TUTORIALS

Graph Neural Networks for Recommendation: Reproducibility, Graph Topology, and Node Representation

Part 2. Graph topology

30 minutes

The 2nd Learning on Graphs Conference (LoG 2023)

TABLE OF CONTENTS

USEFUL RESOURCES

The content of the following slides is taken from:

- Daniele Malitesta, Claudio Pomo, Vito Walter Anelli, Alberto Carlo Maria Mancino, Eugenio Di Sciascio, Tommaso Di Noia: **A Topology-aware Analysis of Graph Collaborative Filtering**. CoRR abs/2308.10778 (2023)

Scan me!

01

MOTIVATIONS

WHY GRAPH CF WORKS WELL?

Algorithmically

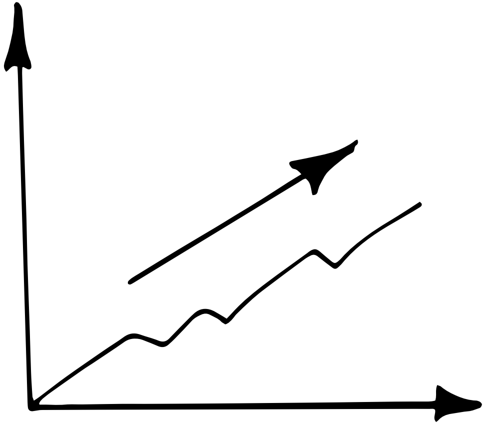
- **Established** approaches such as **LightGCN** and **DGCF** adapt the **GCN layer** to suit the **CF rationale**.
- **Recent** graph-based RSs including **UltraGCN** and **SVD-GCN** go **beyond** the concept of **message-passing**.

Topologically

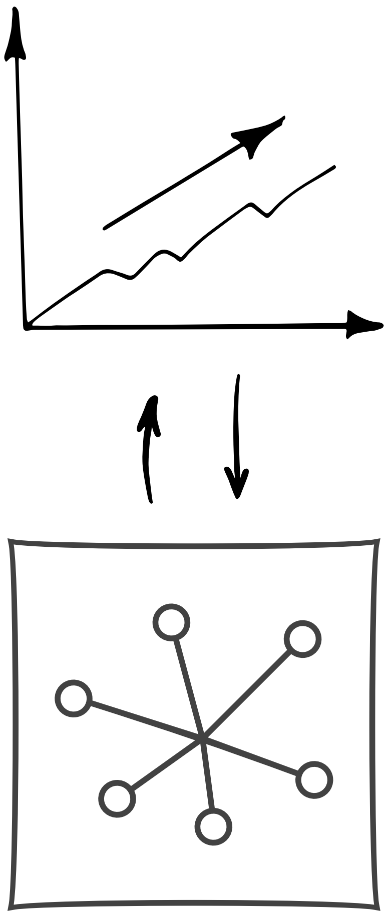
- The **machine learning** literature acknowledges that **graph topology** plays a **crucial role** in **GNNs [Wei et al., Castellana and Errica]**.
- For instance, revisit **network topology** to **enhance** the **learning ability** of **GCNs [Shi et al.]**.

Classical dataset **characteristics** such as **dataset sparsity** may impact the **performance** of **recommendation** models in diverse scenarios and tasks.

[Adomavicius and Zhang, Deldjoo et al.]



The **topological** nature of the **user-item data** and the new **graph CF wave** make it imperative to **unveil** the **dependencies** between **topological** properties and recommendation **performance**.



OUR CONTRIBUTIONS

1) First analysis in the literature

that studies the **impact** of **classical** and **topological** dataset characteristics on the performance of **SOTA graph CF**.

2) Select 4 graph CF approaches

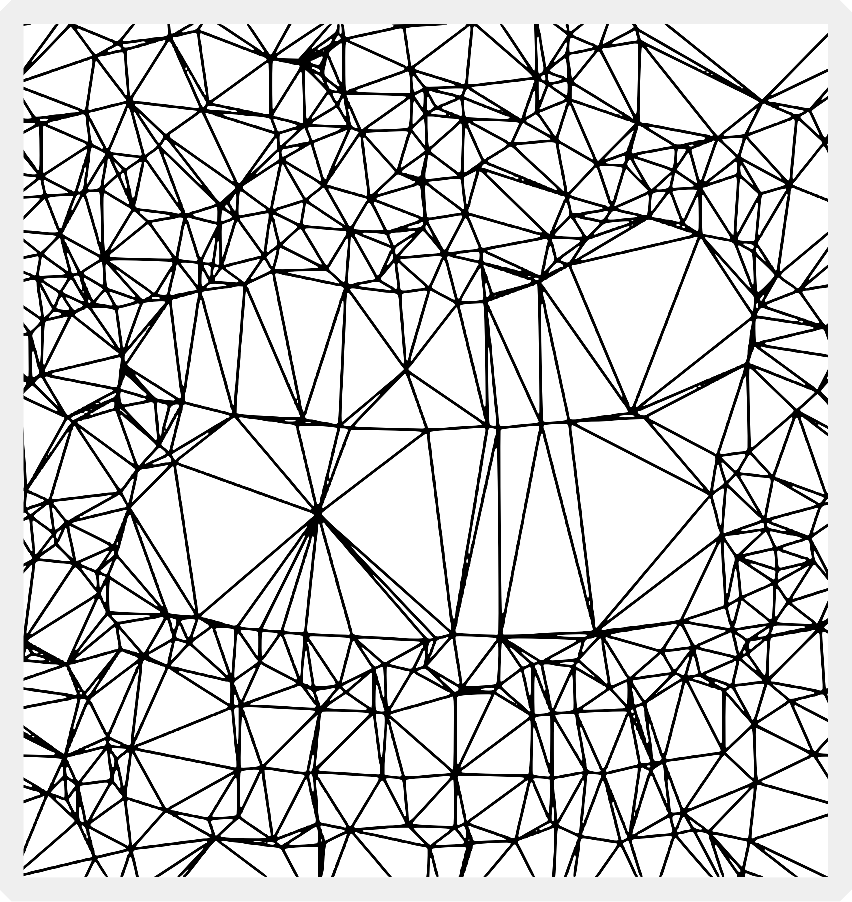
which are **recent** and **across-the-board** in the literature.

3) Build an explanatory model

to calculate the **linear dependences** among **characteristics** and recommendation **metrics**.

4) Validate the explanations

on their **statistical significance**, with **varying** settings of **graph samplings**.



02

TOPOLOGICAL CHARACTERISTICS

NODE DEGREE EXTENSION (1/3)

DEFINITION

Let $N_u^{(l)}$ and $N_i^{(l)}$ be the sets of neighborhood nodes for user u and item i at l distance hops. The extended definition of node degrees for u and i is:

$$\sigma_u = \left|N_u^{(1)}\right|, \qquad \sigma_i = \left|N_i^{(1)}\right|.$$

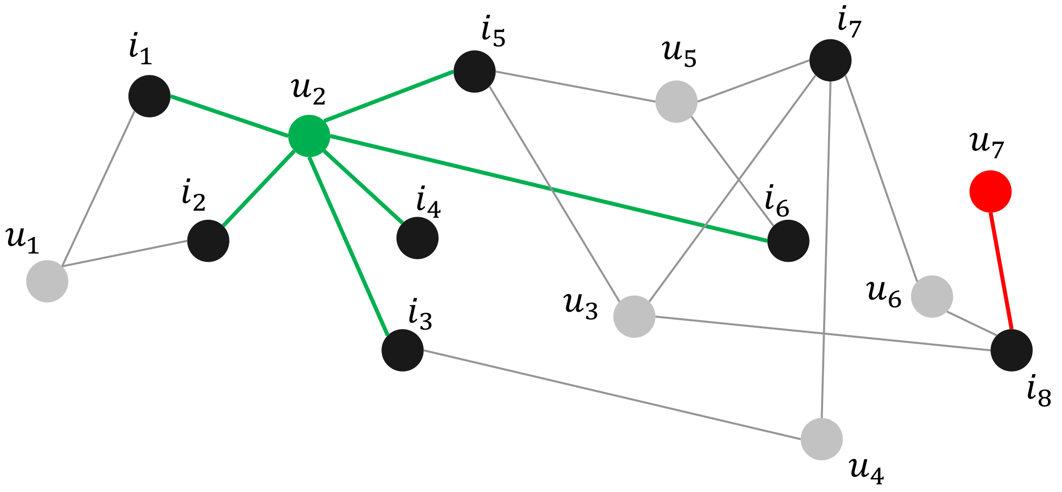
The average user and item node degrees on the whole user and item sets (U and I):

$$\sigma_U = \frac{1}{U}\sum_{u \in U}\left|N_u^{(1)}\right|, \qquad \sigma_I = \frac{1}{I}\sum_{i \in I}\left|N_i^{(1)}\right|.$$

NODE DEGREE EXTENSION (2/3)

RECSYS RE-INTERPRETATION

The **node degree** in the **user-item graph** stands for the **number of items (users) interacted** by a **user (item)**. This is related to the **cold-start issue** in recommendation, where **cold users** denote **low activity** on the platform, while **cold items** are **niche products**.

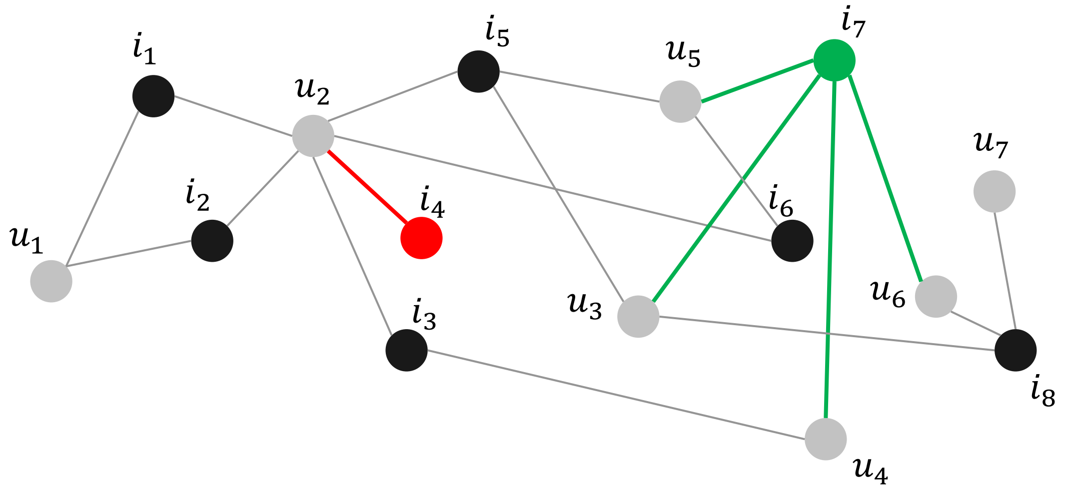


$\sigma_{u_2} = 5$ (active user)
$\sigma_{u_7} = 1$ (non-active user)

NODE DEGREE EXTENSION (3/3)

RECSYS RE-INTERPRETATION

The **node degree** in the **user-item graph** stands for the **number of items (users) interacted** by a **user (item)**. This is related to the **cold-start issue** in recommendation, where **cold users** denote **low activity** on the platform, while **cold items** are **niche products**.



