direction of Work: While preference biases have been the focus of social structure in networks, there is a need to give serious parallel consideration to variability.

This work focuses on introducing monophily in network, can this be exploited with FP ?

But, first we need to understand how FP and homophily work together

What FP pressure when we control monophily/homophily?

Nature Human Behaviour
2018

References of FP Redux paper

Contagions in Social Networks

- Explore effects of FP on diffusion processes (can explore information diffusion ***[Gomez-Rodriguez, Gummadi, Schölkopf, 2014]***)
- Monophilic adoption rules (influenced by friends of friends) make it easier for a contagion to spread instead dying out as a result of FP. ***[Nettasinghe, Krishnamurthy and Lerman, 2018]***
- The disassortativity of the network amplifies the effect of the FP
- Diffusion between different communities (adult content consumption flow)***[Coletto, Silvestri 2016]***
- Can we model viral/sleeping beauties using friendship paradox ***[Gleeson 2016]***

Friendship Paradox based Sampling - Estimating Power-Law Degree Distribution via FP based Sampling ***[Nettasinghe and Krishnamurthy, 2021]***

Theoretical Proof for FP

- Strongest FP in negative assortativity and heterogeneous degree dist. ***[Cantwell, Kirkley, Newmann 2020]***
- More research remains to be done to clarify how attribute assortativity is related to within-node correlation, how these correlations affect behavior, and whether they are caused by something else ***[Kooti, Hodas, Lerman 2014]***

Sentiment Paradox [Zhou, Jin, Zafarani, ICWSM 2020]

- Friends are more positive than me
- **Direction:** self/inherent sentiment, homophily analysis, negative vs positive sentiment, spread of negative sentiment compared to positive sentiment
- Sentiment paradoxes in dynamic social networks as well as the “like’ and “comments” networks will be part of our future studies

Ideas for datasets

Music taste influence:

- Presence of FP in network of users and popularity of songs
- Perception of genres, artists and albums

Minority/Majority Networks with attribute:

- Classwise perception(homophily), latent perception (monophily/inherent trait vs allocated trait) and individual perception(FP)
- Eg, Twitter Communities (Political, Fan Following, Rivalries)

Signed Networks:

- Influence of same group is positive but of different group is negative
- Use FP to weigh the level of influence

Literature Review 2

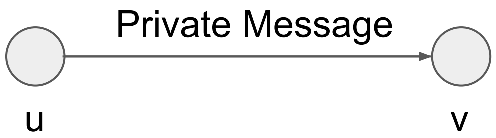
28.04.2021

Scenario A

Idea: Does Friendship paradox's biased perception influence their action?

Case: Will a user increase his/her activity if his/her friends are highly active?

Data: CollegeMsg temporal network (https://snap.stanford.edu/data/CollegeMsg.html)



- Keep a threshold for edges
- As soon as the threshold is reached over time, check the activity in future
- Does activity increase?
- Can we attach the quantity of activity with perception bias?
-

Scenario B

Idea: Who influences me more: more friends or more friend-of-friends?

Case: Which information do I share more? One which is shared more by my friends or one which is shared more by my friends-of-friends ?

Data: Twitter data (https://snap.stanford.edu/data/twitter7.html, http://an.kaist.ac.kr/traces/WWW2010.html)

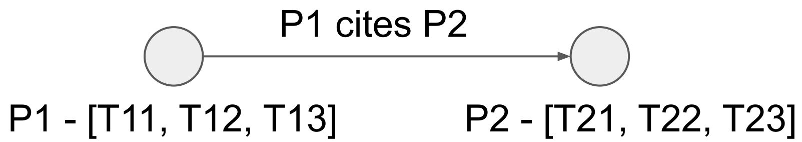
- Check for tweets shared by a user, count the number of times it was shared by his friends and count the number of times it was shared by his friends of friends

Scenario C

Idea: An individual with more friends (under information overload) needs more repeated exposures to spread further the information

Case: Examine information overload in citation networks

Data: MAG (https://www.microsoft.com/en-us/research/project/open-academic-graph/)



- Get topics using abstract of each paper (Fixed list of topics - https://github.com/lingo-iitgn/NLPExplorer)
- Analyse the flow of topics in citation networks
- Do information overload occurs? Does paper hold true?

