Applications

Tutor: Lisette Espín-Noboa

THE WEB CONFERENCE ACM

Overview

Time: 15:30 - 16:30

15:30 - 16:05	Representative samples ● Sampling bias & properties preserved ○ On synthetic networks ○ On real networks
16:05 - 16:40	Ranking inequalities ● Inequality ● Inequity ● Disparity

Literature

Non-exhaustive list of material covered in this section.

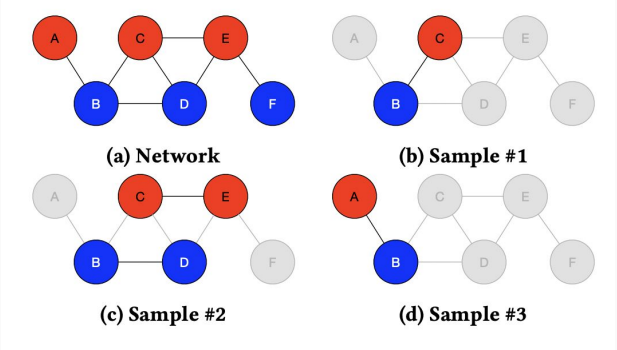
1. Espín-Noboa, L., Wagner, C., Strohmaier, M., & Karimi, F. (2022). Inequality and inequity in network-based ranking and recommendation algorithms. Scientific reports, 12(1), 1-14.
2. Karimi, F., Génois, M., Wagner, C., Singer, P. & Strohmaier, M. Homophily influences ranking of minorities in social networks. Sci. Rep.8 (2018).
3. Stoica, A.-A., Riederer, C. & Chaintreau, A. Algorithmic glass ceiling in social networks: The effects of social recommendations on network diversity. In Proceedings of the 2018 World Wide Web Conference, 923–932 (2018).
4. Fabbri, F., Bonchi, F., Boratto, L. & Castillo, C. The effect of homophily on disparate visibility of minorities in people recommender systems. In Proceedings of the International AAAI Conference on Web and Social Media 14, 165–175 (2020).

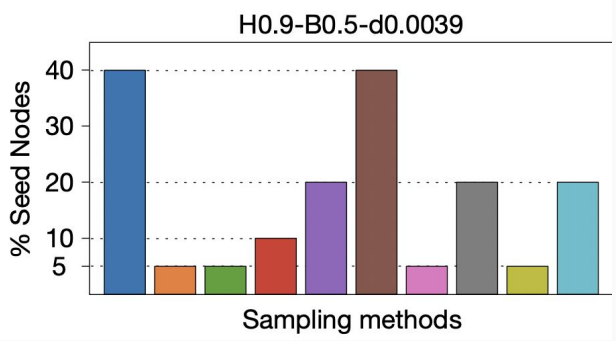
Biases in sampling

Biases in sampling

Real-world social network
(people with attributes)

(very big or proprietary)





Covered in this tutorial

Use case 1
Try multiple sampling methods and identify how representative that sample is compared to the whole network.

Use case 2
Identify the minimum sample size required per sampling method and type of network to achieve at most 20% error in classification.

Espín-Noboa, L., Wagner, C., Karimi, F., & Lerman, K. (2018, April). Towards quantifying sampling bias in network inference. In *Companion Proceedings of the The Web Conference 2018* (pp. 1277-1285).

Exercise

Open **3_exercise.ipynb**

Alternatively, you can open the notebook from Google Colab (you need a Google account): bit.ly/snma2023-notebooks

1. Create 3 `DPAH` graphs
   a. Make sure all of them have the same number of nodes `n`, edge density `d`, fraction of minority `f_m`, activities `plo_M` and `plo_m`, and random seed `seed`.
   b. Make sure they have the same level of homophily within the majority group (e.g., `h_MM=0.5`) and vary only the homophily within the minority group, for example:
      i. Graph 1: `h_MM=0.5` and `h_mm=0.1`
      ii. Graph 2: `h_MM=0.5` and `h_mm=0.5`
      iii. Graph 3: `h_MM=0.5` and `h_mm=0.9`
2. Make 6 random samples using the sampling techniques from `netin.sampling.*`
   a. Make sure they all have the same `pseeds` (sample size)
3. Analysis:
   a. Plot the graphs and the samples.
   b. Plot the representation of groups for each sample
   c. The CDF of the `in_degree` distribution and the CCDF of the `pagerank` distribution.
   d. Which sample looks closest to the full data? Does it depend on `h`?