$\sigma_{i_7} = 4$ (popular item)
$\sigma_{i_4} = 1$ (niche item)

CLUSTERING COEFFICIENT (1/5)

DEFINITION

Let v and w be two nodes from the same partition (e.g., user nodes). Their similarity is the intersection over union of their neighborhoods. By evaluating the metric node-wise:

$$\gamma_v = \frac{\sum_{w \in N_v^{(2)}} \gamma_{v,w}}{\left|N_v^{(2)}\right|} \qquad \text{with} \quad \gamma_{v,w} = \frac{\left|N_v^{(1)} \cap N_w^{(1)}\right|}{\left|N_v^{(1)} \cup N_w^{(1)}\right|},$$

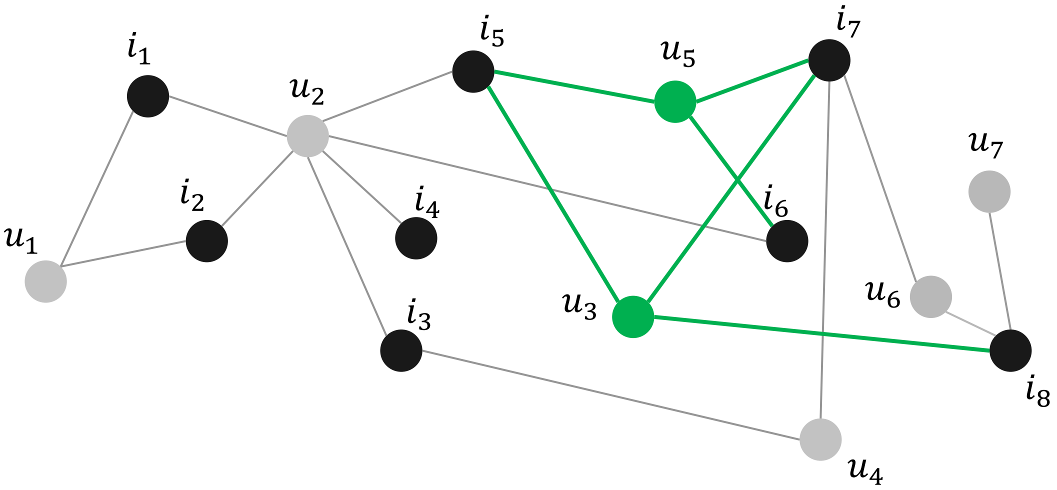
where $N_v^{(2)}$ is the second-order neighborhood set of v. The average clustering coefficient **[Latapy et al.]** on U and I is:

$$\gamma_U = \frac{1}{U} \sum_{u \in U} \gamma_u, \qquad \gamma_I = \frac{1}{I} \sum_{i \in I} \gamma_i.$$

CLUSTERING COEFFICIENT (2/5)

RECSYS RE-INTERPRETATION

High values of the **clustering coefficient** indicate that there exists a **substantial number** of **co-occurrences** among **nodes** of the **same partition**. For instance, **user-wise**, the value **increases** if **several users share** most of their **interacted items**. This intuition **aligns** with the rationale **of collaborative filtering**: two users are likely to **share similar preferences** if they interact with the **same items**.

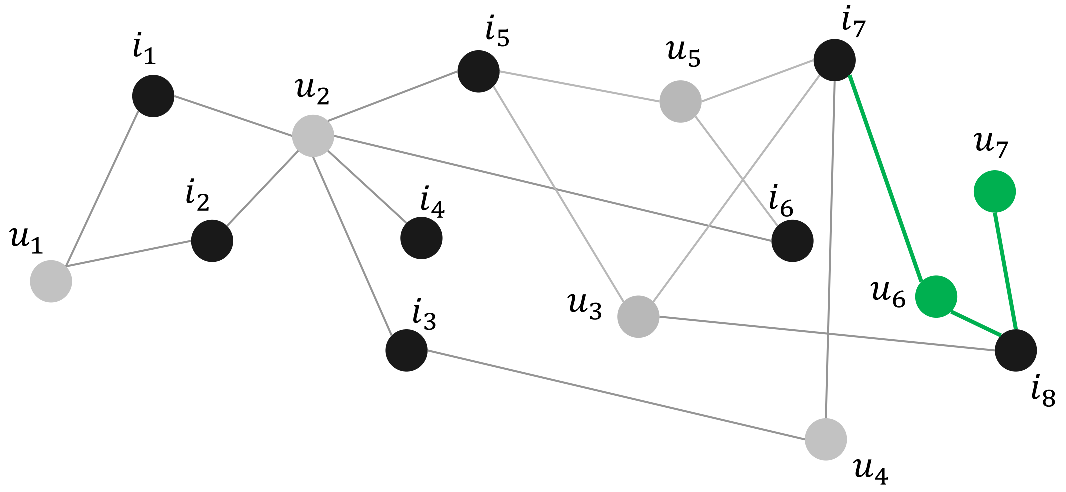


$$\gamma_{u_3,u_5} = \frac{|\{i_5,i_7\}\cap\{i_5,i_6,i_7\}|}{|\{i_5,i_6,i_7\}|} = \frac{2}{3} = 0.667$$

CLUSTERING COEFFICIENT (3/5)

RECSYS RE-INTERPRETATION

High values of the **clustering coefficient** indicate that there exists a **substantial number** of **co-occurrences** among **nodes** of the **same partition**. For instance, **user-wise**, the value **increases** if **several users share** most of their **interacted items**. This intuition **aligns** with the rationale **of collaborative filtering**: two users are likely to **share similar preferences** if they interact with the **same items**.

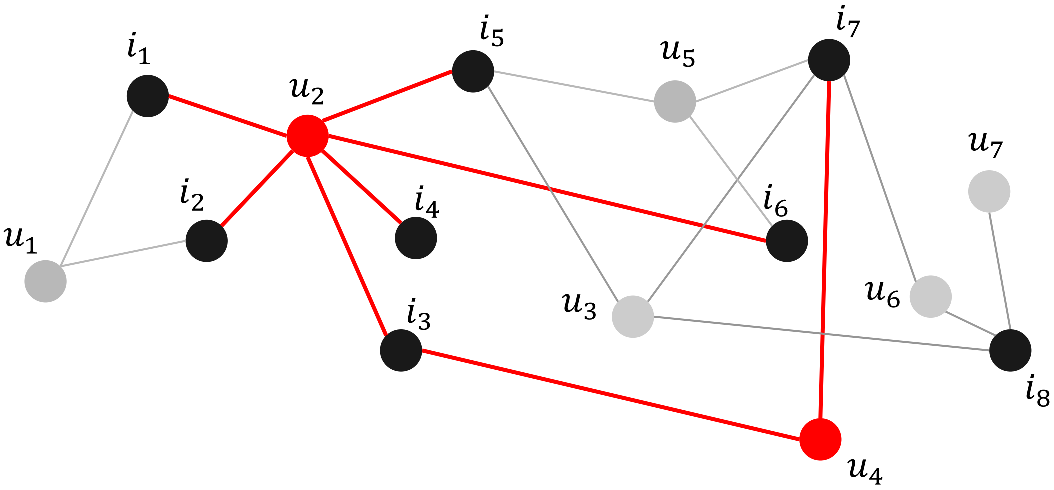


$$\gamma_{u_6,u_7} = \frac{|\{i_7, i_8\} \cap \{i_8\}|}{|\{i_7, i_8\}|} = \frac{1}{2} = 0.500$$

CLUSTERING COEFFICIENT (4/5)

RECSYS RE-INTERPRETATION

High values of the **clustering coefficient** indicate that there exists a **substantial number** of **co-occurrences** among **nodes** of the **same partition**. For instance, **user-wise**, the value **increases** if **several users share** most of their **interacted items**. This intuition **aligns** with the rationale **of collaborative filtering**: two users are likely to **share similar preferences** if they interact with the **same items**.

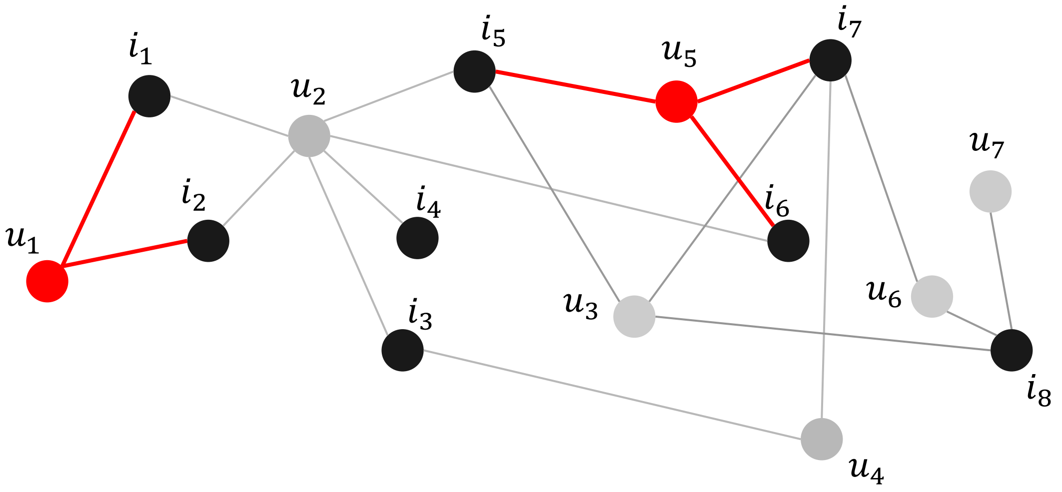


$$\gamma_{u_2,u_4} = \frac{|\{i_1,i_2,i_3,i_4,i_5\}\cap\{i_3,i_7\}|}{|\{i_1,i_2,i_3,i_4,i_5,i_7\}|} = \frac{1}{6} = 0.167$$

CLUSTERING COEFFICIENT (5/5)

RECSYS RE-INTERPRETATION

High values of the **clustering coefficient** indicate that there exists a **substantial number** of **co-occurrences** among **nodes** of the **same partition**. For instance, **user-wise**, the value **increases** if **several users share** most of their **interacted items**. This intuition **aligns** with the rationale **of collaborative filtering**: two users are likely to **share similar preferences** if they interact with the **same items**.



$$\gamma_{u_1,u_5} = \frac{|\{i_1, i_2\} \cap \{i_5, i_6, i_7\}|}{|\{i_1, i_2, i_5, i_6, i_7\}|} = \frac{0}{6} = 0.000$$

DEGREE ASSORTATIVITY (1/3)

DEFINITION

Let $D = \{d_1, d_2, \dots\}$ be the set of **unique node degrees** in the graph, and let e_{d_h,d_k} be the **fraction of edges** connecting **nodes with degrees** d_h and d_k. Then, let q_{d_h} be the **probability distribution** to choose a node with degree d_h **after** having **selected a node** with the **same degree** (i.e., the **excess degree** distribution). The **degree assortativity** coefficient **[M. E. J. Newman]** is:

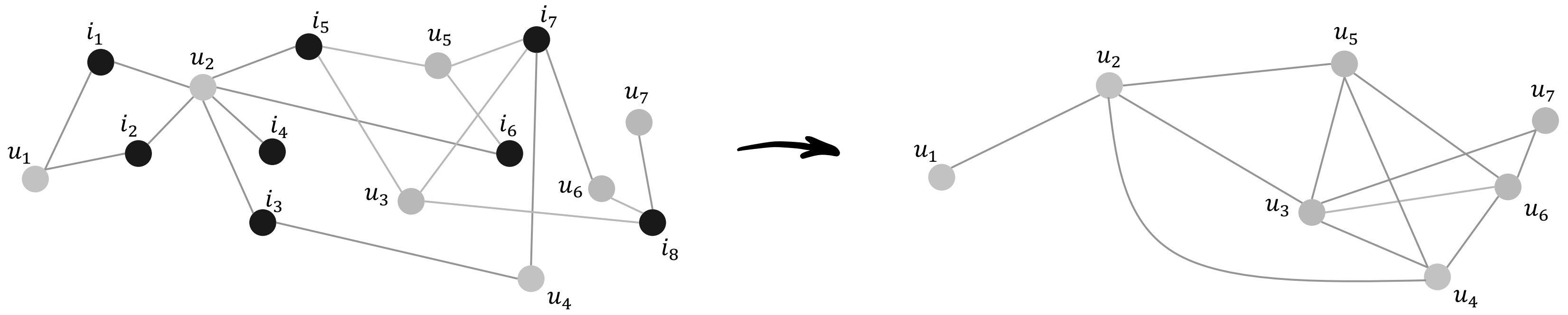
$$\rho = \frac{\sum_{d_h d_k} d_h d_k \left(e_{d_h,d_k} - q_{d_h} q_{d_k}\right)}{std_q^2},$$

where std_q is the standard deviation of the distribution q.

DEGREE ASSORTATIVITY (2/3)

RECSYS RE-INTERPRETATION

The **degree assortativity** calculated user- and item-wise is a **proxy** to represent the **tendency of users** with the **same activity** level on the platform and **items** with the **same popularity** to gather, respectively. To **visualize** the degree assortativity, we need the projected **user-user** and **item-item** graphs.