Scenario D

Idea: Can friendship paradox's perceptions bias help in flagging fraudulent users or misleading information?

Case: Detect fraud users or fake news in social networks

Data: Twitter and Sina Weibo (http://people.duke.edu/~zg70/papers/GANG.pdf, https://www.nature.com/articles/s41598-020-58166-5)

- Check for friends and friends-of-friends of fraudulent users
- User profiling of fake news sharing people
- Who is more likely to share fake news

Competing for Attention in Social Media under Information Overload Conditions

Using Sina Weibo data

An individual with more friends needs more repeated exposures to spread further the information

Vary from epidemic spread model

Can we use friendship paradox to explain this phenomenon ?

Literature Review 3

5.05.2021

Stakeholders----> Me (ego), My friends (close ties), My friends of friends (weak ties)

Factors studied----> Degrees, Attribute, bias, pressure

Friendship Paradox
Feld 1991

User survey: Network vs User perception
Zuckerman and Jost 2001

Conditions for FP
Lerman 2017, Lattanzi 2015

Perception bias explanation
Jackson 2017

Comparing perception models
Lee 2019

Perception bias explanation in directed networks
Lerman 2020

Generalised Friendship Paradox
Eom and Jo 2014

Perception bias explanation
Lerman 2016

Social Media (Twitter)
Hodas 2014

Emotions and Sentiments
Bollen 2017, Zhou 2020

Centrality values
Higham 2018

Applications of Friendship Paradox

Contagion spread
Gomez-Rodriguez 2014

Privacy
Altenburger 2018

Information Diffusion
Nettasinghe 2019

Polling Algorithm
Lerman 2020, Nettasinghe 2021

Potential Gaps:

- Privacy
  Touched but not fully explored

- Community
  Focus on individuals only

- Mitigation
  How to mitigate bias? Why to mitigate the bias?

- Effects of perception bias

- Polling improvement
  Ask random friends, how many of their friends use that hashtag

- Population and Sampling gap
  Experiments are performed on sample, how can we extend the findings to population

- Prediction Framework
  Can we predict behaviours using FP?

More Scenarios...

Why certain group of people do what they do and others don't?