BONUS exercise

Open **3_exercise.ipynb**

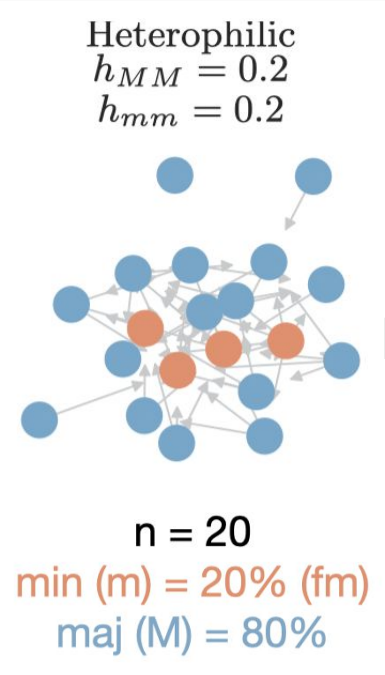
Alternatively, you can open the notebook from Google Colab (you need a Google account): bit.ly/snma2023-notebooks

(30 min)

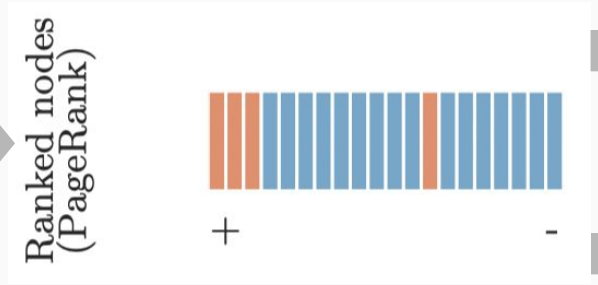
1. Load the `fb_friends` network, and make sure it is a `netin` graph.
2. Get to know the data (`.info()`)
3. Fit the `PA`, `PAH`, and the `PATC` models to the graph.
4. Visualize the graphs.
5. Compare their `degree` and `pagerank` distributions (plot the `pdf` and `cdf`).
6. Choose four sampling techniques and extract a sample for each network using a same sample size. What properties were preserved?