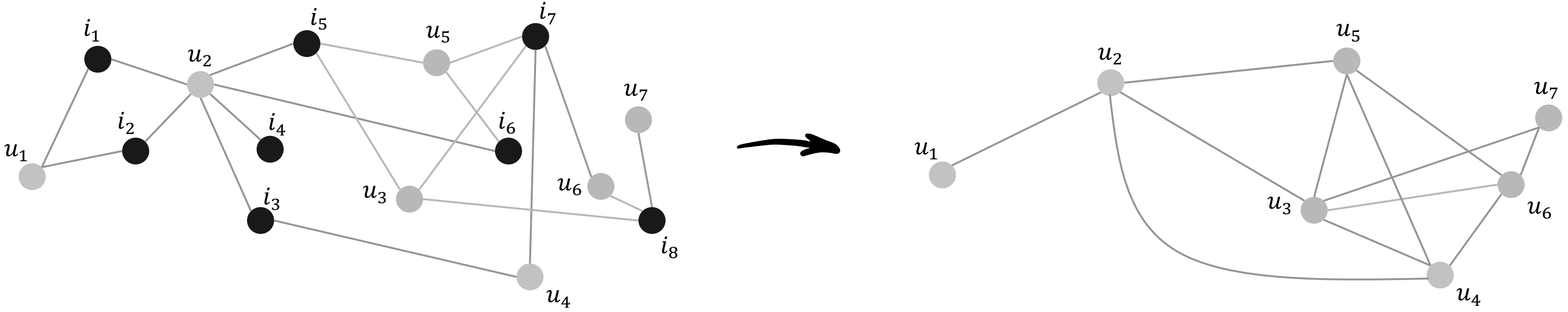
user-item graph

user-user projected graph

DEGREE ASSORTATIVITY (3/3)

RECSYS RE-INTERPRETATION

The **degree assortativity** calculated user- and item-wise is a **proxy** to represent the **tendency of users** with the **same activity** level on the platform and **items** with the **same popularity** to gather, respectively. To **visualize** the degree assortativity, we need the projected **user-user** and **item-item** graphs.



$\rho = -0.191$

(high degree dis-assortativity)

CLASSICAL DATASET CHARACTERISTICS

Space size

Estimates the number of all possible interactions:

$$\zeta = \sqrt{UI}.$$

Shape

Defines the ratio between the number of users and items:

$$\pi = \frac{U}{I}.$$

Density

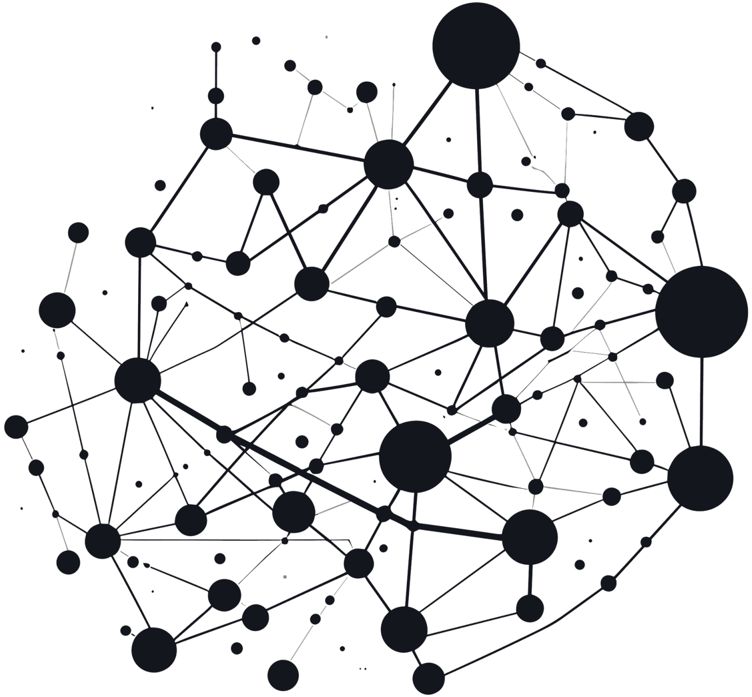
Measures the ratio of actual user-item interactions to all possible interactions:

$$\delta = \frac{E}{UI}.$$

Gini coefficient

Calculates the interactions' concentration for users (items):

$$\kappa_U = \frac{\sum_{u=1}^{U-1} \sum_{v=u+1}^{U} abs(\sigma_u - \sigma_v)}{U \sum_{u=1}^{U} \sigma_u}.$$



03

TOPOLOGY IN GRAPH CF

SELECTED GRAPH CF APPROACHES

LightGCN (SIGIR 2020)

Node degree used to **normalize** the **adjacency matrix** in the message-passing.

DGCF (SIGIR 2020)

Node degree used to **normalize** the **adjacency matrix** in the message-passing.

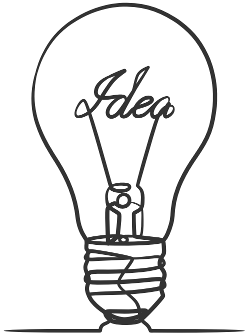
UltraGCN (CIKM 2021)

Node degree used for **normalization** in the **infinite** layer **message-passing**. The model also learns from the **item-projected graph**.

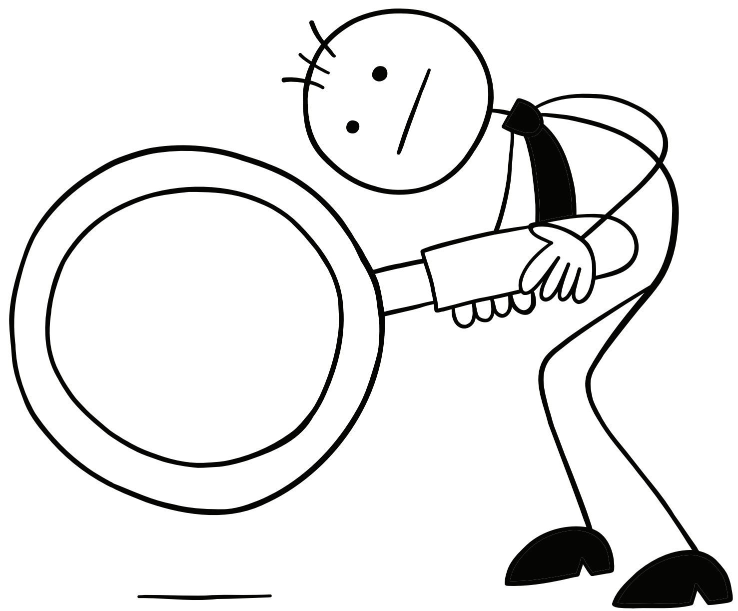
SVD-GCN (CIKM 2022)

Node embeddings involve the **largest singular values** of the **normalized** user-item interaction **matrix**, whose **maximum value** is **related** to the **maximum node degree** of the user-item graph. The model learns from **user-** and **item-projected graphs**.

The **selected graph CF** approaches **explicitly** utilize the **node degree** in the representation learning. However, **clustering coefficient** and **degree assortativity** do **not** have an **evident representation** in the formulations.



- Which **topological aspects** graph-based models can **(un)intentionally capture**?
- Are (topological) dataset **characteristics influencing** the recommendation **performance** of graph CF models?

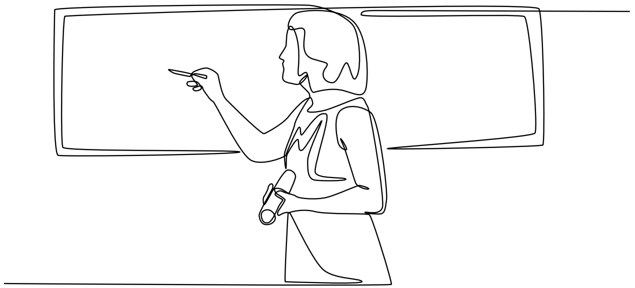


04

PROPOSED ANALYSIS

Our **goal** is to understand whether there **exist dependencies** among (topological) dataset **characteristics** and the **performance** of **graph-based** recommendation **models**.

To this aim, we decide to **build** and **fit** an **explanatory framework**.



EXPLANATORY FRAMEWORK: STEPS

EXPLANATORY FRAMEWORK: STEPS

STEP 1

Select models ✅

EXPLANATORY FRAMEWORK: STEPS

STEP 1	STEP 2
Select models ✅	Select characteristics ✅

EXPLANATORY FRAMEWORK: STEPS

STEP 1	STEP 2	STEP 3
Select models ✅	Select characteristics ✅	Select datasets

DATASETS

	# Users	# Items	# Interactions
Yelp-2018	25,677	25,815	696,865
Amazon -Book	70,679	24,915	846,434
Gowalla	29,858	40,981	1,027,370

EXPLANATORY FRAMEWORK: STEPS

STEP 1

Select models ✅

STEP 2

Select characteristics ✅

STEP 3

Select datasets ✅

EXPLANATORY FRAMEWORK: STEPS

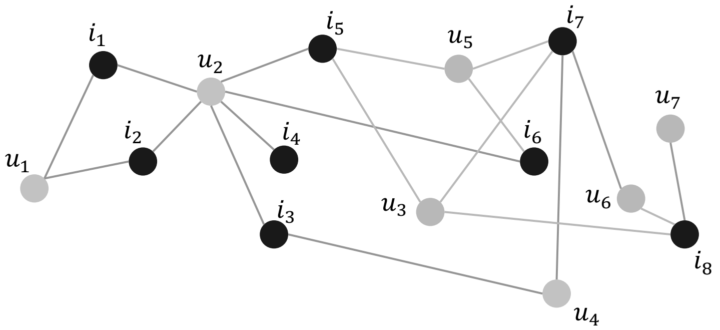
STEP 1	STEP 2	STEP 3	STEP 4
Select models ✅	Select characteristics ✅	Select datasets ✅	Generate sub-datasets

SUB-DATASETS GENERATION

Algorithm 2: Sub-dataset generation.

Input: Bipartite user-item graph $\mathcal{G}$, number of samples M.
Output: M sampled graphs.
$m \leftarrow 1$
$\mathcal{M} = \{\}$
while $m \leq M$ **do**
  $\mu \leftarrow uniform([0.7, 0.9])$
  $\boldsymbol{\mathtt{sampling}} \leftarrow uniform(\{\mathtt{nodeDropout}, \mathtt{edgeDropout}\})$
  $\mathcal{M} \leftarrow \mathcal{M} \cup sample(\mathcal{G}, \mu, \boldsymbol{\mathtt{sampling}})$
  $m \leftarrow m + 1$
end
Return $\mathcal{M}$.