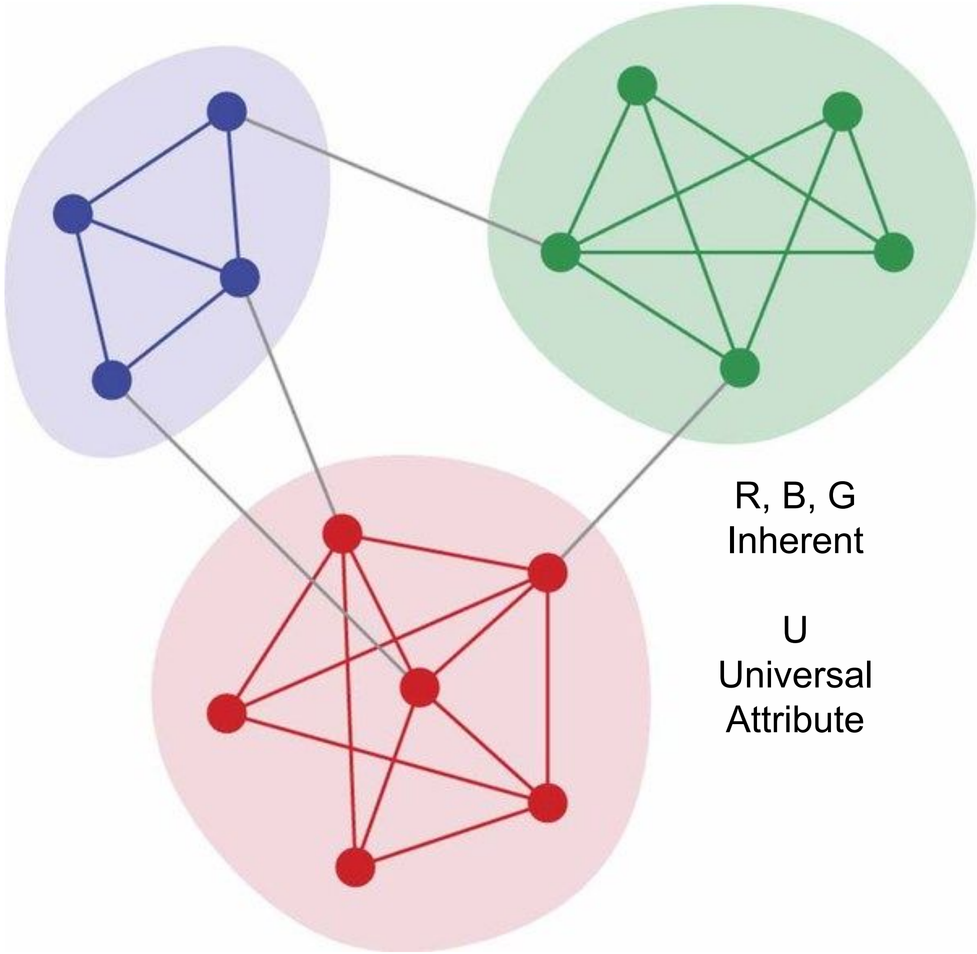
Do people who share fake news are friends with people who don't ?
Does fake news stop at people who have higher number of legitimate friends ?
Impact: Flag fake nodes

Information diffusion in pseudo-cliques vs non-clique/big network with weak links?
Impact: Guess virality

Mitigate perception bias in a social network like Facebook, Twitter
Impact: Reducing bias may lead to true information flow and better decision making

Friendship paradox in communities ??
Impact: Bias amongst communities about common attributes and unique attribute

Do people with most nodes influence the most or is there any other factor involved?
https://blogs.cornell.edu/info4220/2013/04/12/the-friendship-paradox/
Impact: Choosing important nodes for information diffusion



R, B, G
Inherent

U
Universal
Attribute

FP induced bias in communities

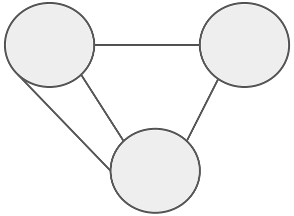
Formalise weighted/non-weighted perception bias

Prove existence of FP in communities

Give conditions for FP in communities

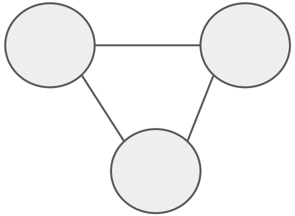
Perception of nodes which connects two communities