Ranking inequalities

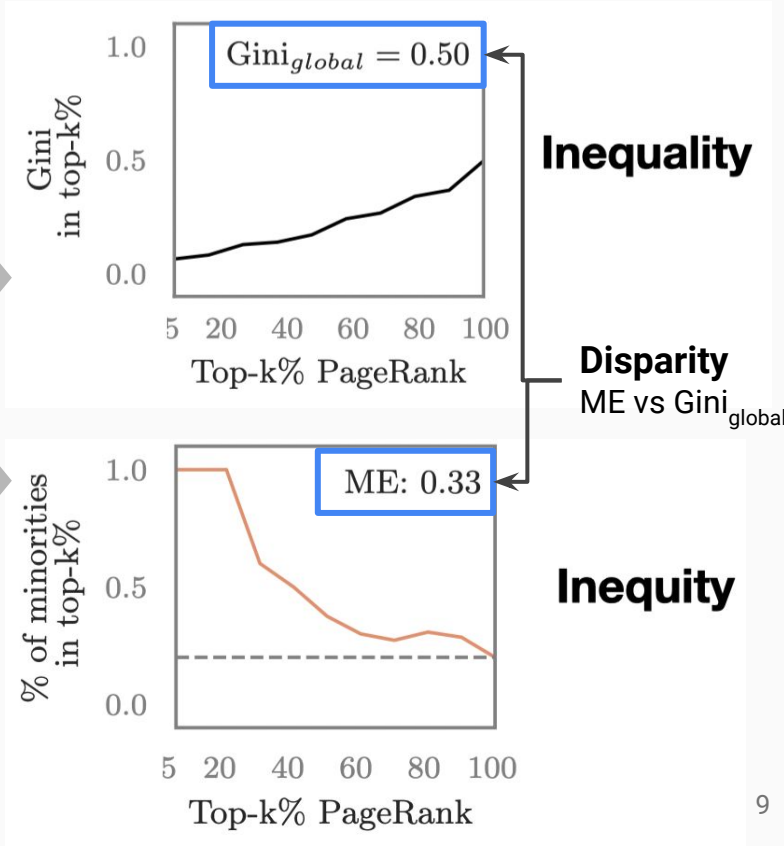
Given a network,



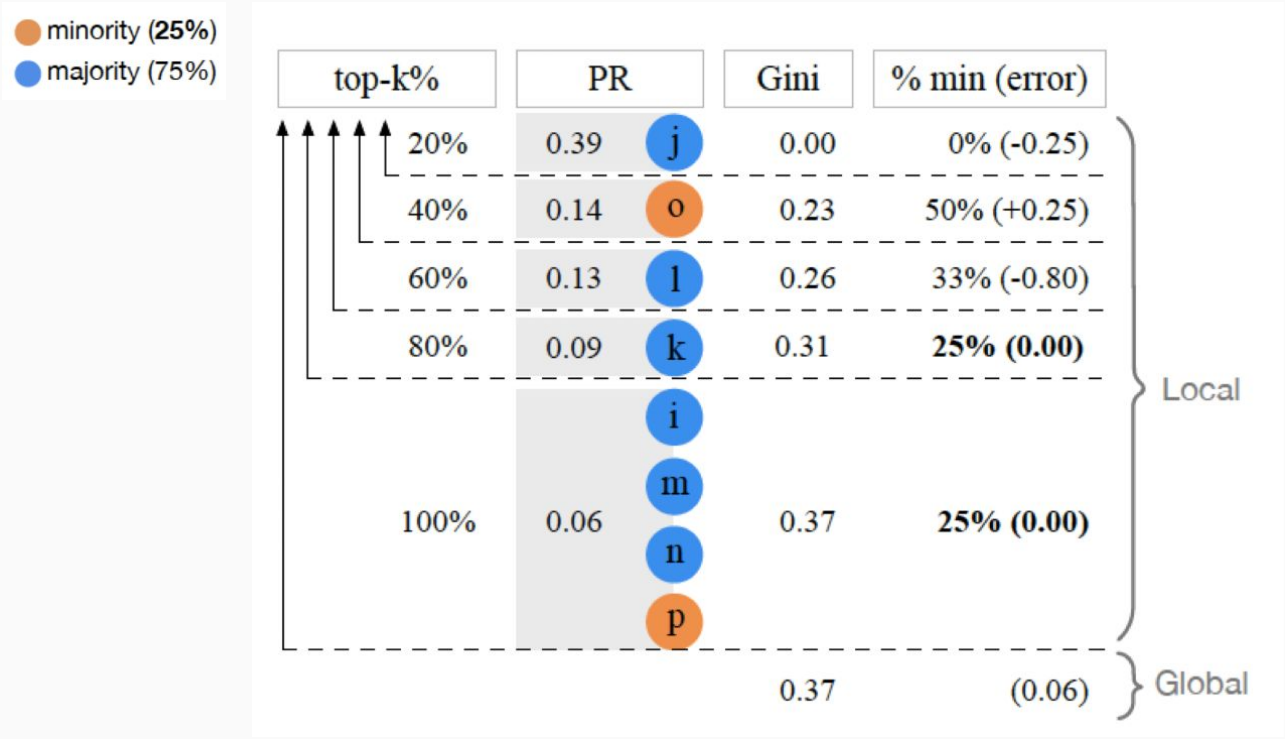
… and a ranking of its nodes



Measure:



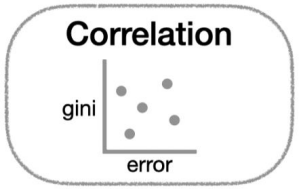
Espín-Noboa, L., Wagner, C., Strohmaier, M., & Karimi, F. (2022). Inequality and inequity in network-based ranking and recommendation algorithms. *Scientific reports*, *12*(1), 1-14.

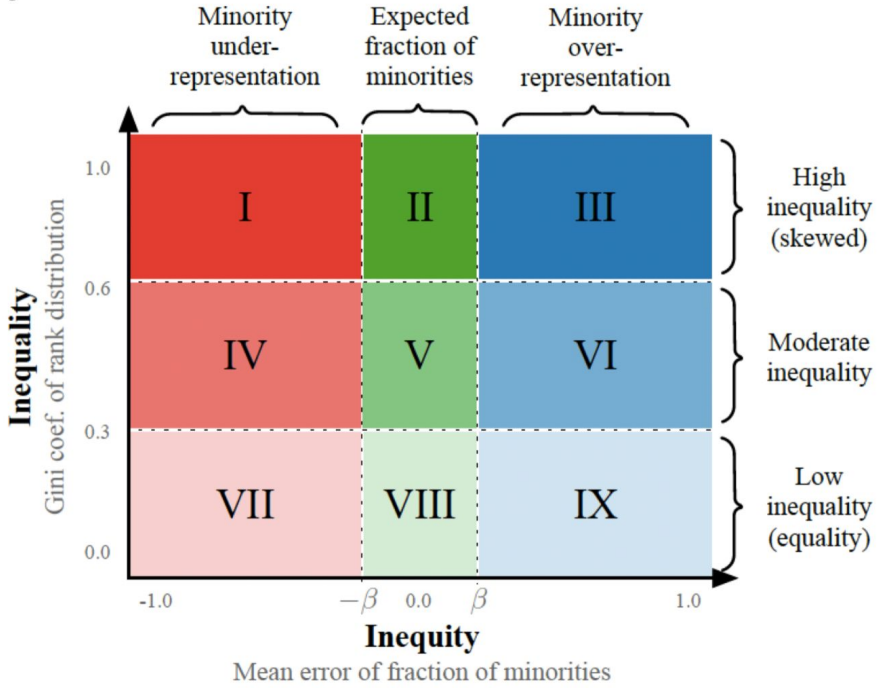


Espín-Noboa, L., Wagner, C., Strohmaier, M., & Karimi, F. (2022). Inequality and inequity in network-based ranking and recommendation algorithms. *Scientific reports*, *12*(1), 1-14.