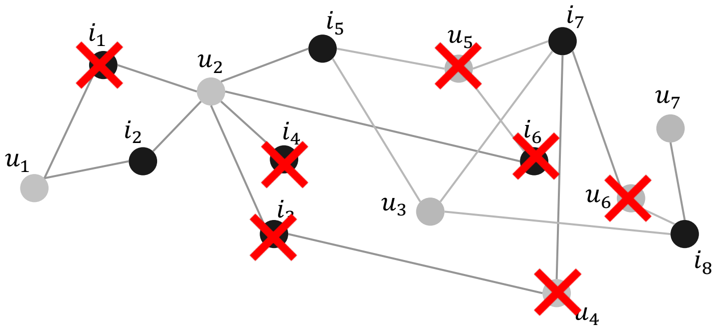
NODE DROPOUT



$$\text{users}\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

items

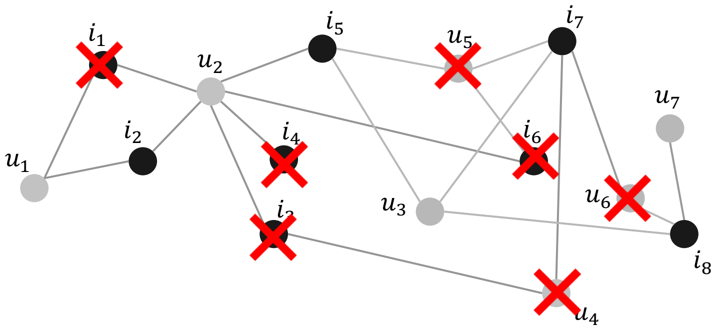
NODE DROPOUT

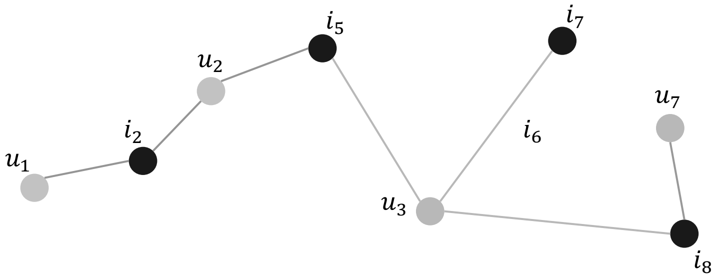


$$\text{users}\quad\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

items

NODE DROPOUT





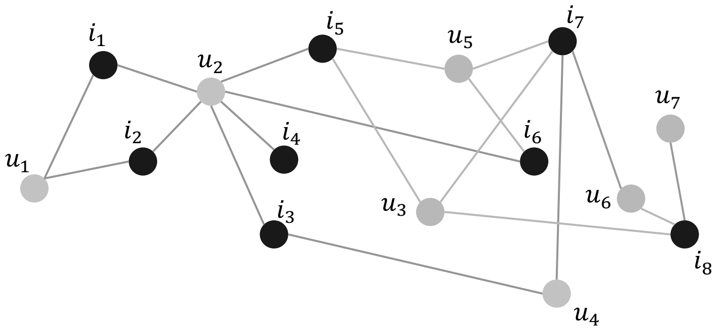
$$\text{users}\quad\begin{bmatrix}1&1&0&0&0&0&0&0\\1&1&1&1&1&1&0&0\\0&0&0&0&1&0&1&1\\0&0&1&0&0&0&1&0\\0&0&0&0&1&1&1&0\\0&0&0&0&0&0&1&1\\0&0&0&0&0&0&0&1\end{bmatrix}$$

items

$$\text{users}\quad\begin{bmatrix}1&0&0&0\\1&1&0&0\\0&1&1&1\\0&0&0&1\end{bmatrix}$$

items

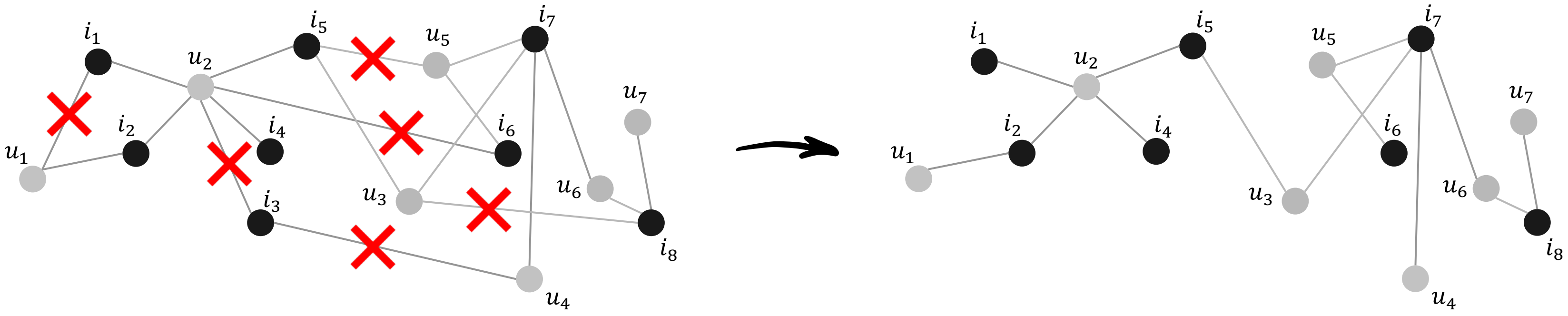
EDGE DROPOUT



$$\text{users}\ \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

items

EDGE DROPOUT



$$\begin{array}{cc} \text{users} & \begin{bmatrix} \boxed{1} & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & \boxed{1} & 1 & 1 & \boxed{1} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & \boxed{1} \\ 0 & 0 & \boxed{1} & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & \boxed{1} & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \\ & \text{items} \end{array}$$

$$\begin{array}{cc} \text{users} & \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \\ & \text{items} \end{array}$$

EXPLANATORY FRAMEWORK: STEPS

STEP 1	STEP 2	STEP 3	STEP 4
Select models ✅	Select characteristics ✅	Select datasets ✅	Generate sub-datasets ✅

EXPLANATORY FRAMEWORK: STEPS

STEP 1

Select models ✅

STEP 2

Select characteristics ✅

STEP 3

Select datasets ✅

STEP 4

Generate sub-datasets ✅

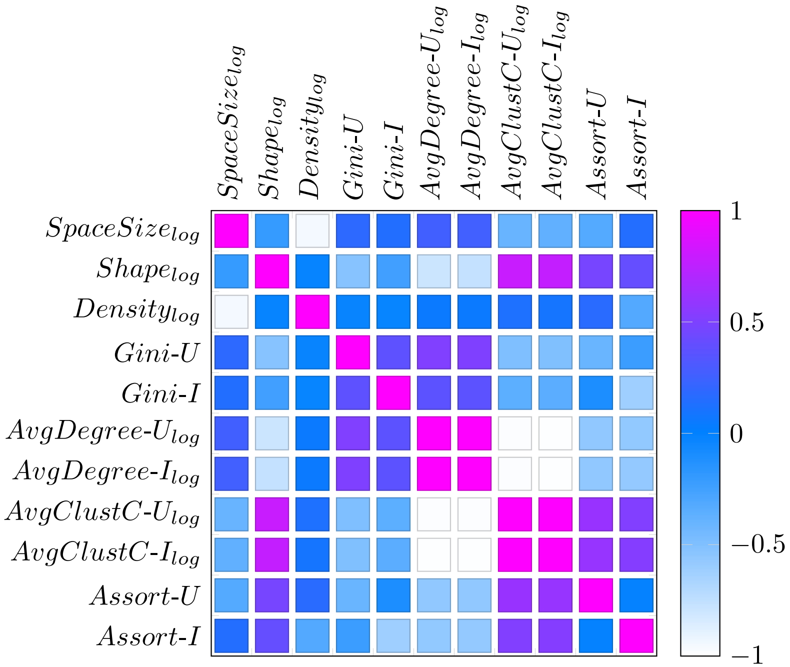
STEP 5

Get datasets' statistics

STATISTICS CALCULATION (1/2)

Type	Characteristics	Symbol	Log10	Shorthand
Classical	Space size	ζ	✓	$SpaceSize_{log}$
	Shape	π	✓	$Shape_{log}$
	Density	δ	✓	$Density_{log}$
	Gini user	$\kappa_{\mathcal{U}}$		$Gini\text{-}U$
	Gini item	$\kappa_{\mathcal{I}}$		$Gini\text{-}I$
Topological	Average degree user	$\sigma_{\mathcal{U}}$	✓	$AvgDegree\text{-}U_{log}$
	Average degree item	$\sigma_{\mathcal{I}}$	✓	$AvgDegree\text{-}I_{log}$
	Average clustering coefficient user	$\gamma_{\mathcal{U}}$	✓	$AvgClustC\text{-}U_{log}$
	Average clustering coefficient item	$\gamma_{\mathcal{I}}$	✓	$AvgClustC\text{-}I_{log}$
	Degree assortativity user	$\rho_{\mathcal{U}}$		$Assort\text{-}U$
	Degree assortativity item	$\rho_{\mathcal{I}}$		$Assort\text{-}I$

STATISTICS CALCULATION (2/2)



EXPLANATORY FRAMEWORK: STEPS

STEP 1

Select models ✅

STEP 2

Select characteristics ✅

STEP 3

Select datasets ✅

STEP 4

Generate sub-datasets ✅

STEP 5

Get datasets' statistics ✅

EXPLANATORY FRAMEWORK: STEPS

STEP 1
Select models ✅

STEP 2
Select characteristics ✅

STEP 3
Select datasets ✅

STEP 4
Generate sub-datasets ✅

STEP 5
Get datasets' statistics ✅

STEP 6
Train models ✅

EXPLANATORY FRAMEWORK: STEPS

STEP 1

Select models ✅

STEP 2

Select characteristics ✅

STEP 3

Select datasets ✅

STEP 4

Generate sub-datasets ✅

STEP 5

Get datasets' statistics ✅

STEP 6

Train models ✅

STEP 7

Evaluate models ✅

EXPLANATORY FRAMEWORK: STEPS

STEP 1

Select models ✅

STEP 2

Select characteristics ✅

STEP 3

Select datasets ✅

STEP 4

Generate sub-datasets ✅

STEP 5

Get datasets' statistics ✅

STEP 6

Train models ✅

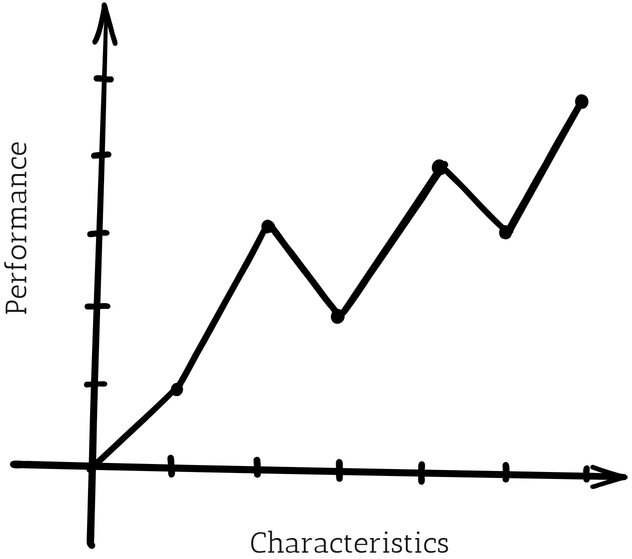
STEP 7

Evaluate models ✅

STEP 8

Generate explanations

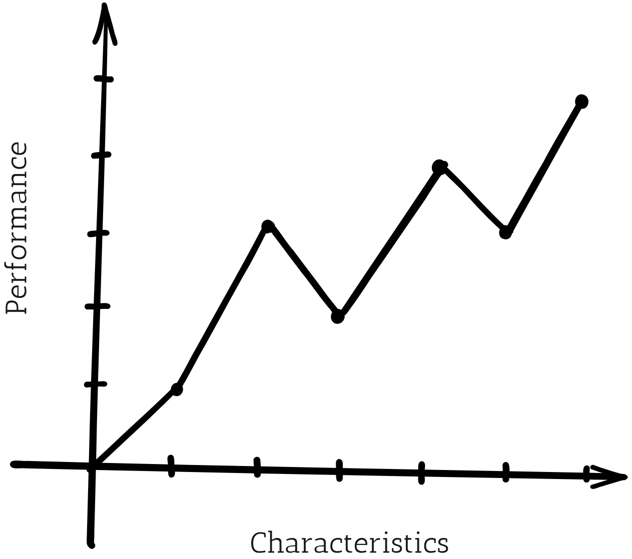
GENERATE (LINEAR) EXPLANATIONS



We aim to fit the following multivariate linear regression model:

$$\boldsymbol{y} = \boldsymbol{\epsilon} + \theta_0 + \theta_c \boldsymbol{X}_c.$$

GENERATE (LINEAR) EXPLANATIONS

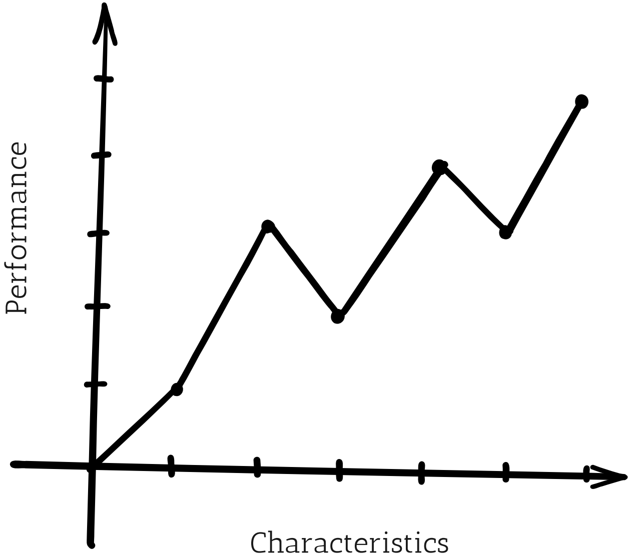


We aim to fit the following multivariate linear regression model:

$$\boldsymbol{y} = \boldsymbol{\epsilon} + \theta_0 + \theta_c \boldsymbol{X}_c.$$