Case 1



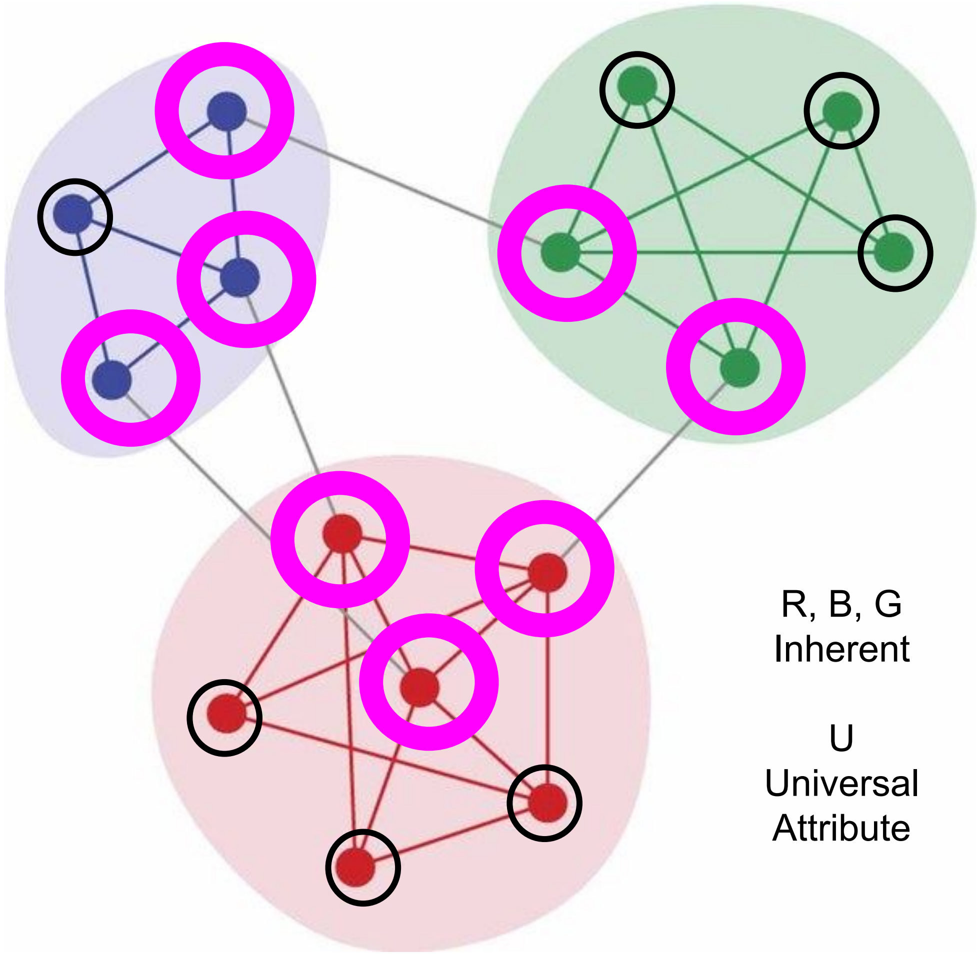
Weighted (Higham 2018)

Case 2



Unweighted

FP induced bias in communities



R, B, G
Inherent

U
Universal
Attribute

Degree analysis of individual nodes

Are pinks popular amongst there community?

Perception of pink and yellow nodes about U

Literature Review 4

12.05.2021

Exposure to opposing views on social media can increase political polarization

Initial Survey

Respondents were offered $11 to provide their Twitter ID and complete a 10-minute survey about their political attitudes, social media use, and media consumption habits (demographics provided by survey firm).

Randomization

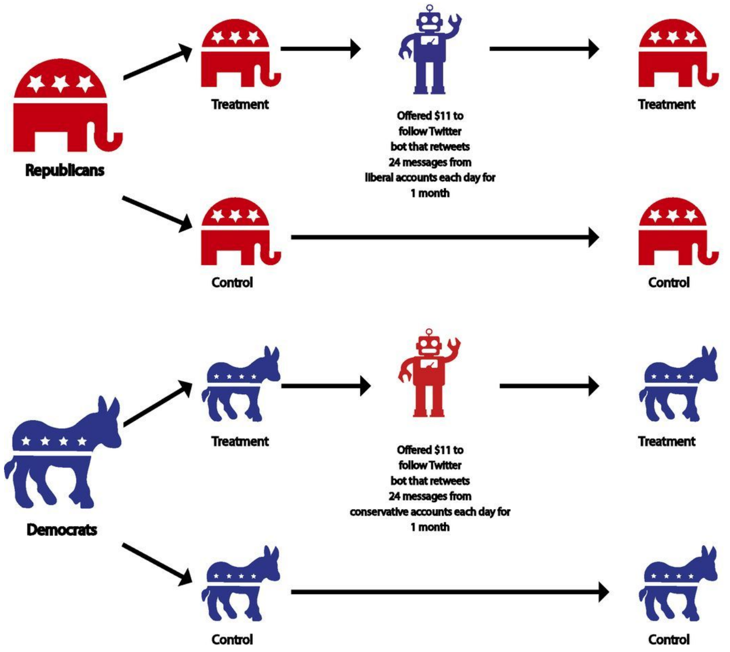
One week later, respondents were assigned to treatment and control conditions within strata created using pre-treatment covariates that describe attachment to party, frequency of Twitter use, and overall interest in current events.