Inequality vs. Inequity

Regions of disparity

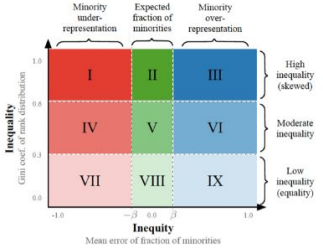




Espín-Noboa et al. 2022

Inequality vs. Inequity in PageRank

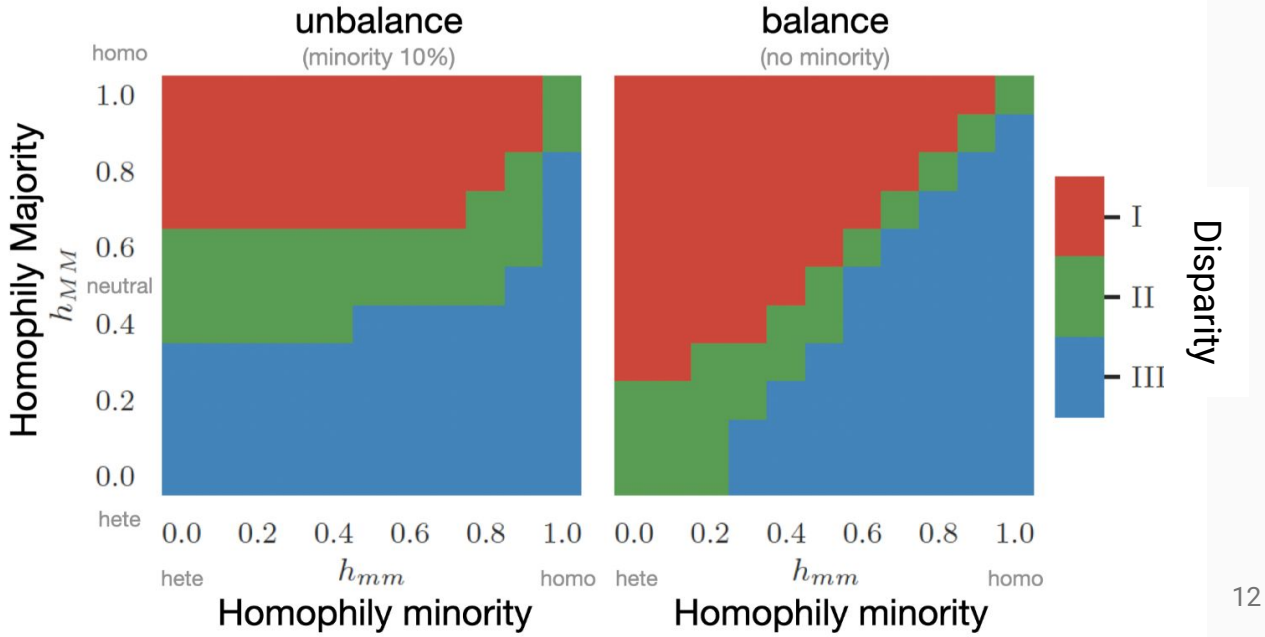
As a function of **H**omophily and **F**raction of **m**inorities



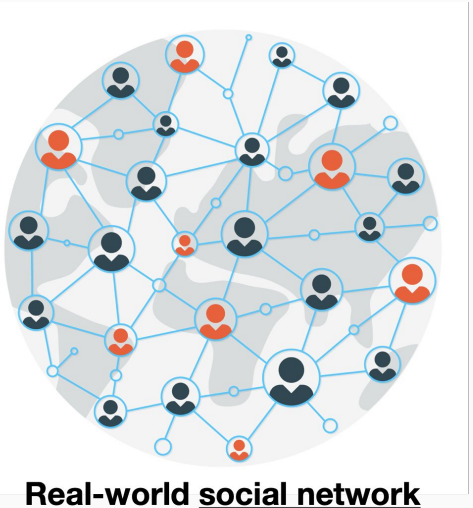
1. In balanced networks, both groups are well represented if hmm = hMM

2. In unbalanced networks, minorities are well represented when majority is neutral and the minority is not too homophilic.

3. In unbalanced and homophilic networks, minorities are well represented when hmm > hMM.



Espín-Noboa et al. 2022



Real-world social network
(people with attributes)

Ranking / RecSys.
(PageRank or WTF)

1. Identify network structure

Fraction min.	Node activity	Density
fm=0.3	γM = γm = 3	d=0.0015

Homophily Maj.	Homophily min.
H_{MM}=0.8	H_{mm}=0.4

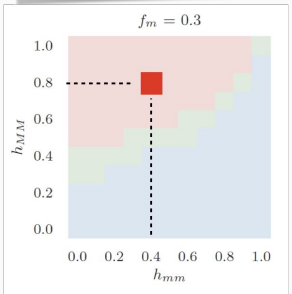
(Inequity is driven by homophily and fraction of minorities)

2. Identify inequality and inequity in ranking


OPEN

Inequality and inequity in network-based ranking and recommendation algorithms

Lisette Espín-Noboa[1,2,3], Claudia Wagner[1,4,5], Markus Strohmaier[1,4,6] & Fariba Karimi[1✉]