Predicted performance

GENERATE (LINEAR) EXPLANATIONS



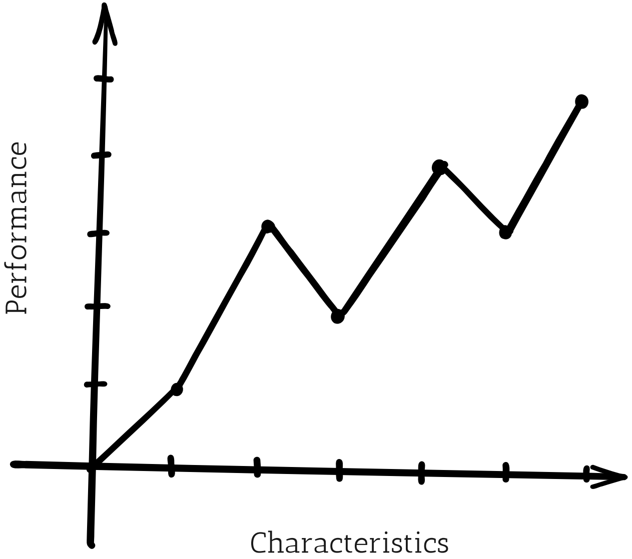
We aim to fit the following multivariate linear regression model:

$$\boldsymbol{y} = \boldsymbol{\epsilon} + \theta_0 + \theta_c \boldsymbol{X}_c.$$

Predicted performance

Prediction error

GENERATE (LINEAR) EXPLANATIONS



We aim to fit the following multivariate linear regression model:

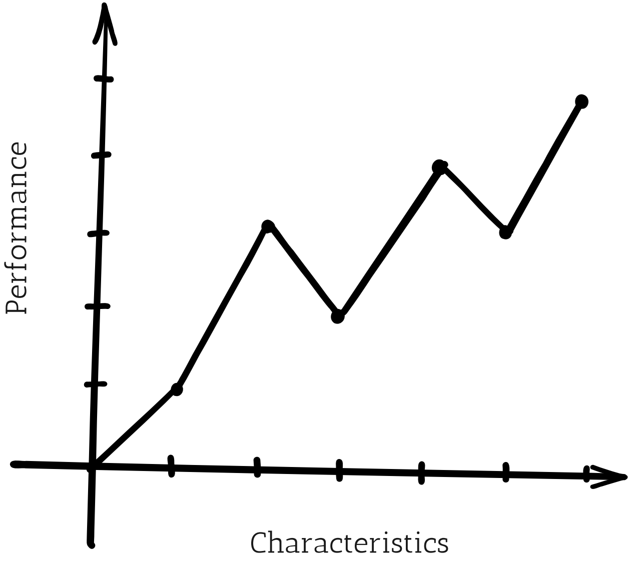
$$\boldsymbol{y} = \boldsymbol{\epsilon} + \theta_0 + \theta_c \boldsymbol{X}_c.$$

Predicted performance

Prediction error

Global bias

GENERATE (LINEAR) EXPLANATIONS



We aim to fit the following multivariate linear regression model:

$$\boldsymbol{y} = \boldsymbol{\epsilon} + \theta_0 + \theta_c \boldsymbol{X}_c.$$

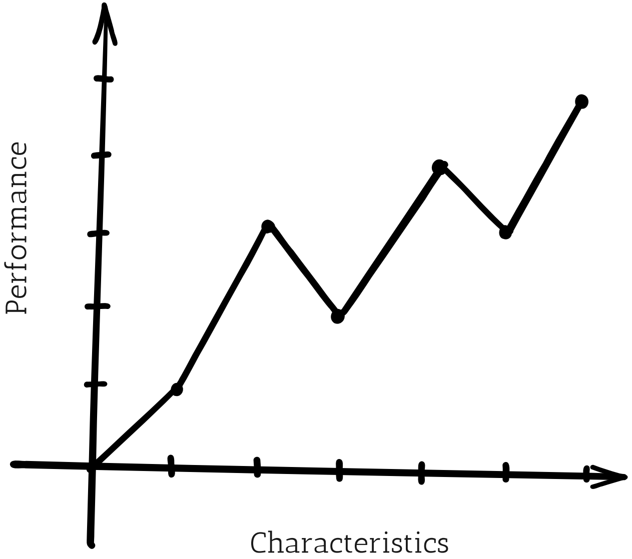
Predicted performance

Prediction error

Global bias

Weight of characteristic c

GENERATE (LINEAR) EXPLANATIONS



We aim to fit the following multivariate linear regression model:

$$\boldsymbol{y} = \boldsymbol{\epsilon} + \theta_0 + \theta_c \boldsymbol{X}_c.$$

Predicted performance

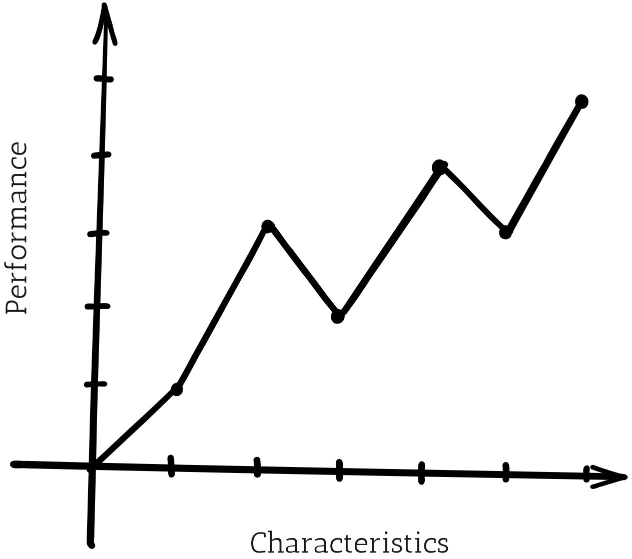
Prediction error

Global bias

Weight of characteristic c

Characteristic c for all samples

GENERATE (LINEAR) EXPLANATIONS



We aim to fit the following multivariate linear regression model:

$$\boldsymbol{y} = \boldsymbol{\epsilon} + \theta_0 + \theta_c \boldsymbol{X}_c.$$

We use the ordinary least squares (OLS) optimization model:

$$(\theta_0^*, \boldsymbol{\theta}_c^*) = \min_{\theta_0, \theta_c} \frac{1}{2} \left|\left| \boldsymbol{y} - \theta_0 - \boldsymbol{\theta}_c \boldsymbol{X}_c \right|\right|_2^2 .$$

EXPLANATORY FRAMEWORK: STEPS

STEP 1

Select models ✅

STEP 2

Select characteristics ✅

STEP 3

Select datasets ✅

STEP 4

Generate sub-datasets ✅

STEP 5

Get datasets' statistics ✅

STEP 6

Train models ✅

STEP 7

Evaluate models ✅

STEP 8

Generate explanations ✅

05

RESULTS

IMPACT OF CHARACTERISTICS

Characteristics' weights
$[\theta_0, \theta_1, \ldots, \theta_c]$

Characteristics	LightGCN		DGCF		UltraGCN		SVD-GCN	
	Yelp2018	**Gowalla**	**Yelp2018**	**Gowalla**	**Yelp2018**	**Gowalla**	**Yelp2018**	**Gowalla**
R^2(adj. R^2)	0.971(0.971)	0.979(0.978)	0.973(0.973)	0.982(0.981)	0.965(0.964)	0.860(0.858)	0.982(0.981)	0.981(0.981)
$Constant$	0.100^{***}	0.121^{***}	0.089^{***}	0.107^{***}	0.061^{***}	0.062^{***}	0.116^{***}	0.135^{***}
$SpaceSize_{log}$	0.070^{***}	0.192^{***}	0.133^{***}	0.237^{***}	-0.059^{***}	0.318^{***}	0.064^{***}	0.114^{***}
$Shape_{log}$	-0.253^{*}	-0.231	-0.282^{**}	-0.220^{*}	0.135	-0.003	-0.193	-0.232^{*}
$Density_{log}$	0.194^{***}	0.298^{***}	0.243^{***}	0.327^{***}	0.026^{*}	0.321^{***}	0.203^{***}	0.234^{***}
$Gini\text{-}U$	0.296^{**}	0.104	0.074	-0.071	-0.043	-0.931^{***}	0.136	0.143
$Gini\text{-}I$	1.362^{***}	0.681^{***}	1.108^{***}	0.560^{***}	0.605^{***}	-0.144	1.138^{***}	0.748^{***}
$AvgDegree\text{-}U_{log}$	0.390^{***}	0.605^{***}	0.518^{***}	0.673^{***}	-0.100^{*}	0.640^{***}	0.364^{***}	0.464^{***}
$AvgDegree\text{-}I_{log}$	0.137^{*}	0.374^{***}	0.235^{***}	0.453^{***}	0.034	0.637^{***}	0.171^{**}	0.231^{**}
$AvgClustC\text{-}U_{log}$	0.613^{***}	0.665^{***}	0.726^{***}	0.783^{***}	-0.077	0.706^{***}	0.613^{***}	0.496^{***}
$AvgClustC\text{-}I_{log}$	0.087	0.332^{*}	0.168	0.373^{**}	0.062	0.671^{**}	0.057	0.215
$Assort\text{-}U$	0.094^{***}	0.024^{*}	0.093^{***}	0.013	0.123^{***}	-0.019	0.080^{***}	0.010
$Assort\text{-}I$	-0.051	-0.031	-0.056^{*}	-0.055	0.001	-0.174^{***}	-0.048^{*}	-0.088^{*}

***p-value ≤ 0.001, **p-value ≤ 0.01, *p-value ≤ 0.05

IMPACT OF CHARACTERISTICS

How good is the linear regression at making predictions?

Characteristics	LightGCN		DGCF		UltraGCN		SVD-GCN	
	Yelp2018	**Gowalla**	**Yelp2018**	**Gowalla**	**Yelp2018**	**Gowalla**	**Yelp2018**	**Gowalla**
R^2(adj. R^2)	0.971(0.971)	0.979(0.978)	0.973(0.973)	0.982(0.981)	0.965(0.964)	0.860(0.858)	0.982(0.981)	0.981(0.981)
Constant	0.100***	0.121***	0.089***	0.107***	0.061***	0.062***	0.116***	0.135***
$SpaceSize_{log}$	0.070***	0.192***	0.133***	0.237***	−0.059***	0.318***	0.064***	0.114***
$Shape_{log}$	−0.253*	−0.231	−0.282**	−0.220*	0.135	−0.003	−0.193	−0.232*
$Density_{log}$	0.194***	0.298***	0.243***	0.327***	0.026*	0.321***	0.203***	0.234***
Gini-U	0.296**	0.104	0.074	−0.071	−0.043	−0.931***	0.136	0.143
Gini-I	1.362***	0.681***	1.108***	0.560***	0.605***	−0.144	1.138***	0.748***
$AvgDegree\text{-}U_{log}$	0.390***	0.605***	0.518***	0.673***	−0.100*	0.640***	0.364***	0.464***
$AvgDegree\text{-}I_{log}$	0.137*	0.374***	0.235***	0.453***	0.034	0.637***	0.171**	0.231**
$AvgClustC\text{-}U_{log}$	0.613***	0.665***	0.726***	0.783***	−0.077	0.706***	0.613***	0.496***
$AvgClustC\text{-}I_{log}$	0.087	0.332*	0.168	0.373**	0.062	0.671**	0.057	0.215
Assort-U	0.094***	0.024*	0.093***	0.013	0.123***	−0.019	0.080***	0.010
Assort-I	−0.051	−0.031	−0.056*	−0.055	0.001	−0.174***	−0.048*	−0.088*

****p-value* $\leq$ *0.001, **p-value* $\leq$ *0.01, *p-value* $\leq$ *0.05*

R^2 is, for many settings, above 95%.