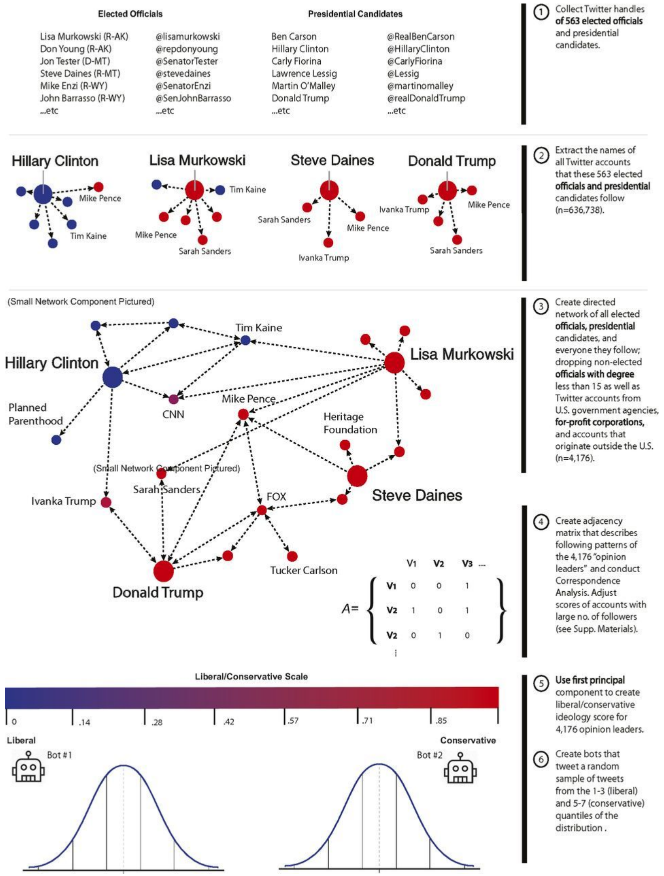
Weekly Surveys

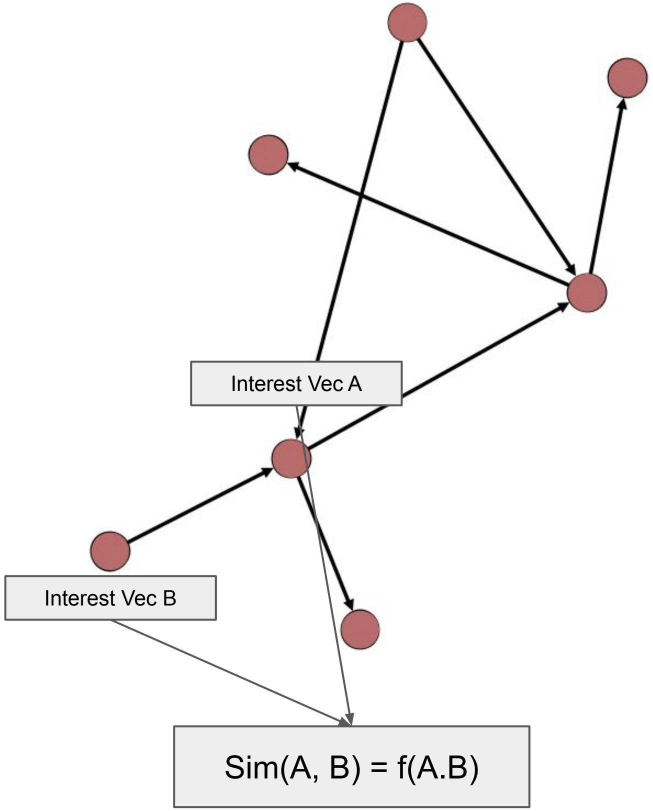
Respondents in treatment conditions informed they are eligible to receive up to $6 each week during the study period for correctly answering questions about the content of messages retweeted by Twitter .Bots.