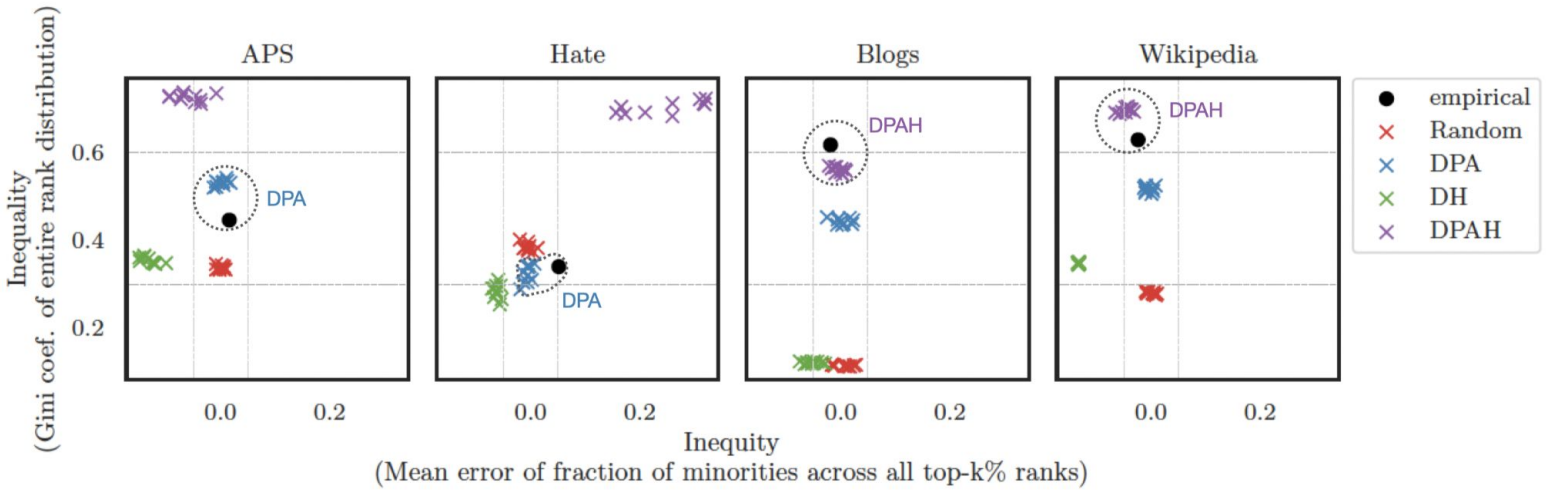


On average minorities are under-represented in top-k's
(Interventions needed)

Espín-Noboa, L., Wagner, C., Strohmaier, M., & Karimi, F. (2022). Inequality and inequity in network-based ranking and recommendation algorithms. *Scientific reports*, *12*(1), 1-14.

Empirical Networks

Model selection (best fit)



Espín-Noboa, L., Wagner, C., Strohmaier, M., & Karimi, F. (2022). Inequality and inequity in network-based ranking and recommendation algorithms. *Scientific reports*, *12*(1), 1-14.

Exercise

Open **4_exercise.ipynb**

Alternatively, you can open the notebook from Google Colab (you need a Google account): bit.ly/snma2023-notebooks

(30 min)

Comparing the effect of homophily in ranking by pagerank

1. Create 9 `DPAH` graphs
   a. Make sure all of them have the same number of nodes `n`, edge density `d`, fraction of minority `f_m`, activities `plo_M=plo_m`, and random seed `seed`.
   b. Make sure they have different values of homophily `h_MM` and `h_mm` as follows:
      i. Graphs 1-3: `h_MM=0.1` and `h_mm \in {0.1, 0.5, 0.9}`
      ii. Graphs 4-6: `h_MM=0.5` and `h_mm \in {0.1, 0.5, 0.9}`
      iii. Graphs 7-9: `h_MM=0.9` and `h_mm \in {0.1, 0.5, 0.9}`
2. Analysis:
   a. Plot the edge-type counts
   b. Plot the probability density function of their `pagerank` distributions
   c. Plot the `inequality` of the `pagerank`
   d. Plot the `inequity` of the `pagerank`
   e. Plot the `disparity` of the `pagerank`.

Bonus exercise

Open **4_exercise.ipynb**

Alternatively, you can open the notebook from Google Colab (you need a Google account): bit.ly/snma2023-notebooks

(30 min)

Comparing the effect of preferential attachment and homophily in ranking by pagerank

1. Generate 3 directed graphs; one for each model: `DPA`, `DH`, and `DPAH`
2. Make sure all of them have the same number of nodes `n`, edge density `d`, fraction of minority `f_m`, activities `plo_M=plo_m`, homophily `h_MM` and `h_mm`, and random seed `seed`, if applicable.
3. Plot the `disparity` of their `pagerank`.

Classification

(not covered today)

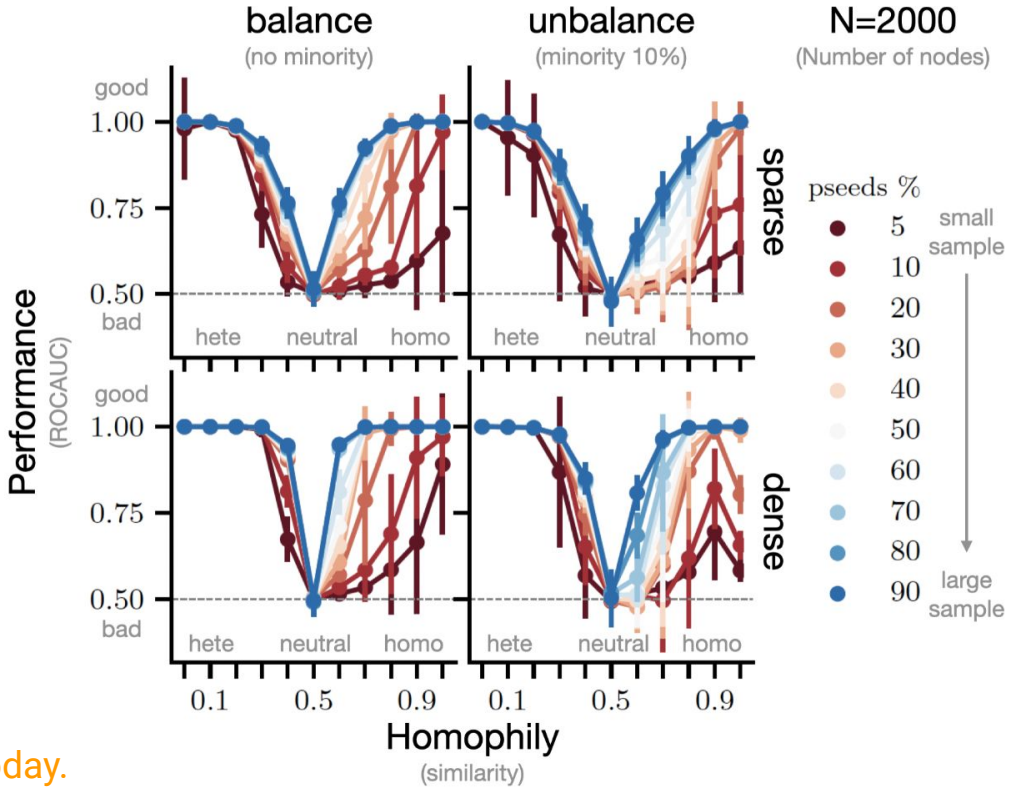
Network structure vs. Classification performance