The regression model can provide **good explanations**.

IMPACT OF CHARACTERISTICS

classical

Characteristics	LightGCN		DGCF		UltraGCN		SVD-GCN	
	Yelp2018	**Gowalla**	**Yelp2018**	**Gowalla**	**Yelp2018**	**Gowalla**	**Yelp2018**	**Gowalla**
R^2(adj. R^2)	0.971(0.971)	0.979(0.978)	0.973(0.973)	0.982(0.981)	0.965(0.964)	0.860(0.858)	0.982(0.981)	0.981(0.981)
Constant	0.100***	0.121***	0.089***	0.107***	0.061***	0.062***	0.116***	0.135***
$SpaceSize_{log}$	0.070***	0.192***	0.133***	0.237***	−0.059***	0.318***	0.064***	0.114***
$Shape_{log}$	−0.253*	−0.231	−0.282**	−0.220*	0.135	−0.003	−0.193	−0.232*
$Density_{log}$	0.194***	0.298***	0.243***	0.327***	0.026*	0.321***	0.203***	0.234***
Gini-U	0.296**	0.104	0.074	−0.071	−0.043	−0.931***	0.136	0.143
Gini-I	1.362***	0.681***	1.108***	0.560***	0.605***	−0.144	1.138***	0.748***
$AvgDegree\text{-}U_{log}$	0.390***	0.605***	0.518***	0.673***	−0.100*	0.640***	0.364***	0.464***
$AvgDegree\text{-}I_{log}$	0.137*	0.374***	0.235***	0.453***	0.034	0.637***	0.171**	0.231**
$AvgClustC\text{-}U_{log}$	0.613***	0.665***	0.726***	0.783***	−0.077	0.706***	0.613***	0.496***
$AvgClustC\text{-}I_{log}$	0.087	0.332*	0.168	0.373**	0.062	0.671**	0.057	0.215
Assort-U	0.094***	0.024*	0.093***	0.013	0.123***	−0.019	0.080***	0.010
Assort-I	−0.051	−0.031	−0.056*	−0.055	0.001	−0.174***	−0.048*	−0.088*

****p-value ≤ 0.001, **p-value ≤ 0.01, *p-value ≤ 0.05*

IMPACT OF CHARACTERISTICS

inverse

direct

Characteristics	LightGCN		DGCF		UltraGCN		SVD-GCN	
	Yelp2018	**Gowalla**	**Yelp2018**	**Gowalla**	**Yelp2018**	**Gowalla**	**Yelp2018**	**Gowalla**
R^2(adj. R^2)	0.971(0.971)	0.979(0.978)	0.973(0.973)	0.982(0.981)	0.965(0.964)	0.860(0.858)	0.982(0.981)	0.981(0.981)
Constant	0.100***	0.121***	0.089***	0.107***	0.061***	0.062***	0.116***	0.135***
$SpaceSize_{log}$	0.070***	0.192***	0.133***	0.237***	−0.059***	0.318***	0.064***	0.114***
$Shape_{log}$	−0.253*	−0.231	−0.282**	−0.220*	0.135	−0.003	−0.193	−0.232*
$Density_{log}$	0.194***	0.298***	0.243***	0.327***	0.026*	0.321***	0.203***	0.234***
Gini-U	0.296**	0.104	0.074	−0.071	−0.043	−0.931***	0.136	0.143
Gini-I	1.362***	0.681***	1.108***	0.560***	0.605***	−0.144	1.138***	0.748***
$AvgDegree\text{-}U_{log}$	0.390***	0.605***	0.518***	0.673***	−0.100*	0.640***	0.364***	0.464***
$AvgDegree\text{-}I_{log}$	0.137*	0.374***	0.235***	0.453***	0.034	0.637***	0.171**	0.231**
$AvgClustC\text{-}U_{log}$	0.613***	0.665***	0.726***	0.783***	−0.077	0.706***	0.613***	0.496***
$AvgClustC\text{-}I_{log}$	0.087	0.332*	0.168	0.373**	0.062	0.671**	0.057	0.215
Assort-U	0.094***	0.024*	0.093***	0.013	0.123***	−0.019	0.080***	0.010
Assort-I	−0.051	−0.031	−0.056*	−0.055	0.001	−0.174***	−0.048*	−0.088*

****p-value* $\leq$ *0.001, **p-value* $\leq$ *0.01, *p-value* $\leq$ *0.05*

Graph CF = **neighborhood** + **latent factors**.

Graph CF behaves as **factorization-based models**.

IMPACT OF CHARACTERISTICS

Characteristics	LightGCN		DGCF		UltraGCN		SVD-GCN	
	Yelp2018	**Gowalla**	**Yelp2018**	**Gowalla**	**Yelp2018**	**Gowalla**	**Yelp2018**	**Gowalla**
R^2(adj. R^2)	0.971(0.971)	0.979(0.978)	0.973(0.973)	0.982(0.981)	0.965(0.964)	0.860(0.858)	0.982(0.981)	0.981(0.981)
$Constant$	0.100***	0.121***	0.089***	0.107***	0.061***	0.062***	0.116***	0.135***
$SpaceSize_{log}$	0.070***	0.192***	0.133***	0.237***	−0.059***	0.318***	0.064***	0.114***
$Shape_{log}$	−0.253*	−0.231	−0.282**	−0.220*	0.135	−0.003	−0.193	−0.232*
$Density_{log}$	0.194***	0.298***	0.243***	0.327***	0.026*	0.321***	0.203***	0.234***
$Gini\text{-}U$	0.296**	0.104	0.074	−0.071	−0.043	−0.931***	0.136	0.143
$Gini\text{-}I$	1.362***	0.681***	1.108***	0.560***	0.605***	−0.144	1.138***	0.748***
$AvgDegree\text{-}U_{log}$	0.390***	0.605***	0.518***	0.673***	−0.100*	0.640***	0.364***	0.464***
$AvgDegree\text{-}I_{log}$	0.137*	0.374***	0.235***	0.453***	0.034	0.637***	0.171**	0.231**
$AvgClustC\text{-}U_{log}$	0.613***	0.665***	0.726***	0.783***	−0.077	0.706***	0.613***	0.496***
$AvgClustC\text{-}I_{log}$	0.087	0.332*	0.168	0.373**	0.062	0.671**	0.057	0.215
$Assort\text{-}U$	0.094***	0.024*	0.093***	0.013	0.123***	−0.019	0.080***	0.010
$Assort\text{-}I$	−0.051	−0.031	−0.056*	−0.055	0.001	−0.174***	−0.048*	−0.088*

****p-value ≤ 0.001, **p-value ≤ 0.01, *p-value ≤ 0.05*

topological

IMPACT OF CHARACTERISTICS

direct

Characteristics	LightGCN		DGCF		UltraGCN		SVD-GCN	
	Yelp2018	**Gowalla**	**Yelp2018**	**Gowalla**	**Yelp2018**	**Gowalla**	**Yelp2018**	**Gowalla**
R^2(adj. R^2)	0.971(0.971)	0.979(0.978)	0.973(0.973)	0.982(0.981)	0.965(0.964)	0.860(0.858)	0.982(0.981)	0.981(0.981)
Constant	0.100***	0.121***	0.089***	0.107***	0.061***	0.062***	0.116***	0.135***
$SpaceSize_{log}$	0.070***	0.192***	0.133***	0.237***	−0.059***	0.318***	0.064***	0.114***
$Shape_{log}$	−0.253*	−0.231	−0.282**	−0.220*	0.135	−0.003	−0.193	−0.232*
$Density_{log}$	0.194***	0.298***	0.243***	0.327***	0.026*	0.321***	0.203***	0.234***
Gini-U	0.296**	0.104	0.074	−0.071	−0.043	−0.931***	0.136	0.143
Gini-I	1.362***	0.681***	1.108***	0.560***	0.605***	−0.144	1.138***	0.748***
$AvgDegree\text{-}U_{log}$	0.390***	0.605***	0.518***	0.673***	−0.100*	0.640***	0.364***	0.464***
$AvgDegree\text{-}I_{log}$	0.137*	0.374***	0.235***	0.453***	0.034	0.637***	0.171**	0.231**
$AvgClustC\text{-}U_{log}$	0.613***	0.665***	0.726***	0.783***	−0.077	0.706***	0.613***	0.496***
$AvgClustC\text{-}I_{log}$	0.087	0.332*	0.168	0.373**	0.062	0.671**	0.057	0.215
Assort-U	0.094***	0.024*	0.093***	0.013	0.123***	−0.019	0.080***	0.010
Assort-I	−0.051	−0.031	−0.056*	−0.055	0.001	−0.174***	−0.048*	−0.088*

****p-value* $\leq$ *0.001, **p-value* $\leq$ *0.01, *p-value* $\leq$ *0.05*

This **confirms** what observed in the **models' formulations**.

High users' interactions correspond to better **performance**.

IMPACT OF CHARACTERISTICS

Characteristics	LightGCN		DGCF		UltraGCN		SVD-GCN	
	Yelp2018	**Gowalla**	**Yelp2018**	**Gowalla**	**Yelp2018**	**Gowalla**	**Yelp2018**	**Gowalla**
R^2(adj. R^2)	0.971(0.971)	0.979(0.978)	0.973(0.973)	0.982(0.981)	0.965(0.964)	0.860(0.858)	0.982(0.981)	0.981(0.981)
$Constant$	0.100***	0.121***	0.089***	0.107***	0.061***	0.062***	0.116***	0.135***
$SpaceSize_{log}$	0.070***	0.192***	0.133***	0.237***	−0.059***	0.318***	0.064***	0.114***
$Shape_{log}$	−0.253*	−0.231	−0.282**	−0.220*	0.135	−0.003	−0.193	−0.232*
$Density_{log}$	0.194***	0.298***	0.243***	0.327***	0.026*	0.321***	0.203***	0.234***
$Gini\text{-}U$	0.296**	0.104	0.074	−0.071	−0.043	−0.931***	0.136	0.143
$Gini\text{-}I$	1.362***	0.681***	1.108***	0.560***	0.605***	−0.144	1.138***	0.748***
$AvgDegree\text{-}U_{log}$	0.390***	0.605***	0.518***	0.673***	−0.100*	0.640***	0.364***	0.464***
$AvgDegree\text{-}I_{log}$	0.137*	0.374***	0.235***	0.453***	0.034	0.637***	0.171**	0.231**
$AvgClustC\text{-}U_{log}$	0.613***	0.665***	0.726***	0.783***	−0.077	0.706***	0.613***	0.496***
$AvgClustC\text{-}I_{log}$	0.087	0.332*	0.168	0.373**	0.062	0.671**	0.057	0.215
$Assort\text{-}U$	0.094***	0.024*	0.093***	0.013	0.123***	−0.019	0.080***	0.010
$Assort\text{-}I$	−0.051	−0.031	−0.056*	−0.055	0.001	−0.174***	−0.048*	−0.088*

****p-value* ≤ 0.001, ***p-value* ≤ 0.01, **p-value* ≤ 0.05

strong
direct

Values for **users-** and **items-**side are **more flattened** for **UltraGCN** and **SVD-GCN**.