Post-Survey

Respondents were offered $12 to repeat the pre-treatment survey one month after initial survey.





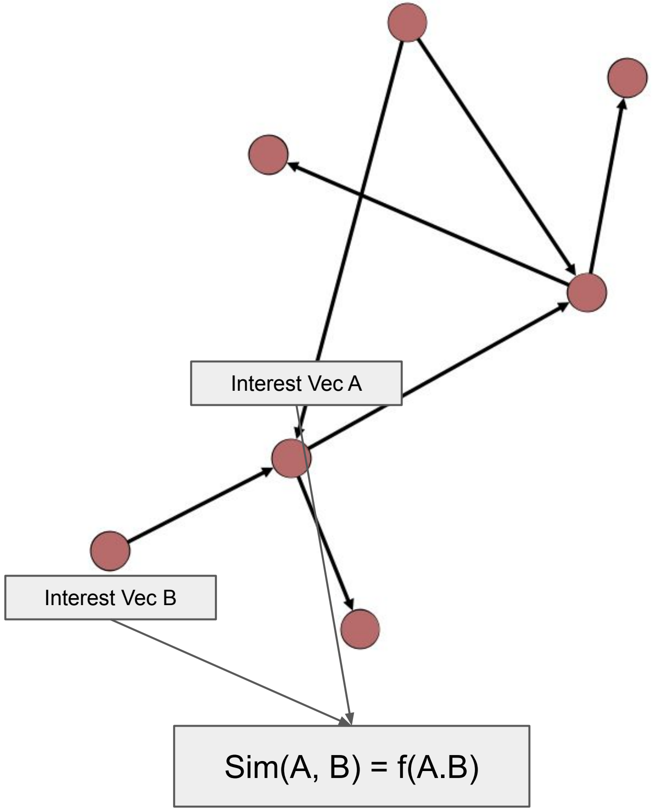


Information adoption and its effects

Adoption Speed
- Number of retweets in certain time
- Number of new topic’s retweets
- Increased retweets in existing topics

Adoption Probability
- Based on “Interest Vector”, calculate the probability to absorb information from friends
- Correlation between adoption speed and similarity (dot product of interest vectors)
- Opinion formation

+ve/-ve Adoption
- Using survey to gauge the polarity of the adoption
- Can we use NLP to get negative and positive adoption ?



Information adoption and its effects

Challenges

- What is Adoption?
    - Topic influence or Tweet frequency
    - New topic (in friends but not in you)
- Role of neighbours and FP
- Define the two sides of adoption for user's point of view
- Define ways to generate interest vector
    - Tweets or Ground truth community
- How can we isolate exchange of information between two nodes ?