1. **Neutral networks** (H=0.5) cannot be classified better than a random classifier.

2. **Homophilic networks** (H>0.5) achieve lower performance than heterophilic networks when samples are small.

3. **Denser networks** achieve higher performance compared to sparse networks.

4. **Network size** mainly affects ROCAUC variance. Larger networks produce more stable results. (not shown here)



Espín-Noboa et al. 2021

Not covered today.

Biases in relational classification

Real-world social network
(people with attributes)

What you get
(people with no attributes)

1. Identify network structure

Inference in OSNs via Lightweight Partial Crawls

Konstantin Avrachenkov
INRIA
Sophia Antipolis,France
k.avrachenkov@inria.fr

Bruno Ribeiro
Dept. of Computer Science
Purdue University
West Lafayette, IN, USA
ribeiro@cs.purdue.edu

Jithin K. Sreedharan
INRIA
Sophia Antipolis,France
jithin.sreedharan@inria.fr

Class balance
B=0.3

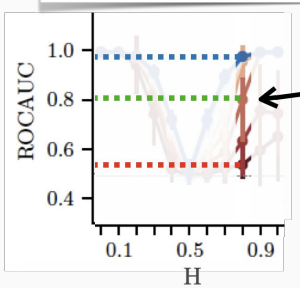
Homophily
H=0.8

2. Identify ROCAUC range for that network

RESEARCH

Explaining Classification Performance and Bias via Network Structure and Sampling Technique

Lisette Espin-Noboa, Fariba Karimi, Bruno Ribeiro, Kristina Lerman and Claudia Wagner