IMPACT OF CHARACTERISTICS

Characteristics	LightGCN		DGCF		UltraGCN		SVD-GCN	
	Yelp2018	Gowalla	Yelp2018	Gowalla	Yelp2018	Gowalla	Yelp2018	Gowalla
R^2(adj. R^2)	0.971(0.971)	0.979(0.978)	0.973(0.973)	0.982(0.981)	0.965(0.964)	0.860(0.858)	0.982(0.981)	0.981(0.981)
Constant	0.100***	0.121***	0.089***	0.107***	0.061***	0.062***	0.116***	0.135***
$SpaceSize_{log}$	0.070***	0.192***	0.133***	0.237***	−0.059***	0.318***	0.064***	0.114***
$Shape_{log}$	−0.253*	−0.231	−0.282**	−0.220*	0.135	−0.003	−0.193	−0.232*
$Density_{log}$	0.194***	0.298***	0.243***	0.327***	0.026*	0.321***	0.203***	0.234***
Gini-U	0.296**	0.104	0.074	−0.071	−0.043	−0.931***	0.136	0.143
Gini-I	1.362***	0.681***	1.108***	0.560***	0.605***	−0.144	1.138***	0.748***
$AvgDegree\text{-}U_{log}$	0.390***	0.605***	0.518***	0.673***	−0.100*	0.640***	0.364***	0.464***
$AvgDegree\text{-}I_{log}$	0.137*	0.374***	0.235***	0.453***	0.034	0.637***	0.171**	0.231**
$AvgClustC\text{-}U_{log}$	0.613***	0.665***	0.726***	0.783***	−0.077	0.706***	0.613***	0.496***
$AvgClustC\text{-}I_{log}$	0.087	0.332*	0.168	0.373**	0.062	0.671**	0.057	0.215
Assort-U	0.094***	0.024*	0.093***	0.013	0.123***	−0.019	0.080***	0.010
Assort-I	−0.051	−0.031	−0.056*	−0.055	0.001	−0.174***	−0.048*	−0.088*

****p-value $\leq$ 0.001, **p-value $\leq$ 0.01, *p-value $\leq$ 0.05*

direct

inverse

LightGCN and **DGCF** have **larger coefficients** that **SVD-GCN**, because they **explicitly propagate** messages.

IMPACT OF CHARACTERISTICS

Characteristics	LightGCN		DGCF		UltraGCN		SVD-GCN	
	Yelp2018	Gowalla	Yelp2018	Gowalla	Yelp2018	Gowalla	Yelp2018	Gowalla
R^2(adj. R^2)	0.971(0.971)	0.979(0.978)	0.973(0.973)	0.982(0.981)	0.965(0.964)	0.860(0.858)	0.982(0.981)	0.981(0.981)
Constant	0.100***	0.121***	0.089***	0.107***	0.061***	0.062***	0.116***	0.135***
$SpaceSize_{log}$	0.070***	0.192***	0.133***	0.237***	−0.059***	0.318***	0.064***	0.114***
$Shape_{log}$	−0.253*	−0.231	−0.282**	−0.220*	0.135	−0.003	−0.193	−0.232*
$Density_{log}$	0.194***	0.298***	0.243***	0.327***	0.026*	0.321***	0.203***	0.234***
Gini-U	0.296**	0.104	0.074	−0.071	−0.043	−0.931***	0.136	0.143
Gini-I	1.362***	0.681***	1.108***	0.560***	0.605***	−0.144	1.138***	0.748***
$AvgDegree\text{-}U_{log}$	0.390***	0.605***	0.518***	0.673***	−0.100*	0.640***	0.364***	0.464***
$AvgDegree\text{-}I_{log}$	0.137*	0.374***	0.235***	0.453***	0.034	0.637***	0.171**	0.231**
$AvgClustC\text{-}U_{log}$	0.613***	0.665***	0.726***	0.783***	−0.077	0.706***	0.613***	0.496***
$AvgClustC\text{-}I_{log}$	0.087	0.332*	0.168	0.373**	0.062	0.671**	0.057	0.215
Assort-U	0.094***	0.024*	0.093***	0.013	0.123***	−0.019	0.080***	0.010
Assort-I	−0.051	−0.031	−0.056*	−0.055	0.001	−0.174***	−0.048*	−0.088*

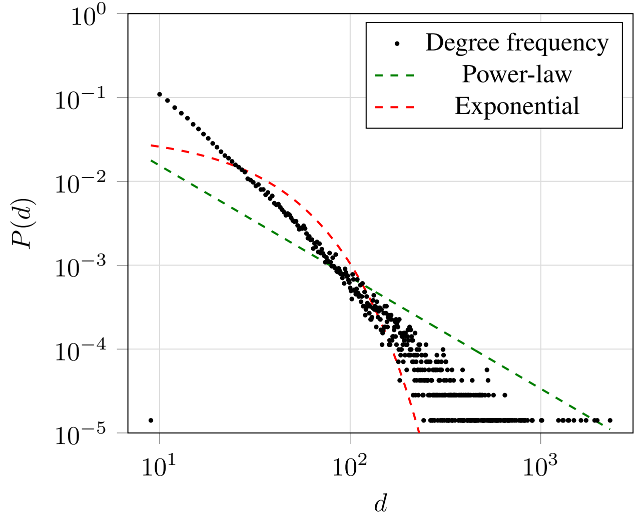
****p-value* $\leq$ *0.001, **p-value* $\leq$ *0.01, *p-value* $\leq$ *0.05*

direct

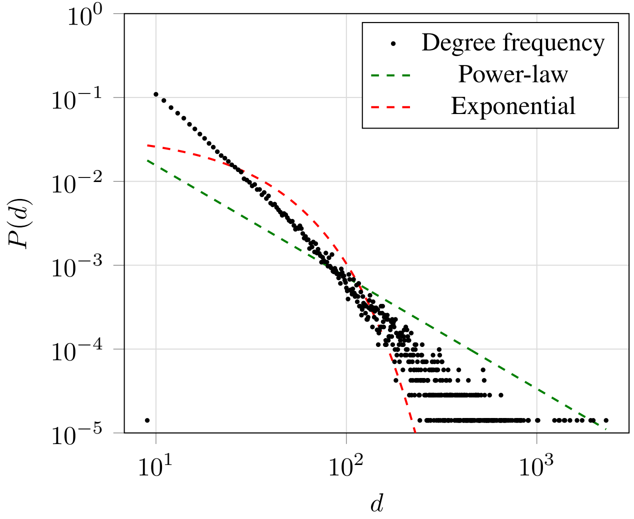
inverse

UltraGCN has **bigger coefficients**, probably because it uses the **infinite-layer** message passing.

Probability distribution of node degrees on Gowalla

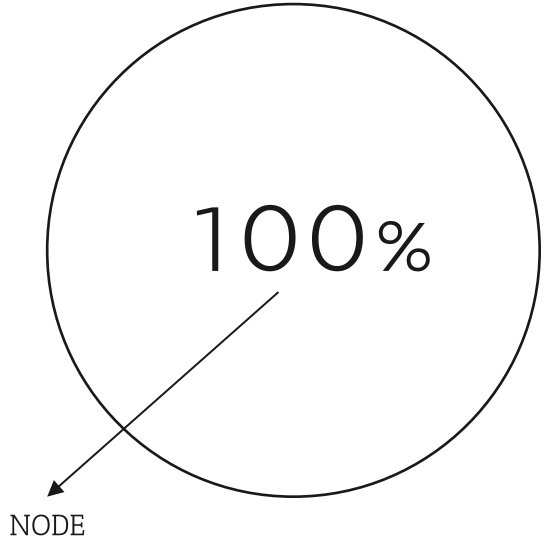


In the **worst-case** scenario, **node-dropout** drops many **high-degree** nodes; **edge-dropout**, drops all the **edges connected** to **several nodes** and thus **disconnect them** from the graph.



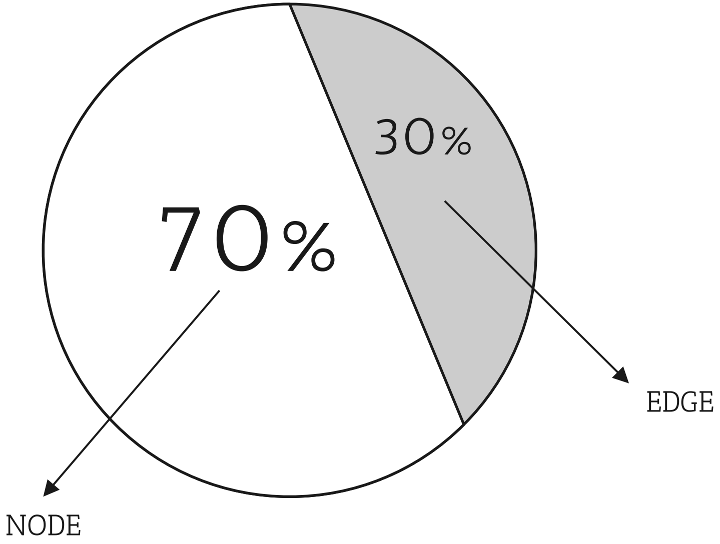
Is this **undermining** the **goodness** of the proposed **explanatory framework**?

INFLUENCE OF NODE- AND EDGE-DROPOUT



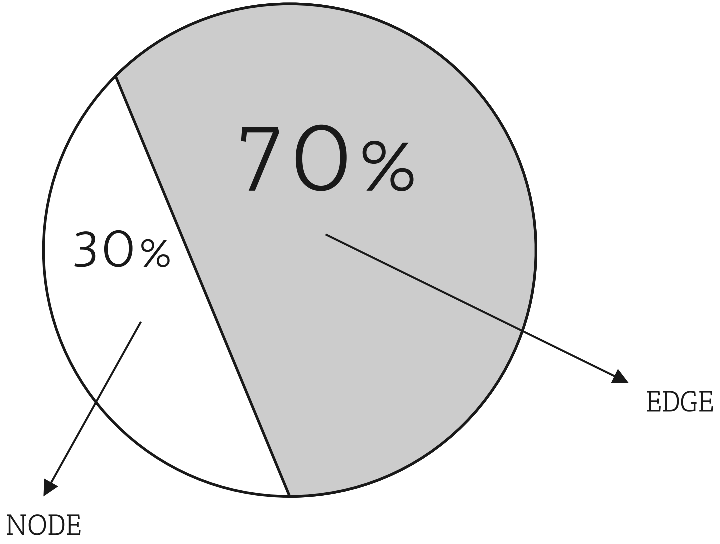
[Setting A] 100% node-dropout

INFLUENCE OF NODE- AND EDGE-DROPOUT



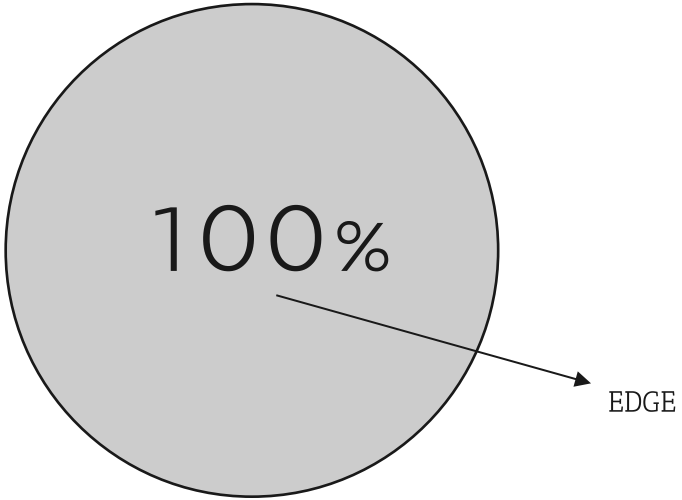
[Setting B] 70% node-dropout
30% edge-dropout

INFLUENCE OF NODE- AND EDGE-DROPOUT



[Setting C] 30% node-dropout
70% edge-dropout

INFLUENCE OF NODE- AND EDGE-DROPOUT



[Setting D] 100% edge-dropout

INFLUENCE OF NODE- AND EDGE-DROPOUT

Datasets' statistics

	Node drop ●●●	Edge drop ○○○	Node drop ●●○	Edge drop ●○○	Node drop ●○○	Edge drop ●●○	Node drop ○○○	Edge drop ●●●
	Average Sampling Statistics **Users:** 5,828 **Items:** 7,887 **Interactions:** 45,620		*Average Sampling Statistics* **Users:** 12,744 **Items:** 17,229 **Interactions:** 97,785		*Average Sampling Statistics* **Users:** 21,730 **Items:** 29,316 **Interactions:** 160,919		*Average Sampling Statistics* **Users:** 28,526 **Items:** 38,467 **Interactions:** 209,659	
Characteristics	**LightGCN**	**SVD-GCN**	**LightGCN**	**SVD-GCN**	**LightGCN**	**SVD-GCN**	**LightGCN**	**SVD-GCN**
R^2(adj. R^2)	0.597(0.583)	0.754(0.745)	0.968(0.967)	0.970(0.969)	0.986(0.985)	0.987(0.987)	0.994(0.994)	0.991(0.991)
Constant	0.179***	0.193***	0.146***	0.159***	0.098***	0.112***	0.062***	0.077***
$SpaceSize_{log}$	0.092	0.037	0.143***	0.064**	0.220***	0.136***	−0.162***	−0.048
$Shape_{log}$	−0.078	−0.118	−0.265	−0.261*	−0.175	−0.303	0.084	−0.157
$Density_{log}$	−0.079**	−0.090**	0.261***	0.195***	0.323***	0.251***	0.072**	0.040
Gini-U	−0.013	0.015	0.184	0.193	0.246	0.224	0.291***	0.225*
Gini-I	0.883***	0.884***	0.856***	0.867***	0.911***	0.940***	0.389***	0.387***
$AvgDegree\text{-}U_{log}$	0.052	0.005	0.536***	0.390***	0.631***	0.539***	−0.132*	0.070
$AvgDegree\text{-}I_{log}$	−0.026	−0.112	0.271**	0.129	0.456***	0.236*	−0.048	−0.087
$AvgClustC\text{-}U_{log}$	0.209	0.168	0.654***	0.508**	0.687**	0.647***	−0.133	0.016
$AvgClustC\text{-}I_{log}$	−0.141	−0.227	0.137	−0.007	0.436	0.172	−0.145	−0.112
Assort-U	0.008	−0.001	0.017	0.008	0.013	−0.002	0.011	−0.003
Assort-I	−0.022	−0.078**	0.028	−0.057	0.059	−0.056	0.012	−0.037

****p-value* ≤ *0.001, **p-value* ≤ *0.01, *p-value* ≤ *0.05*

Node-dropout retains **smaller portions** of the dataset than **edge-dropout**.