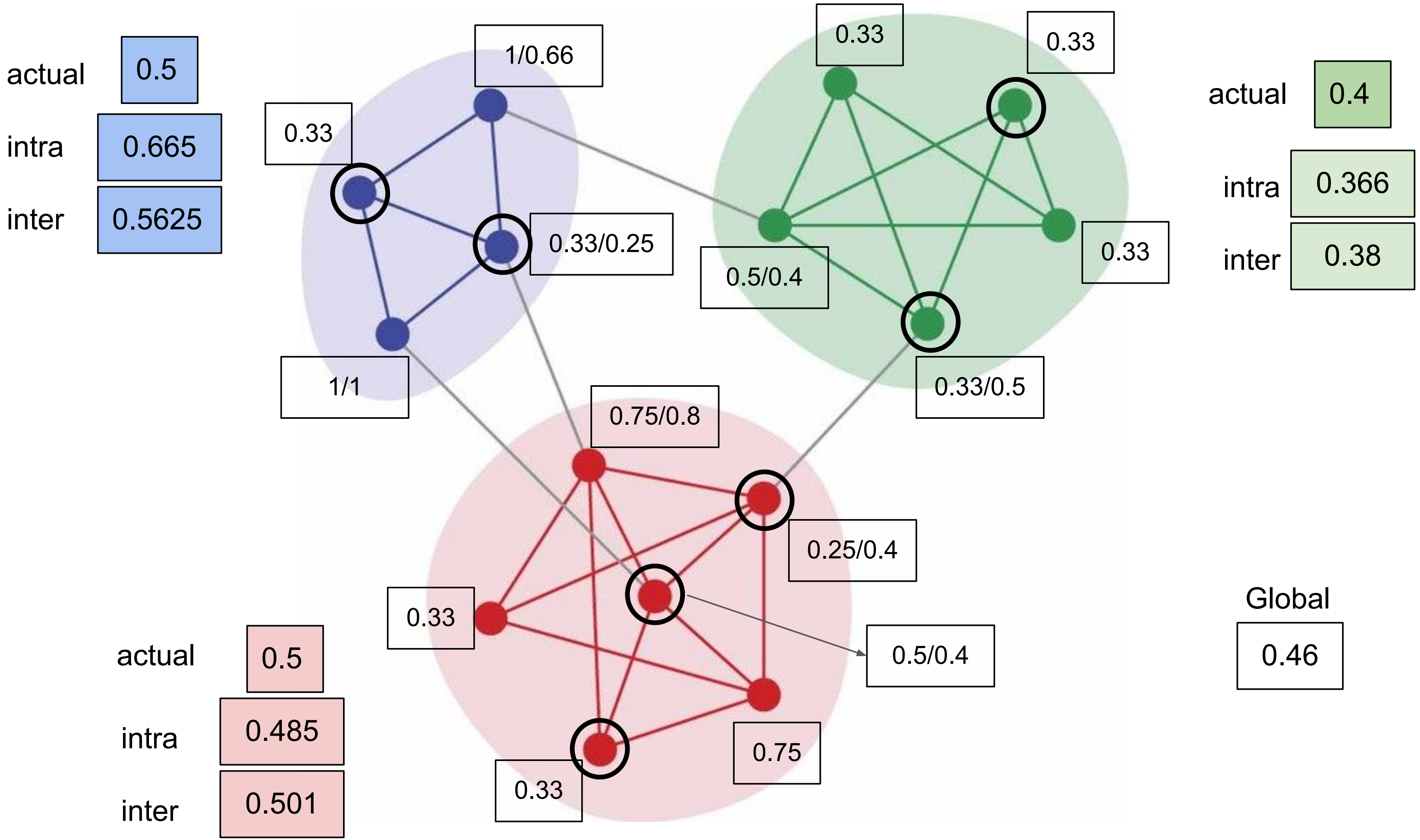
COMMUNITY EPIDEMIOLOGY LABORATORY

Assessing bias in community-based prevalence estimates: towards an unduplicated count of problem drinkers and drug users

General population survey estimates of the overall prevalence of problem drinking and drug use in a community are biased by the exclusion of non-household populations

In short:
Taking samples from some communities while ignoring the others
Underrepresentation of certain communities while overrepresentation of others

Why does that bias occurred ?
What is the perception of ignored community ?
What is the perception of the community which is taken into account ?



actual
0.5
intra
0.665
inter
0.5625
1/0.66
0.33
0.33/0.25
1/1
0.33
0.33
0.33
0.5/0.4
0.33/0.5
actual
0.4
intra
0.366
inter
0.38
0.75/0.8
0.25/0.4
0.33
0.5/0.4
0.33
0.75
actual
0.5
intra
0.485
inter
0.501
Global
0.46