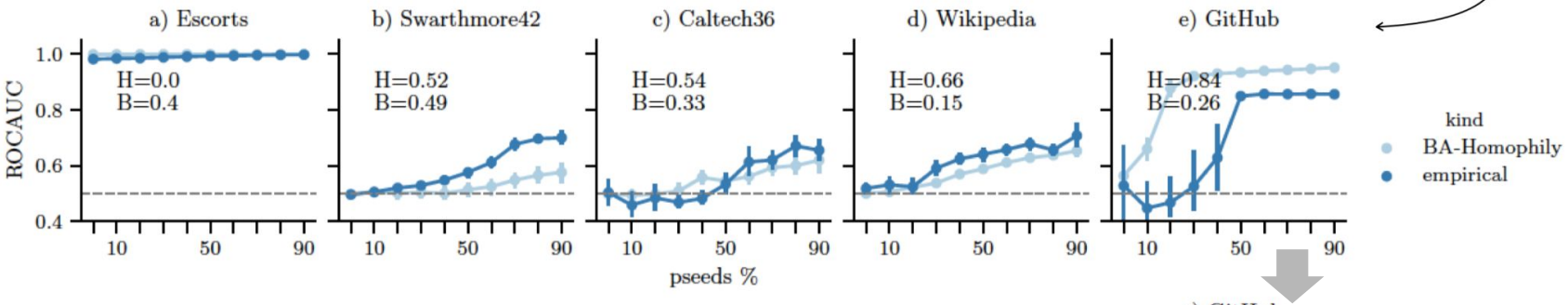


At least 20% of nodes in training sample to achieve ROCAUC > 0.8

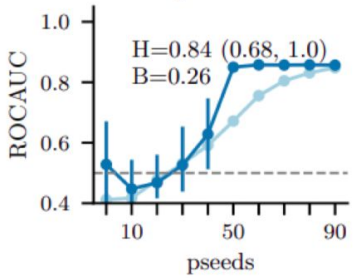
Real-world (empirical) networks

Model fitting

With symmetric homophily
$H_{mm} = H_{MM}$



A more realistic representation
With asymmetric homophily
$H_{mm} != H_{MM}$

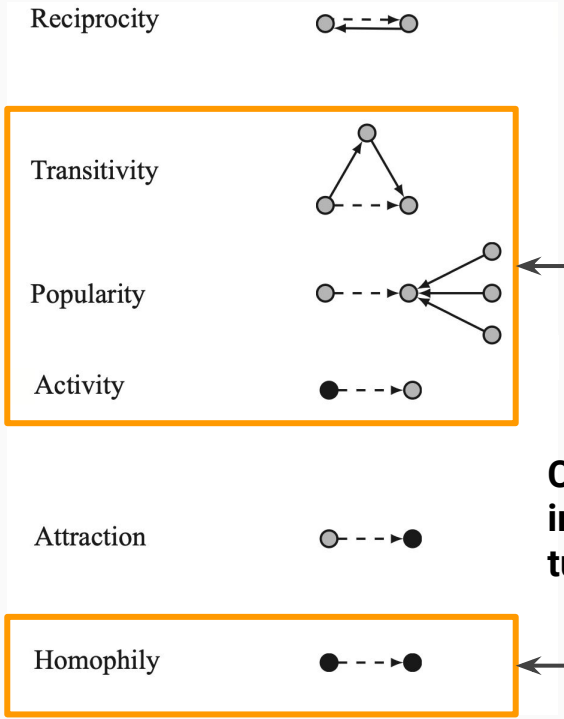


Closing remarks

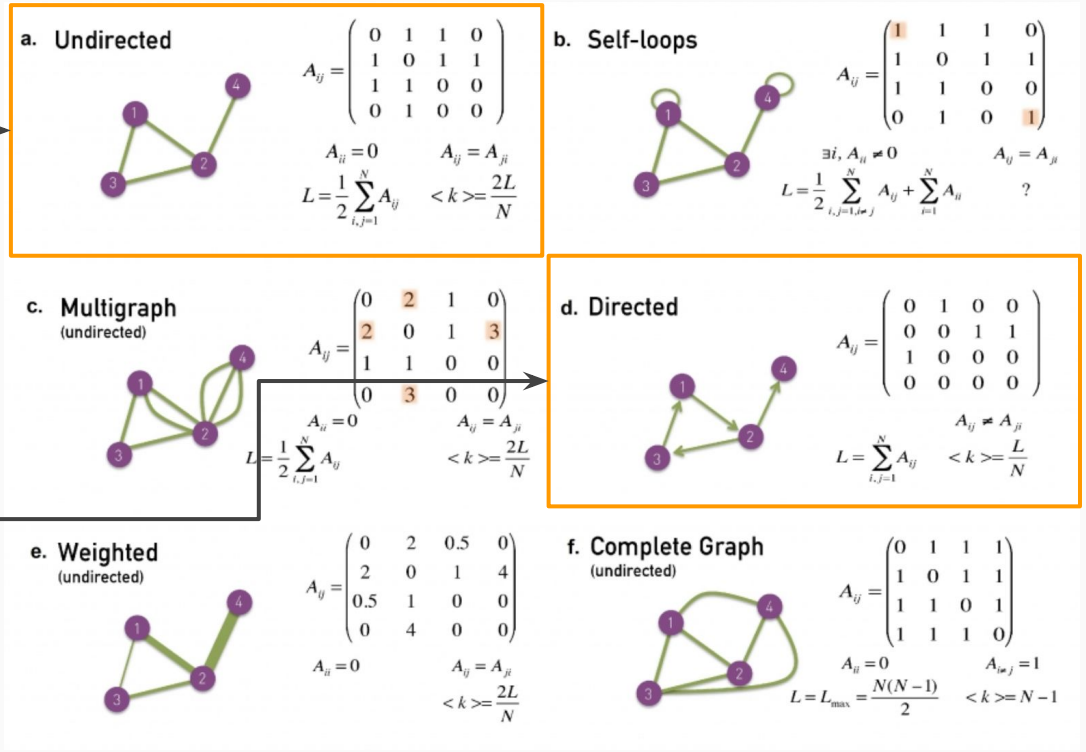
Challenges & open questions

Tutor: Lisette Espín-Noboa

We need more realistic models!