INFLUENCE OF NODE- AND EDGE-DROPOUT

	Node drop ●●●	Edge drop ○○○	Node drop ●●○	Edge drop ●○○	Node drop ●○○	Edge drop ●●○	Node drop ○○○	Edge drop ●●●
	Average Sampling Statistics **Users:** 5,828 **Items:** 7,887 **Interactions:** 45,620		*Average Sampling Statistics* **Users:** 12,744 **Items:** 17,229 **Interactions:** 97,785		*Average Sampling Statistics* **Users:** 21,730 **Items:** 29,316 **Interactions:** 160,919		*Average Sampling Statistics* **Users:** 28,526 **Items:** 38,467 **Interactions:** 209,659	
Characteristics	**LightGCN**	**SVD-GCN**	**LightGCN**	**SVD-GCN**	**LightGCN**	**SVD-GCN**	**LightGCN**	**SVD-GCN**
R^2(adj. R^2)	0.597(0.583)	0.754(0.745)	0.968(0.967)	0.970(0.969)	0.986(0.985)	0.987(0.987)	0.994(0.994)	0.991(0.991)
$Constant$	0.179***	0.193***	0.146***	0.159***	0.098***	0.112***	0.062***	0.077***
$SpaceSize_{log}$	0.092	0.037	0.143***	0.064**	0.220***	0.136***	−0.162***	−0.048
$Shape_{log}$	−0.078	−0.118	−0.265	−0.261*	−0.175	−0.303	0.084	−0.157
$Density_{log}$	−0.079**	−0.090**	0.261***	0.195***	0.323***	0.251***	0.072**	0.040
$Gini\text{-}U$	−0.013	0.015	0.184	0.193	0.246	0.224	0.291***	0.225*
$Gini\text{-}I$	0.883***	0.884***	0.856***	0.867***	0.911***	0.940***	0.389***	0.387***
$AvgDegree\text{-}U_{log}$	0.052	0.005	0.536***	0.390***	0.631***	0.539***	−0.132*	0.070
$AvgDegree\text{-}I_{log}$	−0.026	−0.112	0.271**	0.129	0.456***	0.236*	−0.048	−0.087
$AvgClustC\text{-}U_{log}$	0.209	0.168	0.654***	0.508**	0.687**	0.647***	−0.133	0.016
$AvgClustC\text{-}I_{log}$	−0.141	−0.227	0.137	−0.007	0.436	0.172	−0.145	−0.112
$Assort\text{-}U$	0.008	−0.001	0.017	0.008	0.013	−0.002	0.011	−0.003
$Assort\text{-}I$	−0.022	−0.078**	0.028	−0.057	0.059	−0.056	0.012	−0.037

****p-value ≤ 0.001, **p-value ≤ 0.01, *p-value ≤ 0.05*

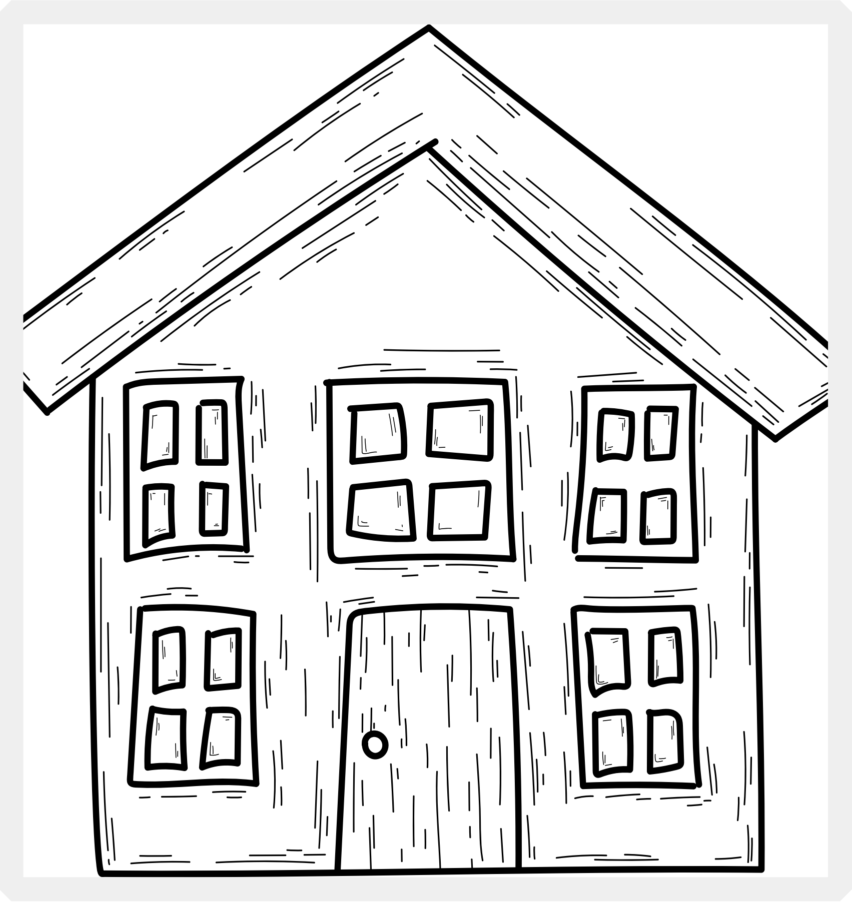
Meaningful learned dependencies (this setting is **like ours**).

INFLUENCE OF NODE- AND EDGE-DROPOUT

	Node drop ●●●	Edge drop ○○○	Node drop ●●○	Edge drop ●○○	Node drop ●○○	Edge drop ●●○	Node drop ○○○	Edge drop ●●●
	Average Sampling Statistics **Users:** 5,828 **Items:** 7,887 **Interactions:** 45,620		*Average Sampling Statistics* **Users:** 12,744 **Items:** 17,229 **Interactions:** 97,785		*Average Sampling Statistics* **Users:** 21,730 **Items:** 29,316 **Interactions:** 160,919		*Average Sampling Statistics* **Users:** 28,526 **Items:** 38,467 **Interactions:** 209,659	
Characteristics	**LightGCN**	**SVD-GCN**	**LightGCN**	**SVD-GCN**	**LightGCN**	**SVD-GCN**	**LightGCN**	**SVD-GCN**
R^2(adj. R^2)	0.597(0.583)	0.754(0.745)	0.968(0.967)	0.970(0.969)	0.986(0.985)	0.987(0.987)	0.994(0.994)	0.991(0.991)
$Constant$	0.179***	0.193***	0.146***	0.159***	0.098***	0.112***	0.062***	0.077***
$SpaceSize_{log}$	0.092	0.037	0.143***	0.064**	0.220***	0.136***	−0.162***	−0.048
$Shape_{log}$	−0.078	−0.118	−0.265	−0.261*	−0.175	−0.303	0.084	−0.157
$Density_{log}$	−0.079**	−0.090**	0.261***	0.195***	0.323***	0.251***	0.072**	0.040
$Gini\text{-}U$	−0.013	0.015	0.184	0.193	0.246	0.224	0.291***	0.225*
$Gini\text{-}I$	0.883***	0.884***	0.856***	0.867***	0.911***	0.940***	0.389***	0.387***
$AvgDegree\text{-}U_{log}$	0.052	0.005	0.536***	0.390***	0.631***	0.539***	−0.132*	0.070
$AvgDegree\text{-}I_{log}$	−0.026	−0.112	0.271**	0.129	0.456***	0.236*	−0.048	−0.087
$AvgClustC\text{-}U_{log}$	0.209	0.168	0.654***	0.508**	0.687**	0.647***	−0.133	0.016
$AvgClustC\text{-}I_{log}$	−0.141	−0.227	0.137	−0.007	0.436	0.172	−0.145	−0.112
$Assort\text{-}U$	0.008	−0.001	0.017	0.008	0.013	−0.002	0.011	−0.003
$Assort\text{-}I$	−0.022	−0.078**	0.028	−0.057	0.059	−0.056	0.012	−0.037

****p-value ≤ 0.001, **p-value ≤ 0.01, *p-value ≤ 0.05*

Not-statistically-significant at the extremes.



06

TAKE-HOME MESSAGES

WHAT WE HAVE LEARNED

01

LATENT-FACTORS *VS.* NEIGHBORHOOD

WHAT WE HAVE LEARNED

01

LATENT-FACTORS *vs.* NEIGHBORHOOD

02

NODE DEGREE

WHAT WE HAVE LEARNED

01

LATENT-FACTORS *vs.* NEIGHBORHOOD

02

NODE DEGREE

03

CLUSTERING COEFFICIENT

WHAT WE HAVE LEARNED

01

LATENT-FACTORS *vs.* NEIGHBORHOOD

02

NODE DEGREE

03

CLUSTERING COEFFICIENT

04

DEGREE ASSORTATIVITY

WHAT WE HAVE LEARNED

01

LATENT-FACTORS *vs.* NEIGHBORHOOD

02

NODE DEGREE

03

CLUSTERING COEFFICIENT

04

DEGREE ASSORTATIVITY

05

NODE- *vs.* EDGE-DROPOUT

REFERENCES

[Malitesta et al.] A Topology-aware Analysis of Graph Collaborative Filtering. CoRR abs/2308.10778 (2023)

[Wei et al.] **Designing the Topology of Graph Neural Networks: A Novel Feature Fusion Perspective. WWW 2022: 1381-1391**

[Castellana and Errica] Investigating the Interplay between Features and Structures in Graph Learning. CoRR abs/2308.09570 (2023)

[Shi et al.] Topology and Content Co-Alignment Graph Convolutional Learning. IEEE Trans. Neural Networks Learn. Syst. 33(12): 7899-7907 (2022)

[Adomavicius and Zhang] Impact of data characteristics on recommender systems performance. ACM Trans. Manag. Inf. Syst. 3(1): 3:1-3:17 (2012)

[Deldjoo et al.] How Dataset Characteristics Affect the Robustness of Collaborative Recommendation Models. SIGIR 2020: 951-960

[Latapy et al.] Basic notions for the analysis of large two-mode networks. Soc. Networks 30(1): 31-48 (2008)

[M. E. J. Newman] Mixing patterns in networks. Phys. Rev. E, 67:026126, Feb 2003. doi: 10.1103/PhysRevE.67.026126

THANKS!

CREDITS: This presentation template was created by **Slidesgo**, and includes icons by **Flaticon**, and infographics & images by **Freepik**

Scan me!