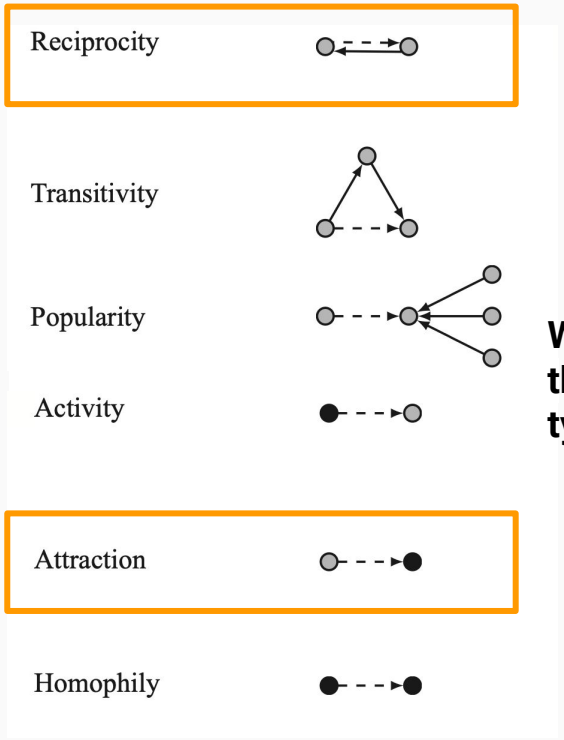
Covered in this tutorial



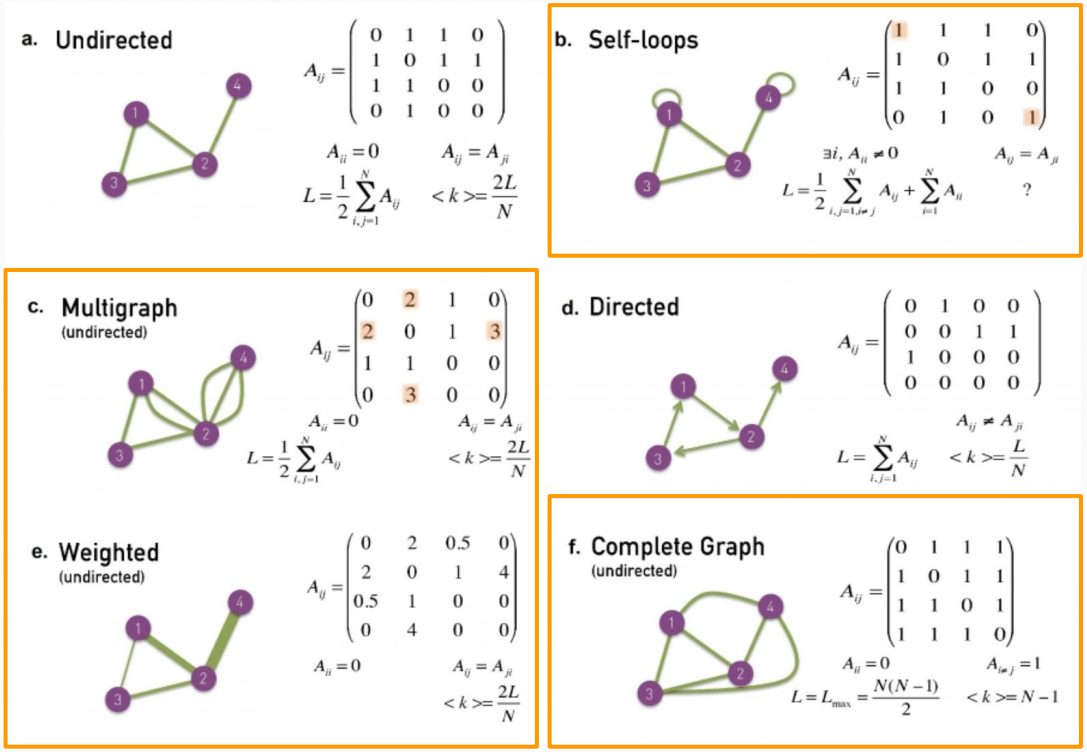
Stadfeld, Christoph, and Viviana Amati. "Network mechanisms and network models." Research Handbook on Analytical Sociology. Edward Elgar Publishing, 2021

http://networksciencebook.com/chapter/2#summary2

We need more realistic models!



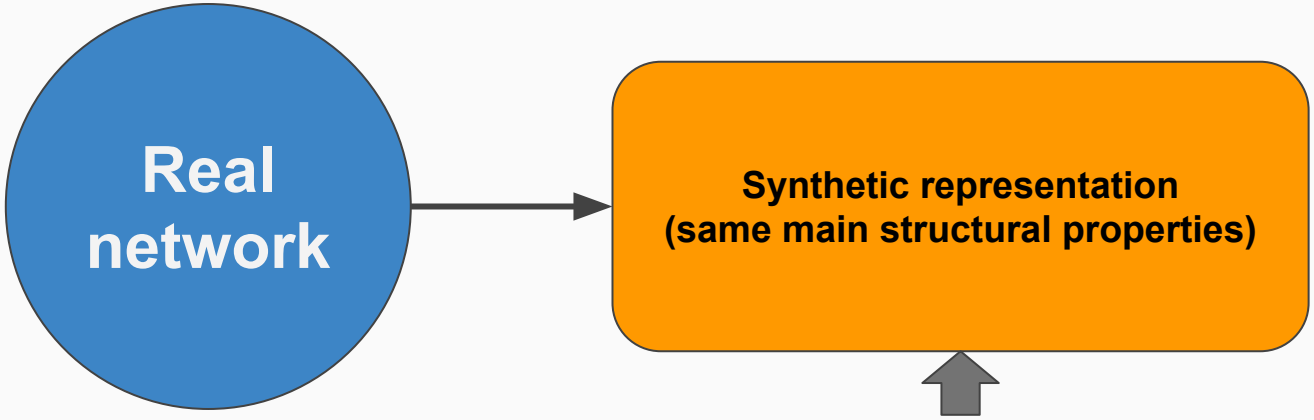
What about these other types



Stadfeld, Christoph, and Viviana Amati. "Network mechanisms and network models." Research Handbook on Analytical Sociology. Edward Elgar Publishing, 2021

http://networksciencebook.com/chapter/2#summary2

Do synthetic networks solve privacy issues for data sharing?



What are the privacy guarantees?

Can a particular (maybe rare) node attribute reveal the identity of an individual in the real network?

We appreciate your feedback.
Thank you very much!

bit.ly/snma2023-survey